Thesis

Blending Model Output with satellite-based and in-situ observations to produce high-resolution estimates of population exposure to wildfire smoke

Submitted by

William Lassman

Department of Atmospheric Science

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2016

Master's Committee:

Advisor: Jeffrey Pierce

Emily Fischer
Russ Schumacher
Sheryl Magzamen
Gabriele Pfister

Abstract

Blending Model Output with satellite-based and in-situ observations to produce high-resolution estimates of population exposure to wildfire smoke

In the western US, emissions from wildfires and prescribed fire have been associated with degradation of regional air quality. Whereas atmospheric aerosol particles with aerodynamic diameters less than 2.5 μm ($PM_{2.5}$) have known impacts on human health, there is uncertainty in how particle composition, concentrations, and exposure duration impact the associated health response. Due to changes in climate and land-management, wildfires have increased in frequency and severity, and this trend is expected to continue. Consequently, wildfires are expected to become an increasingly important source of $PM_{2.5}$ in the western US. While composition and source of the aerosol is thought to be an important factor in the resulting human health-effects, this is currently not well-understood; therefore, there is a need to develop a quantitative understanding of wildfire-smoke-specific health effects. A necessary step in this process is to determine who was exposed to wildfire smoke, the concentration of the smoke during exposure, and the duration of the exposure. Three different tools are commonly used to assess exposure to wildfire smoke: in-situ measurements, satellite-based observations, and chemical-transport model (CTM) simulations, and each of these exposure-estimation tools have associated strengths and weakness.

In this thesis, we investigate the utility of blending these tools together to produce highly accurate estimates of smoke exposure during the 2012 fire season in Washington for use in an epidemiological case study. For blending, we use a ridge regression model, as well as a geographically weighted ridge regression model. We evaluate the performance

of the three individual exposure-estimate techniques and the two blended techniques using Leave-One-Out Cross-Validation. Due to the number of in-situ monitors present during this time period, we find that predictions based on in-situ monitors were more accurate for this particular fire season than the CTM simulations and satellite-based observations, so blending provided only marginal improvements above the in-situ observations. However, we show that in hypothetical cases with fewer surface monitors, the two blending techniques can produce substantial improvement over any of the individual tools.

Acknowledgements

First, I would like to thank to Dr. Bonne Ford for her mentorship and help throughout my master's work.

I'd also like to thank my advisor Dr. Jeff Pierce, as well as my committee: Dr. Emily Fischer, Dr. Russ Schumacher, Dr. Sheryl Magzamen, and Dr. Gabi Pfister for their guidance. I'd like to thank the Pierce group members past and present (Anna, Emily, Jack, Kate, Kim, Landan, and Steve) as well as the Fischer group students (Jake, Jakob, Jared, Steve, Zitely) for the camaraderie and occasional research discussions. Finally, I'd like to thank all of my friends, family, and teammates for keeping their support and patience during this process.

Funding for this research was provided by the NASA Applied Sciences program, number **NNX15AG35G** and the Joint Fire Science program, grant number **JFSP 13-1-01-4**.

This thesis is typset in LaTeX using a document class designed by Leif Anderson.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

Ambient air pollution is one of the leading causes of premature mortality worldwide [1]. Atmospheric aerosols are a major component of air pollution; fine particles with an aerodynamic diameter less than 2.5 μm ($PM_{2.5}$), which are small enough to penetrate deep into the human lung [2], are associated with many of the negative health effects attributed to air pollution. Exposure to $PM_{2.5}$ is associated with development of respiratory diseases through the free radical peroxidation of the lung lining [3] as well as general inflammation injury to the lining of the lung [4]. Furthermore, $PM_{2.5}$ has also been linked to cardiovascular disease through several different disease mechanisms such as systemic inflammation, as well as transport of particles through the lung into the bloodstream [5].

Atmospheric aerosols have many naturally occurring sources and are an important part of the earth-climate system. However, during the industrial revolution, anthropogenic sources of $PM_{2.5}$, such as sulfate, nitrate, and carbonaceous aerosol from combustion processes, increased dramatically in many portions of the world [6]. In many parts of the world, anthropogenic emissions are the dominant source of atmospheric aerosols, and are responsible for most of the resulting health effects attributable to air pollution [7]. A study by Lelieveld et al. [7] estimated that of the 3.3 million premature deaths attributable to outdoor air pollution in 2010, anthropogenic emissions from residential energy, agriculture, power generation, industry, and land traffic are together responsible for 77%, while natural sources are responsible for 18% and biomass burning is responsible for the remaining 5%. [7]. This apportionment of mortality to different sources of air pollution relies on a concentration-response function, which estimate the association between incident disease and environmental exposure. These

concentration-response functions were determined based on long-term exposure to bulk $PM_{2.5}$ mass. Concentration-response functions can be used to estimate changes in all-cause mortality, mortality due to a specific disease (e.g. cardiovascular disease), or morbidity.

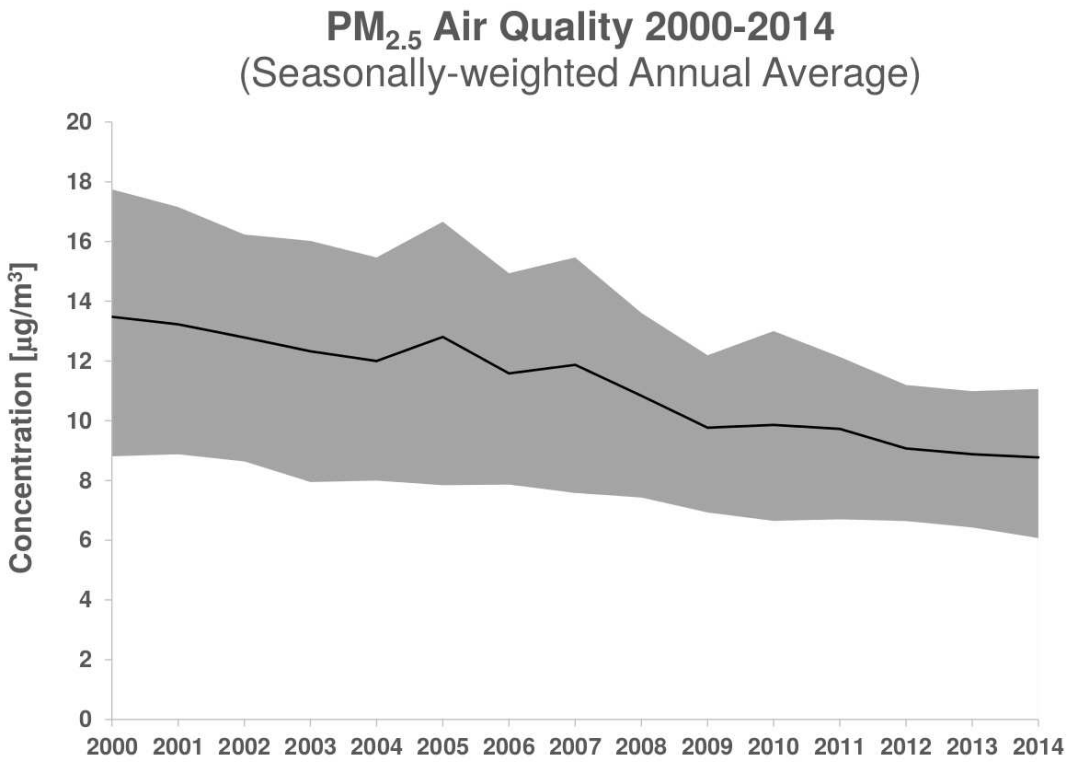

FIGURE 1.1. Annual Average $PM_{2.5}$ concentrations in the USA from 2000-2014. Data taken from the US EPA.

In the United States, $PM_{2.5}$ is considered a criteria pollutant and is regulated under the Clean Air Act; consequently, surface concentrations of $PM_{2.5}$ have declined steadily since the early 1990s [8] resulting in improved health outcomes [9]. As shown in Figure 1.1, seasonally weighted annual average surface $PM_{2.5}$ concentrations have decreased by 30% in the United States in recent years. However, transboundary (i.e. international) and non-anthropogenic sources are not subject to the same controls and are more challenging to mitigate. Figure 1.2 shows the relative contributions of anthropogenic and naturally occurring sources of primary $PM_{2.5}$ in the United States. One of the largest contributors is wildfires. Presently, 5% of

global premature mortality due to air quality has been attributed to emissions from biomass burning, which includes wildfires and prescribed fires [7]). It is important to note that these mortality estimates are based on concentration-response functions for total $PM_{2.5}$ mass, and do not account for different toxicities from different types of aerosol; furthermore, these estimates do not include morbidity.

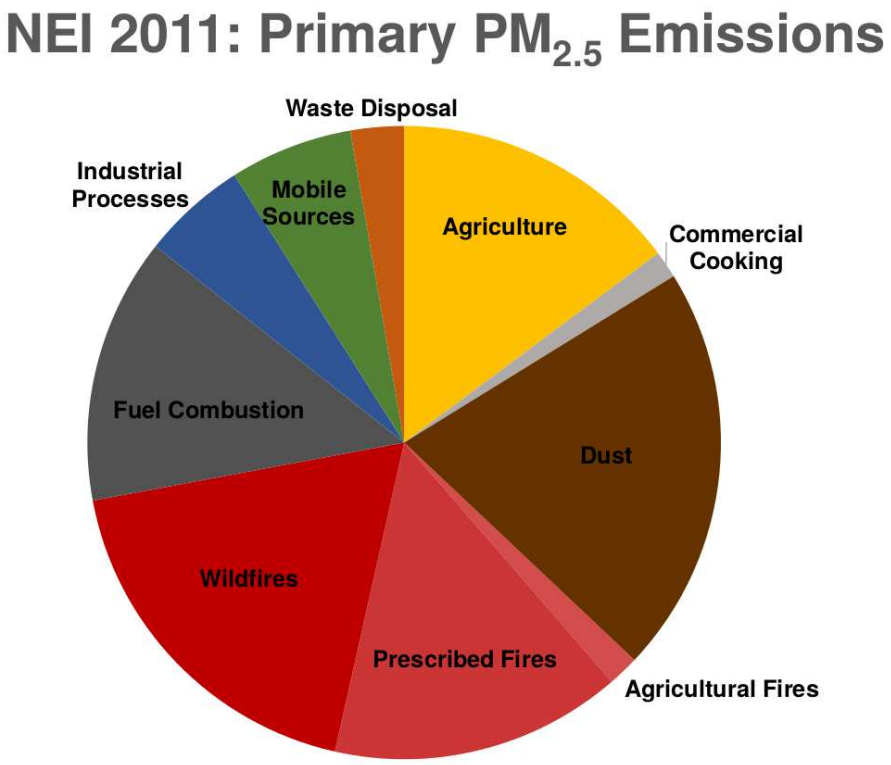

FIGURE 1.2. 2011 National Emissions Inventory (NEI) showing a breakdown of primary $PM_{2.5}$ emissions by source. Data taken from the US EPA.

Wildfires are an important component of the earth-biosphere-atmosphere system. In many ecosystems, fires play an important role in long-term stability of those ecosystems [10]. However, aggressive fire suppression by humans, other land-management practices, and changes in global and regional climate have altered the fire regime in many parts of the western United States [11], resulting in an increase in the number, size, and intensity of wildland fires in the last 60 years. These increases are projected to continue [12], resulting in

an increase in the importance of biomass-burning aerosol in the western US. As anthropogenic emissions of $PM_{2.5}$ continue to decrease, wildfires are projected to become the largest source of $PM_{2.5}$ throughout the western US [13] [14].

Presently, wildfires can produce concentrations of $PM_{2.5}$ that exceed the National Ambient Air Quality Standards for this pollutant (NAAQS; greater than $35\mu g\ m^{-3}$ for 24 hours) [15] both near [16] and far from the fire [17] [18]. While NAAQS exceedences due to wildfire smoke are deemed exceptional events and are not considered a violation of the NAAQS, the health effects of these exceedences remain. During wildfire-smoke events, people have reported symptoms including irritated eyes, irritated respiratory systems, asthma, and acute exacerbations of chronic conditions [19] [20]. While the health effects of ambient $PM_{2.5}$ have been studied extensively (approximately 17,000 publications in ISI Web of Science Search), the wildfire-specific health response is not as well studied (approximately 80 publications in ISI Web of Science). Because wildfire smoke episodes are transient [16], and because the $PM_{2.5}$ originating from biomass burning has a different composition than anthropogenic aerosols [21], the health response to wildfire smoke exposure may be different than to ambient $PM_{2.5}$ exposure. The development of wildfire-specific concentration-response functions is in its infancy. As wildfires become an increasingly important source of $PM_{2.5}$ in the western US, there is a need to develop understanding of health-effects specific to wildfire-smoke $PM_{2.5}$.

Environmental epidemiology uses many different techniques for determining health effects of an environmental pollutant and for creating the concentration-response functions described above. However, most of these study designs require accurate estimates of pollutant exposure, reliable health data, and a large exposed population to ensure detection of a statistically significant change in health outcomes between the hazard and non-hazard periods. Because wildfire smoke events tend to occur in sparsely populated regions, statistical power can

be challenging; to compensate for this, reducing exposure misclassification by producing increasingly accurate estimates of spatiotemporal smoke concentrations is necessary for the epidemiology to be effective. The focus of this study is to produce accurate estimates of exposure to wildfire smoke across large areas, which can be used in an epidemiology study such as a case-crossover study design [22].

Previous studies on health effects due to wildfire-smoke exposure have relied on three different tools to estimate smoke $PM_{2.5}$ exposure: in-situ monitors (e.g., [20]), satellite observations (e.g. [23]), and chemical-transport models (CTMs) (e.g., [24]). Each of these tools has a different set of strengths and weaknesses, and wildfire smoke presents different challenges for all three of these exposure-estimation techniques.

Ground-based in-situ measurements of $PM_{2.5}$ are made routinely throughout the US. The EPA Air Quality System (https://www.epa.gov/aqs) consolidates data from a number of EPA, state, and local monitors. Many different measurement techniques are used, such as Tapered Element Oscillating Microbalances (TEOMs) and Beta Attenuator Mass Monitors (BAMs), which produce hourly measurements, and filter-based measurements, which produce daily measurements. While there are measurement errors and biases associated with each measurement technique, the measurements can provide a high degree of accuracy relative to models and satellites. While in-situ measurements provide accurate information about $PM_{2.5}$ concentrations at a single point in space, the spatial density of sites is inconsistent throughout the western US, and monitors are often insufficient to resolve steep spatial gradients associated with wildfire smoke. Furthermore, some measurement sites only record data every third day which can fail to measure changes in smoke concentrations that often occur on shorter timescales.

Satellite observations complement in-situ measurements due to their ability to make observations over large spatial areas. However, most satellite observational techniques produce estimates of Aerosol Optical Depth (AOD), which is the integrated extinction of light due to the total column aerosol mass and is not necessarily indicative of surface $PM_{2.5}$ concentrations. Satellite observations of AOD have been used in numerous studies of air quality and health (e.g., [23]) and are often combined with models or measurements for these studies (e.g., [25], [26], and [27]).

To fill the gap between satellite and in-situ observations, modelling tools are used (e.g. [25]). A wide range of CTMs are used to estimate surface $PM_{2.5}$ based on information about emissions, chemical processing of emissions, and transport due to meteorology. CTMs produce spatially continuous information about many different species, making these tools attractive for use in health studies [24]. However, CTMs are limited by our understanding of the underlying physical processes, as well as our ability to represent these processes in a computationally practical model. Specifically, accurate modelling of wildfire smoke requires detailed knowledge of fuel loading and moisture, fire behavior, and thermodynamic structure of the atmosphere to correctly predict emission species and mass, as well as pollutant injection height, which are all crucial to predicting surface $PM_{2.5}$ in smoke plumes. With additional uncertainties in meteorology and subgrid-scale parameterizations affecting transport, accurate modelling of wildfire smoke is a challenging endeavor.

Given the limitations of these individual tools, several studies have examined the utility of combining information from each of these tools, as well as other datasets. One study combines CTM simulation results with AOD products from two satellite platforms, in-situ monitor data, as well as other land-use data and distance to fire data in a data-adaptive machine-learning framework consisting of 11 different algorithms [28]. Another study combined

various remote-sensing products with previous days in-situ measurements as well as a dispersion model in a regression model to predict surface $PM_{2.5}$ [29]. An additional study compared the observed health signal using models, monitors, and satellites separately to determine exposure and found that the exact technique used to determine exposure can impact the health response significantly [30]. A review of wildfire-smoke exposure literature noted that while most studies find a signal for respiratory health effects of wildfire smoke exposure, more studies on mortality and cardiovascular morbidity are needed, and a limitation of the studies is the difficulty estimating pollution exposure specific to wildfires [31].

This thesis examines the skill of combining satellite and in-situ observations with CTM model output to reduce exposure misclassification. In Chapter 2, we describe which fires were analyzed, what modelling tools and observational platforms were used, how the data sources were combined, and how the different techniques were evaluated. In Chapter 3, we share results from the different techniques and discuss how and when each technique is applicable. Finally, we discuss future work to improve upon the results, presented in Chapter 4.

CHAPTER 2

Tools and Methods

2.1. Fires Studied

Two different wildfire events were selected to develop the techniques for determining wildfire-smoke exposure: the Washington State 2012 wildfire season and the High Park Fire (Fort Collins, Colorado June 2012). The Washington State fires will be the main focus of this thesis due to the long duration, large exposed population, and while not included here, our epidemiologist collaborators currently have Washington 2012 health data available. The High Park Fire is used in this thesis to evaluate WRF-Chem model setup choices. Maps of fire locations relative to populated areas are shown in Figure 2.1.

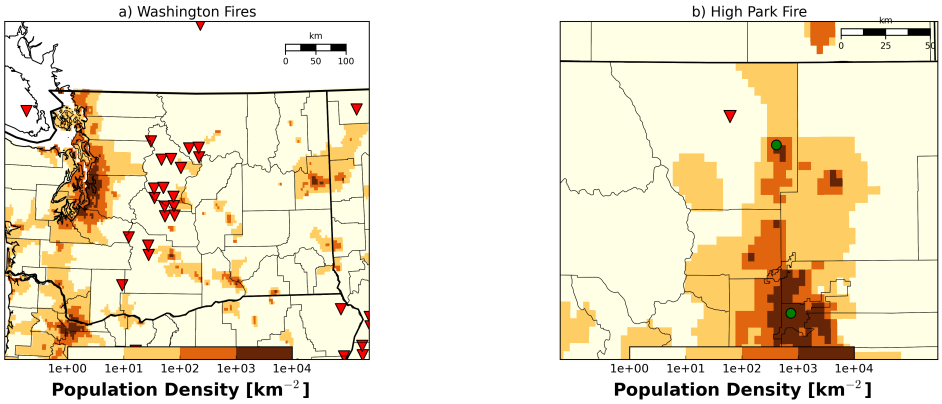

Figure 2.1. Map of major fire locations (triangles) for (a) the Washington wildfire season and (b) High Park Fire relative to population centers (green circles). In panel a, we included 25% of fires that were active 14-21 September 2012, that were large enough to be included in the FINN biomass burning inventory. This choice was made to reduce clutter in the figure. This schematic is only to illustrate where the bulk of the fire activity was located relative to population, a more complete list of fire activity in this time period can be found the FINN biomass burning emissions inventory.

The 2012 Washington wildfires were a series of fires ignited by a lightning storm in the east Cascade mountain range. During September and October, a series of fires burned close

to 260,000 acres. Smoke from the fires severely impacted visibility and air quality in much of the central and eastern parts of the state. It is estimated that the cost of fighting several of the fires exceeded $65 million. During the month of September, there were additional active surface monitors deployed to augment the EPA AQS monitors, that made hourly measurements of $PM_{2.5}$ concentrations. This fire event was selected because the relatively large number of monitors allowed more opportunities to constrain model simulations, as well as produce new estimates of surface $PM_{2.5}$ based on the measured values.

The High Park Fire was ignited by lightning and first detected on June 9th, 2012, and was 100% contained on June 30th 2012. It burned 87,284 acres, becoming the second largest-area fire in the history of Colorado, and resulted in the destruction of 259 homes and 1 fatality (from the fire itself). High Park was selected as a fire to simulate because it was a single isolated fire, which allows us to more easily determine the effect changes in the simulation parameters have on the smoke plume location and concentrations.

2.2. Exposure-Estimation Tools

2.2.1. Chemical Transport Model: WRF-Chem. Computer models are an increasingly popular choice for estimating/forecasting air quality and pollution exposure. Models are able to produce spatially and temporally continuous estimates for many different chemical species and meteorological variables at high resolution. However, models require simplifying assumptions and do not represent all of the physics and chemistry present in the real world. These approximations can compromise the accuracy and usefulness of simulation results. With any model, it is important to understand how these assumptions affect the simulation results.

For our model simulations of wildfires, we use the Weather Research and Forecasting model with Chemistry (WRF-Chem) [32]. The WRF model is an atmospheric fluid-dynamics model that solves fluid-mechanics equations with radiative transfer in the atmosphere and is used in operational weather forecasting as well as for atmospheric research. WRF-Chem uses the WRF meteorological model with chemistry modules simulating the chemical processes in the atmosphere including feedbacks on radiative transfer and cloud processes [33].

For this study, our base-case model settings and inputs are: Global Forecast System (GFS) analysis dataset for meteorological initial/boundary conditions, Model for Ozone And Related chemical Tracers (MOZART) chemical initial/boundary conditions, and MOZART chemistry mechanisms [34] with the Goddard Chemistry Aerosol Radiation and Transport (GOCART) Aerosol model [35], 15 km horizontal resolution, and Yonsei University (YSU) boundary-layer parameterization [36], and the Thompson microphysics parameterization [37]. We use the EPA National Emissions Inventory (NEI) 2011 [38] for anthropogenic emissions, the Model for Emissions of Gases and Aerosols from Nature (MEGAN) [39] to model biogenic emissions, and the Fire Inventory from NCAR (FINN) version 1.5 [40] to model emissions from biomass burning. FINN is a biomass-burning emissions inventory that includes observations of fire locations and radiative power derived from remote sensing tools on a daily timescale, with tabulated information about fuel loading and emission factors to determine the mass of each species emitted. A major advantage of FINN over other biomass-burning emissions inventory is that it is updated regularly with recent fire events, allowing rapid simulation of recent fires. For smoke emissions from biomass burning, WRF-Chem uses an online 1-D plume-rise parameterization to determine pollutant injection height [41]. For our Washington simulations, we begin simulations on June 1st 2012 at 00Z,

and reinitialize our meteorology every twenty four hours, while using the previous day's simulation output to initialize all chemistry fields for the next day.

We perform a series of sensitivity simulations of the High Park Fire to determine the consequences and sensitivity of the model to these choices; we test the sensitivity to the model resolution, boundary-layer parameterization, cloud-microphysics parameterization, meteorology-reanalysis dataset, biomass-burning emissions inventory, and biomass-burning emissions diurnal profile. The details of the sensitivity simulations can be found in Appendix A. The results of the sensitivity study show that the model simulations are highly sensitive to the biomass-burning emissions inventory, diurnal profile of the emissions, and model resolution, and relatively insensitive to the reanalysis dataset and parameterization choices. Due to the scarcity of available surface measurements of $PM_{2.5}$ during the High Park Fire, it was not possible to determine the optimal parameter settings. Furthermore, the choice of parameters is likely to differ based on region, meteorological conditions, and wildfire. We therefore use the base-case simulation settings for the Washington 2012 simulation. Figure 2.2 shows the domain used for the Washington 2012 WRF-Chem simulations.

2.2.2. In-Situ Monitors and Ordinary Kriging. In-situ monitor data for the state of Washington, as well as northern Oregon, and western Idaho and Montana were obtained from the EPA AQS for the months of July-October of 2012. During smoky periods, the Washington Department of Ecology deployed additional $PM_{2.5}$ monitors, which were also included in our analysis. Additional in-situ measurements from Canada were obtained from Dr. Sarah Henderson at the University of British Columbia School of Population and Public Health. All networks of in-situ measurements use a combination of TEOMs and BAMs. The various surface observations are at different temporal resolutions; temporary monitors reported hourly $PM_{2.5}$ readings, while the EPA-network sites provide measurements every

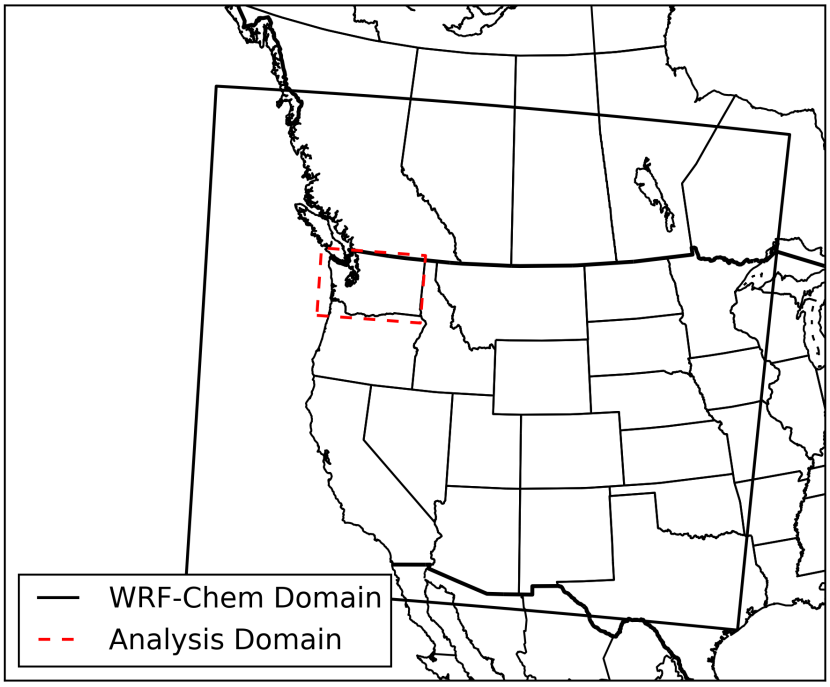

FIGURE 2.2. WRF-Chem simulation domain (black) and inner Analysis domain (red). Both domains use the same 15x15 km grid, but blended surface concentration estimates were only produced inside the analysis domain.

day, or every 3 days. Hourly data are averaged to produce 24-hour-average concentrations. Figure 5 shows a map of all surface monitors included in the analysis; red sites are located inside of our predictive domain where we evaluated model performance, while blue sites were included as inputs to our predictive model to help reduce boundary effects inside the area of interest.

While previous studies (e.g. [2]) have relied on nearest monitor readings to determine exposure, this reduces the sample size of the exposed population, and many health effects (especially cardiovascular effects) may be obscured by a lack of statistical power [42]. Often, an interpolation technique is used to produce spatially continuous estimates of the surface $PM_{2.5}$ based on in-situ monitors in order to increase sample size for a health study [43] [44].

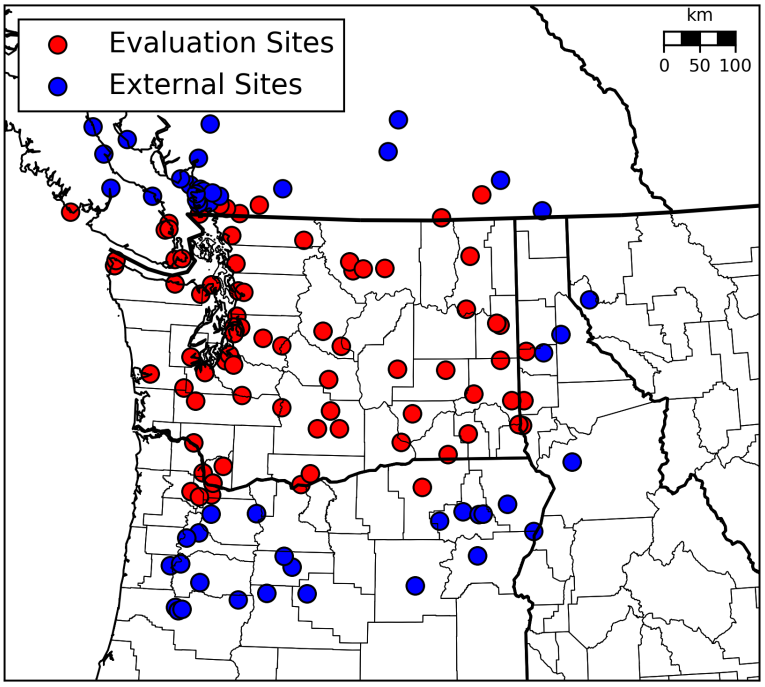

FIGURE 2.3. Surface sites included in the Kriging analysis. Red sites were inside our domain of analysis and were also used to evaluate the various prediction schemes that were used. Blue sites are outside the model domain and were only used as inputs to the kriging model.

Ordinary kriging is a geostatistical interpolation technique that has been used for interpolating air-quality data [45] [46] [47]. Kriging uses the correlation between measurement sites and the sites respective distances to estimate the rate of decay of spatial autocorrelation. This is done by calculating "semivariance" (γ), as defined in Equation 1

$$\gamma(h) = \frac{1}{2}\frac{1}{n(h)}\sum_{i=1}^{n(h)}(z(x_i+h)-z(x_i))^2 \tag{1}$$

where h is the distance, n are all data points at that distance, z is the value being interpolated, x is the point within the domain. Semivariance between each in-situ measurement site is

calculated, and used to construct a "semivariogram", which is semivariance as a function of distance. We then choose an analytical function to model the empirical semivariogram. We use a spherical semivariogram, shown in Equation 2:

$$\gamma(h) = (s - n)\left(\left(\frac{3h}{2r} - \frac{h^3}{2r^3}\right)1_{0,r}(h) + 1_{r,\infty}\right) + n1_{0,\infty}(h) \tag{2}$$

where h is distance, and s, r, and n are shape parameters to optimize the fit. The 'sill' parameter (s) determines the semivariogram function value at infinite distance, the 'nugget' parameter (n) determines the semivariogram function at distance 0, and range parameter r determines the length scale at which semivariance decays to the sill value [45]. The function $1_A(h)$ is defined such that for $1_A(h) = 1\,; h \in A$ and $1_A(h) = 0\,; h \notin A$.

In our interpolation, we use sill and nugget parameter values of 3.0 and 0.1, and a range parameter of 8.0 degrees respectively. Interpolations with ordinary kriging of our data were very sensitive to the choice of semivariogram shape and parameters. We chose to use a spherical semivariogram shape because it performed the best of common semivariogram models, and also because this shape has been shown to perform well in other air-pollution applications [48]. We selected values for the parameters that produced the best estimates according to our evaluation procedure (Section 2.4).

An advantage of kriging over many other spatial-interpolation techniques (e.g. Inverse Distance Weighting) is that kriging accounts for spatial heterogeneity in measurement site placement, so no declustering is required [49]; in other interpolation techniques, a monitoring site's contribution to the interpolation estimate is calculated based on distance to a point where the value is known (e.g. an in-situ monitor site). However, if multiple monitors are close together, they each contribute independently to the interpolated estimate, even though we know that these observations are not independent; this can bias a spatial interpolation

towards regions that are more extensively monitored (i.e. urban areas), over the smaller number of in-situ monitors that are close by. Because kriging uses the spatial autocorrelation between a given surface site and the interpolated location to calculate the monitor‘s contribution to the interpolation, this is implicitly taken into account.

To produce spatially continuous surface-$PM_{2.5}$ estimates, we used the PyKrige module [50] to produce kriged estimates of $PM_{2.5}$ concentrations at every point on our analysis grid domain (Figure 2.2) in the state of Washington. We used 24-hour average $PM_{2.5}$ concentrations and kriged between the values in latitude and longitude for each day. On each day, we used all available measurements; if data were missing at a measurement site on a given day, that monitor was omitted from the kriging analysis for that day. We evaluated the performance of our models at the surface sites that were located inside the model domain (red dots in figure 2.3). The evaluation procedure is described in Section 2.4.

2.2.3. Satellite Observations: MODIS AOD. In this work, we use the MODerate resolution Imaging Spectroradiometer (MODIS) instrument aboard the Terra and Aqua polar-orbiting satellites. Both satellites provide measurements of the entire globe approximately once daily, with the Terra and Aqua overpasses occurring at 10:30 AM and 1:30 PM local time respectively, with a swath width of 2330 km. MODIS instruments consist of radiometer measurements of the Earth at 36 wavelengths.

MODIS Aerosol Optical Depth (AOD) is a measure of optical extinction in the atmosphere due to the total amount of particulate matter in an atmospheric column; in many locations, AOD is often highly correlated with surface $PM_{2.5}$ concentrations. However, there can be variability in the strength of the relationship which depends on many other variables (e.g. [51]).

We use the Level 2 AOD Dark-Target Collection 6 retrieval at 10 km spatial resolution for this analysis [52]. The Aqua and Terra overpasses are averaged together to produce a single AOD observation that is representative of the daytime period. The Level 2 retrieval includes a cloud mask. However, in the case of wildfires, thick smoke often is mistaken for cloud by the retrieval and erroneously masked [26]. Changes were made in Collection 6 to account for this; however some pixels are still flagged as missing when we believe that they represent dense smoke instead of cloud. When masked pixels are surrounded by elevated AOD values due to fire smoke, we assumed that the retrieval failed on these pixels, and we interpolated surrounding pixel values to estimate the AOD in that pixel. If a missing pixel has three adjacent pixels that are elevated above background, then the pixel is given the average of all adjacent non-missing pixels. This process is repeated until gaps that appear to be in the smoke plume are filled. This procedure accounts for approximately 50% of the missing values from the MODIS AOD dataset, making it nearly continuous in regions of high-AOD where we believe dense smoke is located. However, pixels missing due to the presence of real clouds, or other filtering by the algorithm (i.e. not surrounded by high AOD measurements) are left in the dataset because they usually do not have three neighbors that are elevated.

A limitation to our procedure is that it likely underestimates the AOD in regions that were erroneously masked, as surrounding pixels are used to fill these gaps in; if the surrounding pixels had AOD values as high as the missing pixel, they would likely be masked as well. Because AOD is not a direct measure of aerosol mass, we apply this procedure to increase the spatial continuity of the AOD data, though we acknowledge that there are limitations to the AOD data in dense smoke.

Additionally, because we are using composite images from 10:30am and 1:30pm (from Terra and Aqua) as representative of the entire 24-hour time period, we are penalizing MODIS AOD's skill as a predictor of surface $PM_{2.5}$. Given the temporal resolution of our other data sources, this was a necessary step, but we acknowledge that it may reduce the predictive skill of these observations.

Finally, the MODIS AOD values were regridded to the WRF-Chem grid in the Analysis domain (Figure 2.2) using the nearest value.

2.3. Regression Models

2.3.1. Global Regression. We use linear regression for developing a predictive model from the three different exposure datasets: in-situ measurements, satellite-based measurements, and the WRF-Chem model. Linear regression posits that the dependent variable, in our case surface $PM_{2.5}$ concentrations, can be determined by linearly combining N different predictor variables as in Equation 3:

$$\vec{y} = \sum_{i=0}^{N} w_i x_i \tag{3}$$

where the coefficients w_i are chosen to optimize the fit. This is done by minimizing a cost function; the choice of cost function determines the type of regression being implemented.

For our problem, we have three different mostly spatially continuous datasets: WRF-Chem model output, interpolated in-situ measurements based on kriging, and MODIS AOD data.

Thus, the specific form of our linear fit is:

$$PM_{2.5,fit} = A + B * PM_{2.5,krige} + C * PM_{2.5,wrf} + D * AOD \tag{4}$$

where A is the fitted intercept and B, C, and D are the fitted slopes for each predictor variable. We use measured values of $PM_{2.5}$ as our dependent variable to train the regression model at points where measurements were made. We then use the derived coefficients to predict surface concentrations in every grid cell in our domain.

If we reformulated Equation 3 as matrix problem, we arrive at Equation 5

$$\mathbf{X}\vec{w} = \vec{y} \tag{5}$$

A common form of linear regression is "least squares" regression, where the cost function to be minimized is the sum of squared residuals between measured data and values predicted by the linear model (Equation 6).

$$||\mathbf{X}\vec{w} - \vec{y}||^2 \tag{6}$$

However, when different predictor variables are correlated with each other, least squares regression can produce coefficients that are sensitive to small errors. In our case, WRF-Chem $PM_{2.5}$, kriged $PM_{2.5}$, and MODIS AOD can all be indicators of smoke, meaning that the three datasets are likely to be correlated. Therefore, least-squares regression is a poor choice for our system.

An alternative to least squares regression is Tikhonov Regularization, or Ridge Regression. In Ridge Regression, the following expression is used instead:

$$||\mathbf{X}\vec{w} - \vec{y}||^2 + ||\mathbf{\Gamma}\vec{w}|| \tag{7}$$

Here, $\mathbf{\Gamma}$ is the Tikhonov Matrix, which is defined as $\mathbf{\Gamma} = \alpha\mathbf{I}$, where $\mathbf{I}$ is the identify matrix, and α is a scalar. The result of adding this extra term to the cost function is that it penalizes the regression for allowing any one coefficient to grow too large, which prevents the instabilities that arise in Least Square Regression.

We use Leave-One-Out Cross-Validation (Section 2.4) to train our regression coefficients. An alpha (α) parameter value of 0.2 was chosen because it produced the good performance in our evaluation statistics, though the regression model was not sensitive to this parameter.

2.3.2. Geographically Weighted Regression. In the regression analysis described above, one assumption that is implicitly made is that the surface $PM_{2.5}$ is related to the predictor variables in the same way at every point in space. However, we expect different exposure estimates to perform differently in our domain: for example, we expect WRF-Chem to perform differently in regions where emissions are dominated by urban/anthropogenic sources as opposed to biogenic and wildfire emissions, based on the accuracy of the emissions inventories and the representation of relevant processes such as chemistry and deposition. Likewise, we expect an interpolated estimate (e.g. kriging) to perform better in regions with a higher spatial density of surface sites or homogeneous emission sources.

In Geographically Weighted Regression (GWR), we can relax this assumption [53]. GWR is a geostatistical technique that has been used in urban air-quality studies (e.g. [54]) as

well as national-level studies incorporating remote sensing (e.g. [55], [36]), demonstrating its applicability at a wide range of spatial scales. In GWR, we calculate a separate set of regression coefficients at each point in space where a validating dataset can be obtained (in our case, locations of surface monitors where measurements were made) [56]. The coefficients are then interpolated in space (between the surface-monitor locations) to produce unique regression coefficients in every grid cell in our domain, that are tuned to the nearest surface monitors.

In our study, we use ridge regression ($\alpha = 0.2$) at each surface monitor location, using WRF-Chem surface $PM_{2.5}$, kriged surface $PM_{2.5}$, and MODIS AOD in the grid box as inputs, and the in-situ measurements to train a set of unique coefficients for a regression model for each in-situ monitor location. The coefficients are then interpolated using a Gaussian kernel [56] as in Equation 8:

$$G = \exp\left(-\left(\frac{D}{B_w}\right)^2\right) \tag{8}$$

where G is the kernel value of a given surface monitor, D is the distance to that monitor, and B_w is a bandwidth parameter; we use $B_w = 100\,km$ in our study. Using this interpolation procedure, a set of "interpolated" coefficients are calculated for each grid box in the analysis domain (Figure 2.2)

2.4. Model Evaluation Approach

All of our prediction techniques are compared to 24-hour measured $PM_{2.5}$ concentrations from the red surface sites in Figure 2.3. In this section, we describe how we specifically evaluate each of the prediction techniques.

In order to evaluate our WRF-Chem simulations and the MODIS AOD retrievals, we compare the model-predicted surface $PM_{2.5}$ concentrations with $PM_{2.5}$ concentrations measured at the surface monitors. We calculate the slope and R^2 for the aggregated dataset and each surface monitor for both WRF-Chem and MODIS. We also estimate mean absolute error (MAE), and mean bias (MB) for WRF-Chem (we cannot calculate MAE or MB for an AOD-$PM_{2.5}$ comparison).

To evaluate the kriged $PM_{2.5}$ estimates, we use the Leave-One-Out Cross-Validation (LOOCV) technique [57]; a single in-situ monitor is removed from the interpolation scheme, and we perform a kriging interpolation of the $PM_{2.5}$ concentrations measured at the remaining monitors. We compare the result of the kriging interpolation to the measured value by the in-situ monitor value that was removed from the interpolation. This process is repeated for every surface site in the domain. The same sets of statistics are calculated as for WRF-Chem both globally and per surface monitor.

We also evaluate the regression blends of kriging, WRF-Chem, and MODIS using a variation of the LOOCV technique. However, while we remove a single surface monitor for each LOOCV permutation, as above, for this evaluation we do include the WRF-Chem $PM_{2.5}$ and the MODIS AOD at this location. This is to simulate the data available away from surface monitors. To note, we use the LOOCV technique only for evaluation regression blends. When producing actual estimates of $PM_{2.5}$ concentrations in Washington, we use all surface data available.

CHAPTER 3

Results and Discussion

3.1. Sample timeseries data

Figure 3.1 shows timeseries of in-situ measurements of surface $PM_{2.5}$ as well as the kriged $PM_{2.5}$ estimate, WRF-Chem simulated $PM_{2.5}$ and MODIS AOD for two in-situ monitor locations (Figure 3.1a,c) as well as the in-situ monitor locations (Figure 3.1b).

Figure 3.1a shows results corresponding to the location of the surface monitor in Mt. Vernon, WA. Mt. Vernon is located in the northwest part of the state, which based on the in-situ observations of $PM_{2.5}$, does not appear to have experienced major impacts from wildfire smoke at the surface in 2012 (Figure 2.1). Therefore, $PM_{2.5}$ concentrations at this monitor were likely dominated by non-wildfire sources such as biogenic and anthropogenic emissions from the urban corridor near the Puget Sound during our analysis period. Examining the timeseries, we see that there was little variability in the surface observations and concentrations were consistently below 15 μg m^{-3}. WRF-Chem overpredicts $PM_{2.5}$ throughout the time period, but also appears to predict increases in $PM_{2.5}$ concentrations during September, when there were large wildfires in central Washington, that are not corroborated by the measurements. There were increases in MODIS AOD as well; therefore, it is possible that smoke was elevated above the surface at this location, and the model incorrectly transported some of the smoke plume to the surface. There are other possible explanations for the MODIS AOD timeseries showing periods of high AOD that are not reflected in surface observations: aerosols taking up water and thus increasing AOD without increasing dry $PM_{2.5}$ mass, changes in the aerosol size distribution, and local gradients in aerosols within the 10 km MODIS aerosol retrieval.

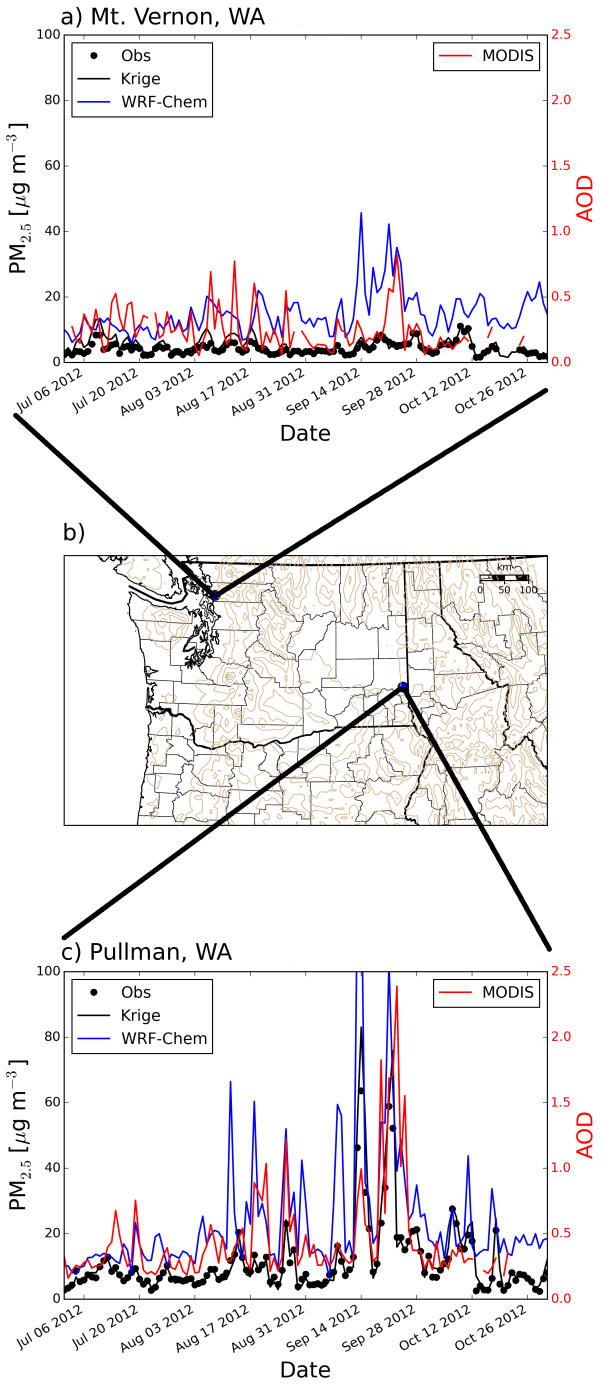

FIGURE 3.1. Time-series plots from (a) Mt. Vernon, WA with both regression models plotted alongside observations in the top panel, and the three raw dataseries in the bottom panel; (b) Map locations of Mt. Vernon and Pullman WA; (c) time-series plot from Pullman WA.

Figure 3.1c shows the same information for Pullman, WA which is in eastern Washington, downwind of the fires in central Washington (Figure 2.1). Pullman was impacted by smoke

in mid-August and in September, as suggested by the timeseries of both the observed $PM_{2.5}$ and the MODIS AOD. Like in Mt. Vernon, WRF-Chem appears to overestimate background surface $PM_{2.5}$; however, the positive bias during non-fire periods does not appear to be as high. WRF-Chem also appears to overestimate increases in $PM_{2.5}$ associated with wildfire smoke.

To assess the performance of these techniques throughout the domain, we compared estimated surface $PM_{2.5}$ concentrations to in-situ observations at every site throughout the domain (domain shown in Figure 2.2 and site locations shown in Figure 2.3). Figure 3.2 shows the estimate from each predictive model (including the GRR and GWR blend methods) plotted against surface observations, with corresponding slope, R^2, Mean Absolute Error (MAE) and Mean Bias (MB) consolidated in Table 3.1. Figure 3.3 shows the R^2 value of each technique at every surface site. These plots will be discussed further in the following subsections.

TABLE 3.1. Performance statistics for each exposure estimate.

Exposure Estimate	**Slope**	**R^2**	**Mean Absolute Error** [μg m^{-3}]	**Mean Bias** [μg m^{-3}]
WRF-Chem	0.67	0.25	11.45	10.22
Kriging	0.70	0.69	2.09	0.00
MODIS AOD	0.01	0.18	–	–
Global Ridge Regression	0.70	0.69	2.14	0.01
Geo-Weighted Ridge Regression	0.78	0.66	2.40	0.37

3.2. Evaluation of the three individual datasets

3.2.1. WRF-Chem Surface $PM_{2.5}$. Figure 3.2a shows the comparison of WRF-Chem simulated $PM_{2.5}$ (section 2.2.1) to surface observations. As shown in Table 3.1, WRF-Chem estimates have a slope of 0.67, R^2 of 0.25, and MAE of 11.4 and MB of 10.2 μg m^{-3}. These statistics suggest that our WRF-Chem simulation over-predicts concentrations throughout

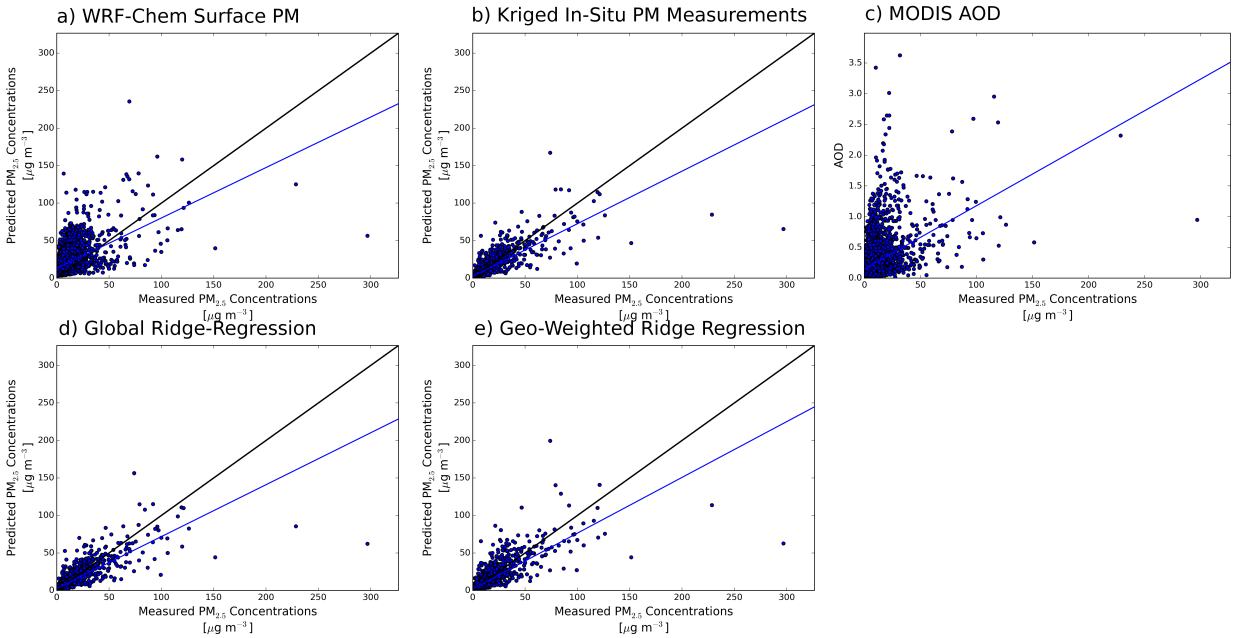

FIGURE 3.2. One-to-one plots of: (a) WRF-Chem, (b) Kriging, (c) MODIS AOD, (d) Global Ridge Regression, (e) Geo-Weighted Ridge Regression predictions of surface $PM_{2.5}$ vs. measured concentrations.

WA during this time period (June-September 2012). The slope of less than 1 is driven by (1) an underprediction of surface $PM_{2.5}$ for the few highest concentrations that were recorded and (2) a high $PM_{2.5}$ bias in non-smoke regions. Estimates from our WRF-Chem simulation are on average biased high, in large part due to the overprediction of $PM_{2.5}$ in non-fire regions.

Figure 3.3d shows the WRF-Chem $PM_{2.5}$ average bias at each in-situ monitor. We see that much of the positive bias is driven by the large number of monitors in western Washington that were not impacted by smoke. The R^2 values (Figure 3.3a) are close to zero at many of these monitors as well, showing that WRF-Chem is not capturing the variability at those surface monitors. Based on this, it appears that poor WRF-Chem performance for our simulation can be attributed, in part, to non-wildfire emissions, such as anthropogenic emissions, biogenic emissions, MOZART boundary conditions, or the choice

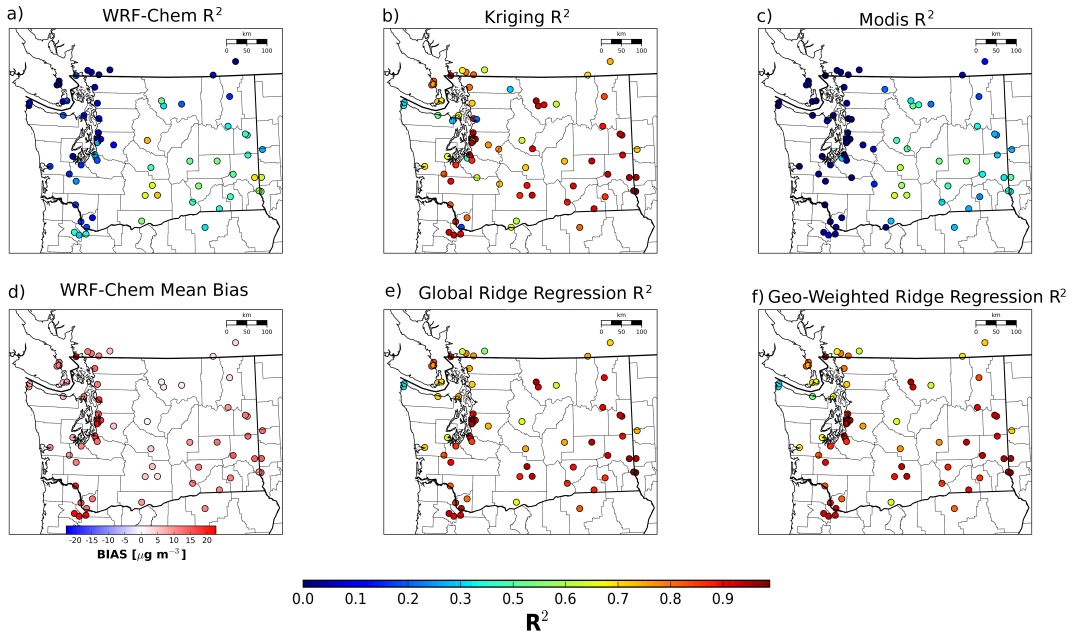

FIGURE 3.3. A-C, E, and F: R^2 values calculated at each surface monitor for: (a) WRF-Chem simulated surface $PM_{2.5}$ concentrations, (b) kriged in-situ observations of $PM_{2.5}$ concentrations, (c) MODIS AOD, (e) Global Ridge Regression estimated surface $PM_{2.5}$ concentrations, (f) Geo-Weighted Ridge Regression predicted surface $PM_{2.5}$. We excluded all monitors from evaluation for over 60% of days with data missing from either MODIS or observations, which is reflected in the relatively fewer monitors in the GRR and GWR figures. (d) contains the mean-bias associated with WRF-Chem estimates at each in-situ measurement site (and uses a different color scale).

of aerosol microphysics scheme. In contrast, the mean absolute bias is much closer to zero, and R^2 values are much higher in the central part of the state that was most-severely impacted by smoke and where mean measured concentrations were higher. In the eastern third of the state, WRF-Chem $PM_{2.5}$ estimates from our simulation are also biased high with respect to the surface observations, but R^2 values remain high. This is consistent with our observations in Pullman WA, where we observed in the timeseries (Figure 3.1c) that WRF-Chem overpredicts ambient and wildfire $PM_{2.5}$, but seems to capture the variability in surface observations. We conclude from this that, in spite of the high bias, our WRF-Chem simulations do offer useful information about wildfire-smoke $PM_{2.5}$ although it struggles to

represent $PM_{2.5}$ concentrations near the Puget Sound, which was not greatly affected by smoke.

To explore this bias due to non-smoke emissions further, an additional WRF-Chem simulation was performed with biomass burning emissions turned off (not shown). Even with no fires, WRF-Chem $PM_{2.5}$ estimates were biased high in the Puget Sound region, leading us to conclude that the overprediction in eastern WA was not due to simulation of the fire emissions, but to the representation of anthropogenic and/or biogenic emissions or chemical boundary conditions. It is well known that the choice of emissions inventory is a critical parameter in air-quality simulations [58]; however, other air-quality studies of the Puget Sound region (e.g. [59]) found good agreement between model-predicted $PM_{2.5}$ concentrations and in-situ measurements using the NEI in the CMAQ model, suggesting that our problems are not inherently due to the NEI. Other air-quality studies using WRF-Chem to predict surface $PM_{2.5}$ and ozone concentrations have found that choice of chemical mechanism (e.g. aqueous phase oxidation of SO_2) [60] and representation of atmosphere-ocean interactions [61] can also play a large role in surface $PM_{2.5}$ concentrations. Changing the model configuration (as discussed in Appendix A), can produce a wide range of $PM_{2.5}$ concentrations. Therefore, we could likely optimize our model simulation to better predict surface concentrations; however, it may be unlikely that the optimal model settings would be the same in western Washington along the coast, central Washington along the Cascade mountains, and in the plains of eastern Washington.

As previously mentioned, one of the main challenges for determining wildfire-specific health effects is overcoming exposure misclassification [31]. Other studies that have attempted to use WRF-Chem to characterize exposure to wildfire smoke (e.g. [24]) have confirmed that exposure misclassification can occur when using models to estimate exposure and the extent

of exposure misclassification is difficult to quantify. In particular, the Alman et al. [24] study had to contend with a comparatively smaller number of in-situ $PM_{2.5}$ monitors in their study domain against which to evaluate their exposure estimates. While it is possible in principle to use WRF-Chem to generate highly accurate surface $PM_{2.5}$ estimates, this may require optimizing the WRF-Chem simulation to better account for parameters and processes that are more representative of each region. This would likely also require enough in-situ monitors to comprehensively evaluate WRF-Chem performance in the domain. While this may be feasible in Washington due to the dense network of monitors, our goal is to produce a method that is transferrable to different regions and fires. Therefore, we also analyzed how routine observations could be used to estimate exposure and improve upon these model estimates.

3.2.2. Kriged Surface $PM_{2.5}$. The analogous one-to-one plot for our kriged $PM_{2.5}$ estimates (section 2.2.2) and R^2 map are shown in Figures 3.2b and Figure 3.3b, respectively. Using kriging to estimate $PM_{2.5}$ appears to outperform our WRF-Chem simulation of $PM_{2.5}$ throughout the domain with a slope of 0.7, R^2 of 0.69, MAE of 2.1 g/m3, and MB of 0.0. The slope less than one is driven by, as with our WRF-Chem simulation, an underprediction of the highest observed concentrations. This is expected, because interpolation techniques cannot predict local maxima or minima that do not impact a measurement site used in the interpolation. On average, however, kriging is able to produce estimates that are close to the actual value, given the low MAE and bias (calculated using our evaluation technique from section 2.4).

In Figure 3.3b, we note that kriging generally has the highest R^2 values in regions where surface monitors are closest together. This is an expected result; the spatial autocorrelation of $PM_{2.5}$ decays with distance, so sites in closer proximity are better able to capture the

expected variability than more distant ones. However, there are individual sites that appear to have low R^2 values despite close proximity to other monitors. This low correlation can occur if a monitor is near a local source of $PM_{2.5}$, especially if these surface monitors have a low $PM_{2.5}$ variability range. Because the kriging model produces estimates with a lower error and bias, we consider this to be a more-accurate estimate than our WRF-Chem simulations for the 2012 Washington fires.

While interpolation has been shown to be useful for characterizing $PM_{2.5}$ concentrations for ambient air pollution (e.g. [47] [62]), relatively few studies have attempted to use interpolation to characterize smoke concentrations for wildfires, largely because the highly concentrated plumes of smoke often produce spatial $PM_{2.5}$ concentration gradients that are too steep for in-situ monitor networks to resolve [63]. Studies that do use in-situ monitors to determine wildfire impacts on air quality either limit their study area to close to a monitor [63], or combine $PM_{2.5}$ monitor data with other sources of information (e.g. [28]). For the 2012 Washington fires, we are able to take advantage of Washington's comprehensive in-situ monitor network that was augmented by the deployment of extra monitors. Furthermore, the number and distribution of large wildfires (Figure 2.1) produced large areas of smoke impact instead of isolated plumes, which possibly explains why our interpolation is able to represent surface concentrations with better-than-expected accuracy. However, we are still concerned about the presence of local maxima or minima in $PM_{2.5}$ concentrations in between surface monitors, and therefore decided to incorporate other source of information, such as the satellite observations and WRF-Chem output into our exposure estimates.

3.2.3. MODIS AOD. The one-to-one plot for MODIS AOD (section 2.2.3) and in-situ $PM_{2.5}$ concentrations is shown in Figure 3.2c. Because MODIS AOD is not the same dimension as $PM_{2.5}$, the one-to-one plot does not lend itself to the same interpretation

as for WRF-Chem and Kriging. However, we can still use the figure to assess how well AOD might predict surface $PM_{2.5}$. First, we notice that there is a slope of 0.01 (Table 3.1); physically, this means that an increase in 0.01 AOD on average produces an increase of 1 μg m^{-3} in $PM_{2.5}$ concentrations. We can see, however, that the observations tend to be bifurcated around the line (either above or below), with few data points lying along the line. This variability in the AOD:$PM_{2.5}$ ratio may be due to variability in (1) the vertical profile of aerosol, e.g. due to lofted smoke, (2) the ambient relative humidity driving variability in aerosol water, and (3) the aerosol size distribution or composition driving variability in the mass extinction efficiency, and/or (4) temporal/spatial variability in the 24-hr $PM_{2.5}$ not captured by the resolution of the satellite observation. The R^2 value for MODIS AOD is 0.18. In Figure 3.3c, we see that, like with WRF-Chem $PM_{2.5}$, R^2 values for AOD are close to zero in western Washington, but at some surface monitors in the central and eastern portion of the state that were more-heavily smoke impacted, R^2 values are as high as 0.6.

Our use of MODIS AOD to estimate surface $PM_{2.5}$ concentrations produces an expected level of performance: MODIS AOD has been successfully used to identify wildfire smoke in previous studies (e.g. [23] [26]). With the exception of thick smoke occasionally being erroneously classified as a cloud [26], AOD is well suited for identifying large changes in atmospheric aerosol loading, which explains its good performance in central and eastern Washington. Contrastingly, low variability in aerosol loading in western Washington, coupled with the presence of fog and stratus clouds, as well as uncertainties in the aerosol retrieval in coastal regions, produces lower performance. However, because this region was not impacted by smoke, we may not need to rely on MODIS to estimate population exposure.

Even in smoke-impacted regions, we do not expect MODIS to explain all of the variance in the surface $PM_{2.5}$ concentrations, as there are still challenges associated with distinguishing

surface $PM_{2.5}$ from aerosol mass that is above the surface and not directly relevant to human health. Other studies have attempted to deal with this uncertainty in different ways; the Rappold et al. ([23]) study was conducted at a coarse spatial resolution and used a binary exposure variable (e.g. smoke or no-smoke based on fractional coverage of counties). At this level of spatial resolution, exposure misclassification is likely to occur from many sources, and the contribution due to uncertainty in smoke plume height may be relatively minor in comparison. The van Donkelaar et al. ([26]) study combines satellite AOD with output from a chemical transport model to try to further constrain population-level exposure estimates. Therefore, in the next part of our study, we investigate combining MODIS AOD with WRF-Chem output and kriged in-situ measurements of $PM_{2.5}$ to produce spatially continuous estimates of $PM_{2.5}$.

3.3. Regression Blends

3.3.1. Global Ridge Regression. For our first combination of model and observations, we aggregated our three datasets (e.g. kriged $PM_{2.5}$ measurements, WRF-Chem $PM_{2.5}$ estimates, and MODIS AOD) from all in-situ monitor locations into a ridge regression blend (GRR, section 2.3.1). As shown in Table 3.1, the GRR blend produces summary statistics (slope = 0.70, R^2=0.69, MAE = 2.14, MB=0.01) that are almost identical to the kriging estimate. If we examine the coefficients in our regression model, as shown in Table 3.2, we see that the GRR model is weighted very heavily towards the kriging model. If we use the mean AOD:$PM_{2.5}$ ratio from Figure 3.2c of 0.01 AOD/[μg m^{-3}] to convert the GRR coefficient for AOD into the same units as the other GRR coefficients, we find that the relative weights are 0.91 for kriging, 0.07 for WRF-Chem, and 0.01 for MODIS AOD. The result of these coefficients is that the GRR model-predicted $PM_{2.5}$ is very close to the

kriging-predicted values, and the performance statistics are therefore very similar as well. The heavy reliance on the kriged $PM_{2.5}$ estimates in the GRR is due to the kriging estimates outperforming both our WRF-Chem simulation and MODIS AOD at most locations. The R^2 values across the different measurement locations for the GRR are also quite similar to the kriged $PM_{2.5}$ estimate (Figure 3.2d,b). The only discernable difference is the absence of a few surface monitor locations in the GRR figure (Figure 3.3d), which is due to a lack of sufficient MODIS AOD observations at those locations.

It is important to note that these three inputs to the regression model are expected to be highly correlated. While we address instabilities in the regression model by using Ridge Regression, one unavoidable complication is that the regression coefficients are not necessarily unique, due to multicollinearity. Information that is contained in multiple datasets is only added to the regression model once. Therefore, a set of coefficients could exist that relies more on either our WRF-Chem $PM_{2.5}$ or MODIS AOD to produce the same estimate of surface $PM_{2.5}$.

TABLE 3.2. Regression coefficients for GRR fit. MODIS AOD and intercept coefficients are listed, but are in different units than kriging and WRF-Chem coefficients. The values in parentheses were converted to the same units by dividing the coefficient by the average measured surface $PM_{2.5}$ concentration, and for MODIS AOD, multiplying by the slope (Figure 3.2c)

Regression Coefficient	**Mean Value**	**Standard Deviation**
Intercept [μg m^{-3}] ([μg m^{-3}/μg m^{-3}])	-0.78 (-0.00)	0.04 (0.00)
Kriging [μg m^{-3}/μg m^{-3}]	0.91	0.00
WRF-Chem [μg m^{-3}/μg m^{-3}]	0.07	0.00
MODIS AOD [μg m^{-3}/AOD] ([μg m^{-3}/μg m^{-3}])	0.87 (0.01)	0.04 (0.00)

3.3.2. Geographically Weighted Ridge Regression. For our second combination of model and observations, we used Geographically Weighted Regression (GWR, described in 2.3.2). Table 3.1 shows the performance of the GWR model (slope = 0.76, R^2=0.66, MAE = 2.40, MB=0.37). GWR is similar to GRR, but rather than using one set of coefficients for the whole domain, a separate set of regression coefficients is calculated at each surface monitor location. The coefficients are then interpolated using a Gaussian kernel with a bandwidth parameter (equation 8) of 100 km. The bandwidth parameter can be thought of as a length scale at which the regression coefficients calculated a surface monitor influence the GWR coefficients throughout the domain. We varied the bandwidth to test the sensitivity and found that increasing the bandwidth reduced the slope further from unity, but improved the R^2. We chose 100 km to maximize the performance of these two statistics. Although the resulting R^2 is lower than the R^2 determined from the kriging estimate and the Global Ridge Regression model and the MAE and MB are slightly higher, the slope is much closer to one. And to note, these increases in MAE and MB from 2.1 to 2.4 and 0 to 0.4 μg m^{-3} (relative to using the kriged $PM_{2.5}$) are small, relative to the concentrations seen during the wildfire smoke events (e.g. 80 μg m^{-3}).

Figure 3.3e shows the R^2 values at each site. For sites near other sites, the GWR appears to perform the same or better than GRR, while isolated sites suffer a slight decrease in performance. This may be due to the interpolation of regression coefficients, which adds another layer of estimation to the system. Figure 3.4 contains a plot of the intercept and three regression coefficients calculated at each in-situ monitor in the domain and rescaled to sum to one. In Figure 3.4b, we see that at most of the in-situ monitor locations, GWR relies heavily on the kriged $PM_{2.5}$ measurements to produce the best estimate of surface $PM_{2.5}$ concentrations. However, in Figure 3.4c and d, we see that there is a non-zero contribution

from both WRF-Chem and MODIS in the central and eastern parts of the state, where both datasets were shown to explain more of the variance in the system (Figure 3.3). Interestingly, WRF-Chem appears to have a negative coefficient at some of the smoke-impacted monitors in central Washington. These monitors are all located near the Puget Sound so we expect a close correlation with the urban sites near the Puget Sound; however, these monitors are separated from the urban areas by a mountain range and in reality, these surface monitors are likely to be decoupled from the anthropogenic signal, despite the close proximity. Likewise, because WRF-Chem overpredicts anthropogenic emissions, it is possible that WRF-Chem predictions in this region also contain an anthropogenic signal, and the negative coefficient allows the model to remove this signal from the predicted surface concentrations at these locations.

GWR allows the model more degrees of freedom to account for strengths/weaknesses of the three inputs in different regions. which is an advantage over GRR. A disadvantage of GWR is that the total amount of data that is used to calibrate model coefficients at each individual site is lower than the amount of data used to calibrate the single set of coefficients in GRR, where the data from all individual sites are pooled together. Thus, it is possible that the fits at the individual sites for GWR are less robust than the global fit for GRR. Furthermore, while sites outside the analysis domain (external sites in Figure 2.3) were used as inputs to the kriging dataset, they were not incorporated into the GWR analysis, so analysis grid boxes near the boundaries of the domain have fewer regression coefficients contributing to the GWR calculation, and are also less reliable.

Our techniques perform comparably to other studies that have also tried to estimate smoke concentrations. The Yao et al. [29] study of the 2003-2012 fire seasons in British Columbia, which combined prior days $PM_{2.5}$ concentrations with the NOAA Hazard Mapping

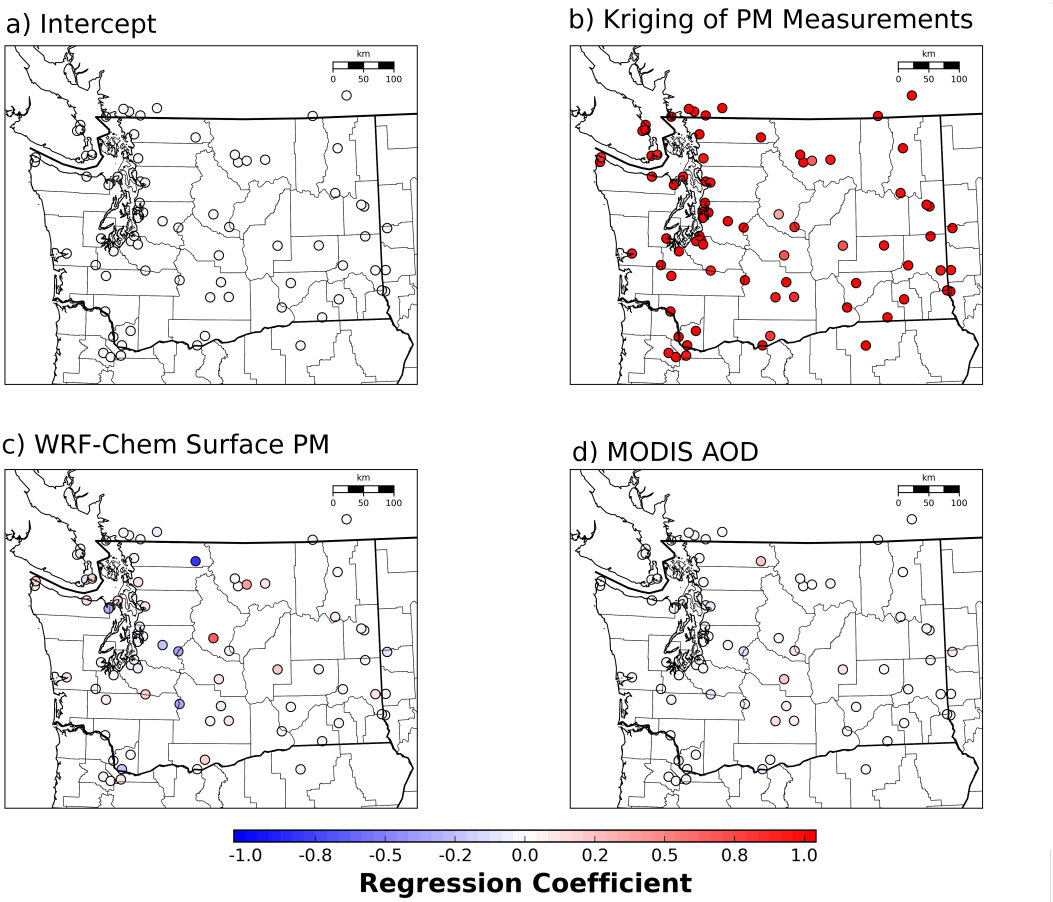

FIGURE 3.4. GWR Regression Coefficients (a) intercept, (b) kriged surface $PM_{2.5}$ concentrations, (c) WRF-Chem surface $PM_{2.5}$ concentrations, (d) MODIS AOD) calculated at each surface monitor, and normalized to sum to 1. The MODIS AOD coefficient and intercept are in different units (μg m^{-3}/AOD and μg m^{-3} respectively), and were changed to the same units as the kriging and WRF-Chem coefficients (as in Table 3.2) before the normalization was performed.

System (HMS) product, and MODIS-measured Fire Radiative Power (FRP), reported correlation coefficients of 0.84 (R^2 of 0.71), and a Normalized Root Mean Square Error (NRMSE) of 55.6% whereas our GWR model produces an R^2 of 0.66 and a NRMSE of 2%. Our study likely produced lower NRSME than the Yao et al. [29] study because we input same day $PM_{2.5}$ in-situ measurements from other sites, rather than yesterdays measurement at the same site. We expect that on average, $PM_{2.5}$ concentrations are highly autocorrelated in time, which explains why the Yao et al. study produced high R^2 values; however, during

wildfires that produce broad, regional smoke plumes, measurements may be decoupled from yesterdays measurement, which contributes to their relatively high NRMSE. Because we use a kriging interpolation to inform our regression model, we are able to capture these broad, regional smoke plumes which reduces our NRMSE, though not including yesterdays same-site measurement penalizes our surface-$PM_{2.5}$-estimate R^2 value. Finally, we expect that even if identical techniques were used in various studies, the performance of the techniques will vary for different fires and locations.

Another study, Reid et al. [28], combined many different types of land-use and GIS data with in-situ and satellite observations and CTM output in a multi-method, machine-learning framework. The Reid et al. [28] study found that a generalized boosting model (GBM) with 29 different variables input produced their best estimates, with a slope of 0.99, R^2 of 0.80, and a root mean square error (RMSE) of 1.5 μg m^{-3}. In comparison to this study, we are able to achieve similar performance with our error statistic (MAE of 2.4 μg m^{-3} for GWR), but our slope and R^2 do not achieve the same level of performance. In large part, this is because the GBM uses more input variables (29 vs. our 3), allowing for more degrees of freedom to train their model. Additionally, they used a suite of 11 sophisticated statistical models and selected the top performer, whereas we use only one regression model. An advantage of our techniques is that they allow us to investigate the spatial dependence of different input datasets, which may be more difficult with esoteric machine-learning techniques. Machine-learning approaches as applied by Reid et al. [28], use a wide range of environmental variables as inputs to their model such as wind direction, planetary boundary layer height, and distance to nearest fire, as well as land-use and distance to highways. While we expect some of these to be related to surface $PM_{2.5}$ surface concentrations, many of these variables (boundary layer height, U wind and V wind) are not necessarily independent, and others

(distance to highway) do not have a clear physical explanation for their ability to predict wildfire smoke concentrations, which limits our ability to gain a physical understanding of the system from these types of models. However, they do produce highly accurate results, which in the context of a health study is the primary objective. As in our study, they found that WRF-Chem simulation output was relatively uncorrelated with their best predictor variable (GASP AOD).

AOD as a metric to assess surface $PM_{2.5}$ is inherently limited by the challenge of distinguishing between surface and lofted smoke. Likewise in many situations, in-situ monitors are not numerous enough and fire plumes frequently are too narrow for an interpolation approach to add much value to a blended model. Therefore, both Reid et al. [28] and our approach, as well as the Alman et al. [24] approach could benefit from improved WRF-Chem simulations of wildfire smoke.

3.4. Estimate Technique Performance Simulated with Fewer In-Situ Measurements

In the case of the 2012 Washington Fires, kriging the surface-based $PM_{2.5}$ measurements produces estimates that generally far outperform WRF-Chem surface $PM_{2.5}$ and MODIS AOD datasets as concentration predictors throughout time and space. Therefore, the improvement from blending WRF-Chem $PM_{2.5}$ and MODIS AOD with the kriged $PM_{2.5}$ estimate is minor. However, not every region and time period has as comprehensive of a surface monitor network as Washington state during the 2012 fire period. In a more-sparsely monitored region, satellite, and in situ observations may add more value to the regression model than for the case discussed here. Therefore, we explored how the slope, R^2, and mean error of the kriging, GRR, and GWR methods vary as the number of surface $PM_{2.5}$ monitors in our domain varies

from 12 to 212 (which corresponds to using all measurements). For each number of monitors, we perform 100 trials where we randomly select the sets of surface monitors (e.g. for the 50-monitor case, we randomly select the 50 monitors and perform all subsequent analyses 100 different times). We include all monitors (evaluation and external monitors in Figure 2.3) as candidates for random selection as inputs for the models. For each selection of monitors, we perform a kriging fit of the surface-monitor data and compute a GRR and GWR. These 100 trials provide us with an estimate of how the different predictive models depend on the choice and number of monitors. We evaluate each of these predictions at all evaluation sites in Figure 2.3 as was done with the complete dataset (Section 2.4), regardless of the number of input sites to comprehensively assess performance throughout the domain of interest. Figure 3.5 shows the resulting R^2 value, slope, and MAE for the kriging, GRR, and GWR fits as a function of the number of surface sites in domain. In this analysis, we see that, on average, the GWR performs best with respect to slope for any number of input monitors. Furthermore, GWR achieves the highest R^2 when fewer than 50 monitors (corresponding to 25% of total available monitors) are input into the model. Kriging and GRR both achieve higher R^2 values than GWR when 120 monitors (i.e. 60% of total available monitors) have been input to the model. With respect to MAE, kriging performs best for any number of sites. However, all three models maintain an MAE between 2.0 and 4.0 μg m^{-3}, which is small relative to the concentrations that are produced by wildfire smoke.

A limitation of this analysis is that we have not optimized any of these estimation strategies for the new number of in-situ monitors. The kriging weights are calculated based on the semivariogram (Equation 2). The purpose of the semivariogram is to capture the dependence of physical semivariance (i.e. spatial autocorrelation) as a function of distance, and should theoretically be independent of the number of in-situ monitors used to inform the

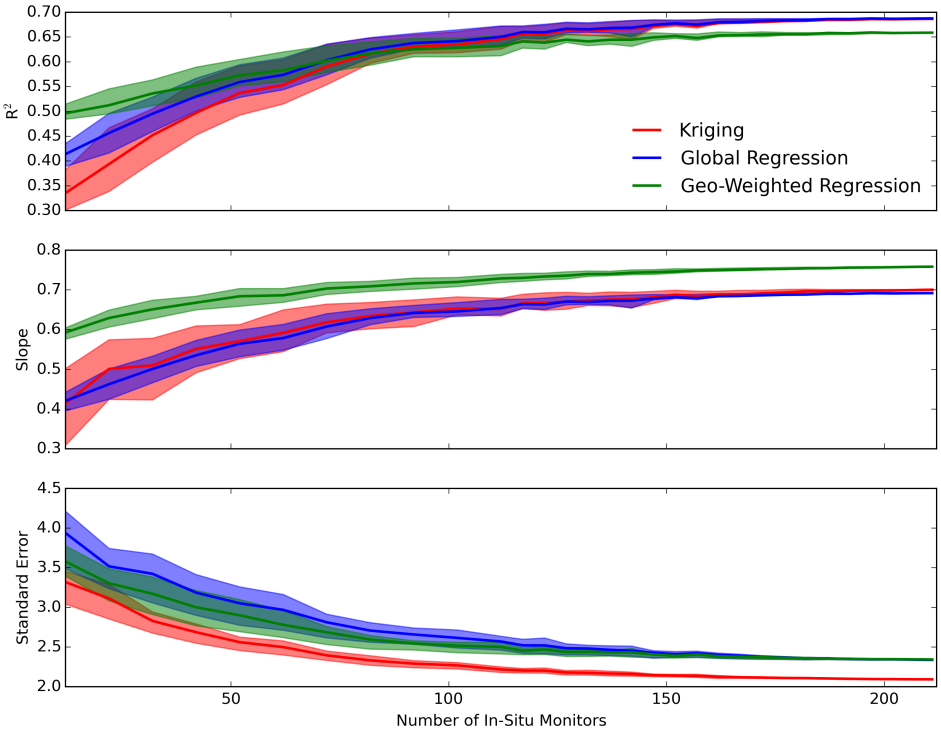

FIGURE 3.5. R^2, slope, and standard error (top to bottom) of Kriging, Global Ridge Regression, and Geographically Weighted Regression models as a function of number of sites. The shaded region represents between the 25th and 75th percentiles, with the line representing the mean from the random trials. All three models experience an increase in performance as sites are added; the Global Regression and Kriging models converge after 100 (or 50%) of the surface monitors are added to the calculation. GWR performs best with respect to slope and R^2 initially, but the other two models catch up as monitors are added.

model. Therefore, the performance of the kriging estimate is unlikely to change. Likewise, the only free parameter in the GRR estimate is the α parameter in the ridge regression matrix. We do not expect reducing the number of surface monitors to change the correlation between the three input datasets, and therefore the previous value should be sufficient.

In contrast, GWR has a bandwidth parameter, which determines the length-scale of regression coefficients influence throughout the domain. As fewer surface monitors are available, we expect the optimal bandwidth parameter to change. However, the bandwidth

depends not only on the number of in-situ monitors, but the configuration of those monitors. To fully explore this for an arbitrary choice of in-situ monitors requires a prohibitively expensive calculation; therefore, the results we present here can be thought of as a lower-bound for the GWR performance for few surface monitors.

3.5. "Smoke" vs. $PM_{2.5}$ Concentrations

The motivation of this thesis was to characterize wildfire-smoke-specific health effects by producing accurate estimates of smoke concentrations. However, our methods have focused on predicting total $PM_{2.5}$ concentrations, including contributions from non-wildfire sources. This is because we rely on in-situ observations to evaluate our different approaches. Furthermore, when we use observations (in-situ and satellite) in our analysis, they contain $PM_{2.5}$ from all sources, not only wildfires.

To account for this in our health analysis (not included in this thesis), we produce estimates of the non-wildfire-smoke contribution to $PM_{2.5}$ concentrations, and subtract this from our estimates of total $PM_{2.5}$. The details of how these estimates were obtained are described in Appendix C.

CHAPTER 4

Conclusions and Future Work

This thesis describes the development of new methods for determining population-level exposure to wildfire smoke by combining information from existing datasets that have previously been used for determining exposure, and discusses a strategy for evaluating these methods.

First, we use WRF-Chem to simulate smoke production and transport from wildfires during the 2012 Washington fire season, and use the model output to estimate surface $PM_{2.5}$ concentrations. We evaluate the simulations with in-situ data; the result of these model-observation comparisons suggest that WRF-Chem is able to capture variability in surface $PM_{2.5}$ measurements in central and eastern Washington which were impacted by smoke, but overestimates surface $PM_{2.5}$ concentrations in the Puget Sound urban area.

Next, we produce estimates of surface $PM_{2.5}$ concentrations by ordinary kriging of the in-situ monitors. We evaluate this estimate using LOOCV. According to our evaluation strategy, kriging of in-situ observations produces estimates with substantially more skill than our WRF-Chem simulation, a surprising result, given that wildfires usually produce thin, highly concentrated smoke plumes, that we did not expect a network of in-situ monitors would fully resolve. For this particular fire season, the large number of fires produced broad plumes of smoke which the dense network of monitors were able to resolve.

We then examined the utility of MODIS AOD in estimating wildfire smoke concentrations. We found that MODIS AOD performance varied throughout the region; central and eastern Washington MODIS AOD performed as expected; with some post-processing to fill erroneously masked pixels due to thick smoke, AOD was correlated with surface $PM_{2.5}$ in smoke-impacted

regions. MODIS AOD performed comparatively much worse in the western part of the state due to low variability in surface $PM_{2.5}$, as well as stratus clouds and fog reducing sample size. However, because western Washington was not impacted by smoke during this time-period, MODIS AOD was not a necessary tool in this region.

In order to generalize our methods to many different regions and fires, we attempt to combine our interpolation estimates with our WRF-Chem simulations and MODIS AOD measurements in two regression frameworks. Global ridge regression (GRR) produces linear weights for WRF-Chem surface $PM_{2.5}$, kriged in-situ measurements of $PM_{2.5}$, and MODIS AOD, and combines them to estimate surface $PM_{2.5}$. Because our kriging estimates outperformed WRF-Chem and MODIS, the GRR defaulted to the kriging estimates and failed to improve performance. When we subset the data by surface monitor and interpolate coefficients between in-situ monitors (i.e. GWR), we see that a more geographically flexible regression framework changes estimates more than adding only WRF-Chem and MODIS to a global regression framework, and though it improves the estimate with respect to the slope, it produces minor decreases in R^2 and small increases in MAE.

Finally, we investigate the performance of the top three performers (kriging alone, GRR and GWR) with fewer surface monitors to calibrate the model. For situations with fewer in-situ monitors, GWR and GRR both outperform kriging, with GWR performing the best of the three. As more in-situ monitors are included, the three models converge to a similar solution, with GWR producing estimates with a better slope, and slightly worse R^2 value. Because most regions of the western US are not monitored as comprehensively as Washington, we expect that GWR will be a flexible, easily implemented technique that can take advantage of any available in-situ measurements, as well as MODIS AOD (or other satellite observations) and CTM simulations.

4.1. Next Steps and Future Tools/Datasets

In this analysis, we have relied on in-situ observations, chemical-transport modelling, and satellite observations. Each of these tools is continually being improved, which will likely result in improvements in our ability to predict smoke.

As we show in Appendix A, the performance of WRF-Chem for estimating surface $PM_{2.5}$ is sensitive to the many different possible model configurations. We believe that for our study, improvements to the skill of the WRF-Chem input to our regression model will allow the GRR and GWR to rely more heavily on the WRF-Chem input, and therefore improve predictive skill, especially in regions where kriging does not perform well (i.e. where there are fewer monitors). This may require a degree of optimization to determine the best simulation settings for a particular region and/or fire event.

Independent of the model configuration settings, as our understanding of how to simulate wildfire smoke improves, we expect our ability to represent the impact on air quality in a chemical-transport model to improve as well. The NOAA FIREX and the JFSP FASMEE field campaigns will take place in the next few years, which will use a combination of modelling, aircraft, and in-situ measurements to characterize wildfires and prescribed burns. One of the aims of these field campaigns is to develop better understanding of emissions and plume rise, which are both known sources of error in modelling wildfires with WRF-Chem [64]. The findings from these field campaigns may lead to advancements in our modelling capabilities, allowing us to make more-accurate predictions of surface $PM_{2.5}$ concentrations.

With respect to in-situ monitors, the development of low-cost $PM_{2.5}$ sensors is an active area of research. Within the foreseeable future, the spatial density of in-situ observations may increase through these low-cost sensors, which suggests that more regions may become

as extensively monitored as Washington, allowing for robust interpolation of surface $PM_{2.5}$ during smoke events.

In this analysis, we have focused on polar-orbiting MODIS instruments, which make twice-daily measurements with a high degree of spectral detail. Other studies have found good predictive skill using GASP AOD from aboard GOES satellites to determine areas that are impacted by smoke [28]. GOES satellites are located in geostationary orbit, and provide nearly continuous observations of the western US.. We did not test using this dataset in our regressions. However, we expect that for our particular study region and time period, GASP will experience many of the same challenges we experienced with MODIS (i.e. extensive cloud cover and low variability in surface $PM_{2.5}$ in western Washington, as well as erroneous cloud-masking of dense smoke). Because our analysis was done at daily resolution, we used the MODIS AOD retrieval instead. However, in order to extend our analysis to higher temporal resolution (section 4.2), a geostationary platform will be essential for tracking changes to smoke plumes within a single day.

Additionally, other techniques for filling in missing pixels in MODIS and GASP AOD data have been demonstrated to work, such as Radial Basis Functions [28] which allow missing pixels to have a local maximum. Given that these missing pixels are masked because the retrieval mistakes the thick smoke for a cloud, this may be an appropriate consideration.

In addition to the current GOES satellites, the next generation NOAA geostationary satellite for the western US (GOES-R) is expected to launch in November 2016. GOES-R will provide the spectral and spatial detail of MODIS on a geostationary platform, allowing fire observations over western North America throughout both day and night, and smoke observations throughout the day. The advancements in smoke observation that this platform

could provide may greatly enhance our ability to incorporate satellite observations into smoke prediction models, especially at sub-daily timescales.

Finally, as Reid et al. [28] showed, an ensemble of sophisticated statistical frameworks for combining data, as well more input variables, can lead to improved performance. A machine-learning approach to these sorts of problems allow for different types of model inputs, as well as different statistical approaches, which can ensure that the best performing tools and best performing datasets are used for a given fire.

4.2. Estimation of other Exposure Metrics

Throughout our analysis, we have focused on predicting daily average $PM_{2.5}$ concentrations. However, daily average concentrations may not be the best predictor of health outcomes (e.g. mortality or hospitalizations). Stakeholders in this project have expressed interest in exploring the connection between health outcomes on other $PM_{2.5}$ metrics, such as daily maximum $PM_{2.5}$ concentrations, or the number of hours above a specified threshold $PM_{2.5}$ concentration.

While it is trivial to produce these estimates from WRF-Chem simulation output, interpolating metrics such as these is not straightforward. One approach could be to perform our analysis at hourly temporal resolution instead of daily, and then calculate these associated metrics in each grid box. This will limit the number of in-situ observations we can interpolate because not every surface monitor reports hourly values, and would force us to modify our approach of incorporating MODIS-Aqua and MODIS-Terra imagery for the satellite inputs to the model given their limited number of overpasses per day. While we could incorporate GASP AOD products, which are available at sub-daily timescales, we would still have no satellite information available at night.

Another approach could be to use the diurnal information we have at in-situ monitors that do report hourly data to infer how the concentrations vary with respect to the 24-hour average concentrations and extrapolate to parts of the domain that are highly correlated with these surface monitors. This would require additional assumptions about how spatially correlated different parts of the domain are, but may allow us to use many of the methods and datasets developed in this study.

Additionally, in this analysis, we have focused on producing predictions of bulk $PM_{2.5}$ mass associated with wildfire smoke. However, the composition of the particles may influence the toxicity and the health response [65]. Wildfire smoke can contain organic carbon aerosols, black carbon soot aerosols, and mineral dust with metals. The aerosols can be freshly emitted or aged, which affects the oxidation state of many of the compounds in the particle phase. Additionally, other non-particle species such as ozone, formaldehyde, benzene, or other toxic volatile organic compounds may also be present in the smoke plume. These chemical species have known long-term and short-term health effects; however, we may be erroneously attributing these health effects to wildfire $PM_{2.5}$ instead. Furthermore, these gaseous species may interact with the human lung, producing nonlinearities in the health response to $PM_{2.5}$ [66]. Additional studies are required to determine how to estimate the concentrations of the many gas-phase species that may impact human health, and additional epidemiological techniques may be required to disentangle the potentially complicated interactions of these many pollutants.

4.3. Additional Wildfire-Smoke Cases

The analysis in this thesis was performed on a single set of fires within a single U.S. state (Washington). Our results suggest that the predictive utility of each of the three tools:

in-situ measurements, satellite measurements, and models, varies between fires and regions. Furthermore, the availability of in-situ measurements for use in an interpolation model varies depending on the region. Therefore, it is important to try to repeat this analysis on fires in different locations and years.

From a more-pragmatic standpoint, we might expect the health effects of smoke exposure to vary between populations. The endpoint of this research is use of these exposure-estimation techniques in a health study on exposure to wildfire smoke. More health studies of different wildfires need to be done, and for therefore, there is a need to determine exposure for multiple wildfires.

The Colorado 2012 wildfire season would make for an interesting comparison study because there were a sparser set of in-situ monitors in the smoke-affected area relative to the Washington case studied here. As we showed earlier, the information from the model and satellites are more important when in-situ measurements are sparse. Therefore, analyzing Colorado 2012 fires would be an opportunity to test these methods in a system where additional in-situ monitors are not available. However, we will be limited in our ability to evaluate our exposure estimates or quantify exposure misclassification due to a lack of in-situ monitor sites. Because the 2012 fire season was composed of isolated fires, the modeling system is more challenging.

Conversely, in 2015, the Colorado Front Range was impacted by long-range transport of smoke from Canada and the Pacific Northwest. Unlike 2012, the smoke plumes were much broader, with the entire front range experiencing similar concentrations, so this smoke event may be easier to model. While less challenging in terms of assessing exposure, this case is interesting from an epidemiology perspective because the smoke that impacted Colorado

was emitted thousands of kilometers away, and aged smoke make have different health effects than fresh smoke.

Another candidate fire is the Station Fire in Los Angeles, 2009. During the Station Fire, smoke impacted the downtown area, exposing a large population. A large exposed population may produce a health signal that is more robust against noise as compared to the smaller populations exposed during the 2012 Washington and Colorado events. Therefore, this fire is an opportunity to find a strong health signal across a large and diverse population. However, given the complex emissions, topography, and other meteorological phenomena in and around Los Angeles, this fire will present a major challenge for accurately assessing exposure.

Finally, the 2015 Washington fires are an interesting follow-up study because much of the population will be the same as during the 2012 event, allowing for comparisons across fire seasons. From an exposure assessment perspective, this is also the next logical step because the procedure will be the same, or very similar, to the methods outlined in this thesis.

Bibliography

[1] GBD 2013 Risk Factors Collaborators, M. H. Forouzanfar, L. Alexander, H. R. Anderson, V. F. Bachman, S. Biryukov, M. Brauer, R. Burnett, D. Casey, M. M. Coates, A. Cohen, K. Delwiche, K. Estep, J. J. Frostad, K. C. Astha, H. H. Kyu, M. Moradi-Lakeh, M. Ng, E. L. Slepak, B. A. Thomas, J. Wagner, G. M. Aasvang, C. Abbafati, A. Abbasoglu Ozgoren, F. Abd-Allah, S. F. Abera, V. Aboyans, B. Abraham, J. P. Abraham, I. Abubakar, N. M. E. Abu-Rmeileh, T. C. Aburto, T. Achoki, A. Adelekan, K. Adofo, A. K. Adou, J. C. Adsuar, A. Afshin, E. E. Agardh, M. J. Al Khabouri, F. H. Al Lami, S. S. Alam, D. Alasfoor, M. I. Albittar, M. A. Alegretti, A. V. Aleman, Z. A. Alemu, R. Alfonso-Cristancho, S. Alhabib, R. Ali, M. K. Ali, F. Alla, P. Allebeck, P. J. Allen, U. Alsharif, E. Alvarez, N. Alvis-Guzman, A. A. Amankwaa, A. T. Amare, E. A. Ameh, O. Ameli, H. Amini, W. Ammar, B. O. Anderson, C. A. T. Antonio, P. Anwari, S. Argeseanu Cunningham, J. Arnlv, V. S. A. Arsenijevic, A. Artaman, R. J. Asghar, R. Assadi, L. S. Atkins, C. Atkinson, M. A. Avila, B. Awuah, A. Badawi, M. C. Bahit, T. Bakfalouni, K. Balakrishnan, S. Balalla, R. K. Balu, A. Banerjee, R. M. Barber, S. L. Barker-Collo, S. Barquera, L. Barregard, L. H. Barrero, T. Barrientos-Gutierrez, A. C. Basto-Abreu, A. Basu, S. Basu, M. O. Basulaiman, C. Batis Ruvalcaba, J. Beardsley, N. Bedi, T. Bekele, M. L. Bell, C. Benjet, D. A. Bennett, H. Benzian, E. Bernab, T. J. Beyene, N. Bhala, A. Bhalla, Z. A. Bhutta, B. Bikbov, A. A. Bin Abdulhak, J. D. Blore, F. M. Blyth, M. A. Bohensky, B. Bora Baara, G. Borges, N. M. Bornstein, D. Bose, S. Boufous, R. R. Bourne, M. Brainin, A. Brazinova, N. J. Breitborde, H. Brenner, A. D. M. Briggs, D. M. Broday, P. M. Brooks, N. G. Bruce, T. S. Brugha, B. Brunekreef, R. Buchbinder, L. N. Bui, G. Bukhman, A. G. Bulloch, M. Burch, P. G. J. Burney, I. R.

Campos-Nonato, J. C. Campuzano, A. J. Cantoral, J. Caravanos, R. Crdenas, E. Cardis, D. O. Carpenter, V. Caso, C. A. Castaeda-Orjuela, R. E. Castro, F. Catal-Lpez, F. Cavalleri, A. avlin, V. K. Chadha, J.-C. Chang, F. J. Charlson, H. Chen, W. Chen, Z. Chen, P. P. Chiang, O. Chimed-Ochir, R. Chowdhury, C. A. Christophi, T.-W. Chuang, S. S. Chugh, M. Cirillo, T. K. D. Claen, V. Colistro, M. Colomar, S. M. Colquhoun, A. G. Contreras, C. Cooper, K. Cooperrider, L. T. Cooper, J. Coresh, K. J. Courville, M. H. Criqui, L. Cuevas-Nasu, J. Damsere-Derry, H. Danawi, L. Dandona, R. Dandona, P. I. Dargan, A. Davis, D. V. Davitoiu, A. Dayama, E. F. de Castro, V. De la Cruz-Gngora, D. De Leo, G. de Lima, L. Degenhardt, B. del Pozo-Cruz, R. P. Dellavalle, K. Deribe, S. Derrett, D. C. Des Jarlais, M. Dessalegn, G. A. deVeber, K. M. Devries, S. D. Dharmaratne, M. K. Dherani, D. Dicker, E. L. Ding, K. Dokova, E. R. Dorsey, T. R. Driscoll, L. Duan, A. M. Durrani, B. E. Ebel, R. G. Ellenbogen, Y. M. Elshrek, M. Endres, S. P. Ermakov, H. E. Erskine, B. Eshrati, A. Esteghamati, S. Fahimi, E. J. A. Faraon, F. Farzadfar, D. F. J. Fay, V. L. Feigin, A. B. Feigl, S.-M. Fereshtehnejad, A. J. Ferrari, C. P. Ferri, A. D. Flaxman, T. D. Fleming, N. Foigt, K. J. Foreman, U. F. Paleo, R. C. Franklin, B. Gabbe, L. Gaffikin, E. Gakidou, A. Gamkrelidze, F. G. Gankp, R. T. Gansevoort, F. A. Garca-Guerra, E. Gasana, J. M. Geleijnse, B. D. Gessner, P. Gething, K. B. Gibney, R. F. Gillum, I. A. M. Ginawi, M. Giroud, G. Giussani, S. Goenka, K. Goginashvili, H. Gomez Dantes, P. Gona, T. Gonzalez de Cosio, D. Gonzlez-Castell, C. C. Gotay, A. Goto, H. N. Gouda, R. L. Guerrant, H. C. Gugnani, F. Guillemin, D. Gunnell, R. Gupta, R. Gupta, R. A. Gutirrez, N. Hafezi-Nejad, H. Hagan, M. Hagstromer, Y. A. Halasa, R. R. Hamadeh, M. Hammami, G. J. Hankey, Y. Hao, H. L. Harb, T. N. Haregu, J. M. Haro, R. Havmoeller, S. I. Hay, M. T. Hedayati, I. B. Heredia-Pi, L. Hernandez, K. R.

Heuton, P. Heydarpour, M. Hijar, H. W. Hoek, H. J. Hoffman, J. C. Hornberger, H. D. Hosgood, D. G. Hoy, M. Hsairi, G. Hu, H. Hu, C. Huang, J. J. Huang, B. J. Hubbell, L. Huiart, A. Husseini, M. L. Iannarone, K. M. Iburg, B. T. Idrisov, N. Ikeda, K. Innos, M. Inoue, F. Islami, S. Ismayilova, K. H. Jacobsen, H. A. Jansen, D. L. Jarvis, S. K. Jassal, A. Jauregui, S. Jayaraman, P. Jeemon, P. N. Jensen, V. Jha, F. Jiang, G. Jiang, Y. Jiang, J. B. Jonas, K. Juel, H. Kan, S. S. Kany Roseline, N. E. Karam, A. Karch, C. K. Karema, G. Karthikeyan, A. Kaul, N. Kawakami, D. S. Kazi, A. H. Kemp, A. P. Kengne, A. Keren, Y. S. Khader, S. E. A. H. Khalifa, E. A. Khan, Y.-H. Khang, S. Khatibzadeh, I. Khonelidze, C. Kieling, D. Kim, S. Kim, Y. Kim, R. W. Kimokoti, Y. Kinfu, J. M. Kinge, B. M. Kissela, M. Kivipelto, L. D. Knibbs, A. K. Knudsen, Y. Kokubo, M. R. Kose, S. Kosen, A. Kraemer, M. Kravchenko, S. Krishnaswami, H. Kromhout, T. Ku, B. Kuate Defo, B. Kucuk Bicer, E. J. Kuipers, C. Kulkarni, V. S. Kulkarni, G. A. Kumar, G. F. Kwan, T. Lai, A. Lakshmana Balaji, R. Lalloo, T. Lallukka, H. Lam, Q. Lan, V. C. Lansingh, H. J. Larson, A. Larsson, D. O. Laryea, P. M. Lavados, A. E. Lawrynowicz, J. L. Leasher, J.-T. Lee, J. Leigh, R. Leung, M. Levi, Y. Li, Y. Li, J. Liang, X. Liang, S. S. Lim, M. P. Lindsay, S. E. Lipshultz, S. Liu, Y. Liu, B. K. Lloyd, G. Logroscino, S. J. London, N. Lopez, J. Lortet-Tieulent, P. A. Lotufo, R. Lozano, R. Lunevicius, J. Ma, S. Ma, V. M. P. Machado, M. F. MacIntyre, C. Magis-Rodriguez, A. A. Mahdi, M. Majdan, R. Malekzadeh, S. Mangalam, C. C. Mapoma, M. Marape, W. Marcenes, D. J. Margolis, C. Margono, G. B. Marks, R. V. Martin, M. B. Marzan, M. T. Mashal, F. Masiye, A. J. Mason-Jones, K. Matsushita, R. Matzopoulos, B. M. Mayosi, T. T. Mazorodze, A. C. McKay, M. McKee, A. McLain, P. A. Meaney, C. Medina, M. M. Mehndiratta, F. Mejia-Rodriguez, W. Mekonnen, Y. A. Melaku, M. Meltzer, Z. A. Memish, W. Mendoza, G. A. Mensah, A. Meretoja, F. A. Mhimbira, R. Micha, T. R.

Miller, E. J. Mills, A. Misganaw, S. Mishra, N. Mohamed Ibrahim, K. A. Mohammad, A. H. Mokdad, G. L. Mola, L. Monasta, J. C. Montaez Hernandez, M. Montico, A. R. Moore, L. Morawska, R. Mori, J. Moschandreas, W. N. Moturi, D. Mozaffarian, U. O. Mueller, M. Mukaigawara, E. C. Mullany, K. S. Murthy, M. Naghavi, Z. Nahas, A. Naheed, K. S. Naidoo, L. Naldi, D. Nand, V. Nangia, K. M. V. Narayan, D. Nash, B. Neal, C. Nejjari, S. P. Neupane, C. R. Newton, F. N. Ngalesoni, J. d. D. Ngirabega, G. Nguyen, N. T. Nguyen, M. J. Nieuwenhuijsen, M. I. Nisar, J. R. Nogueira, J. M. Nolla, S. Nolte, O. F. Norheim, R. E. Norman, B. Norrving, L. Nyakarahuka, I.-H. Oh, T. Ohkubo, B. O. Olusanya, S. B. Omer, J. N. Opio, R. Orozco, R. S. Pagcatipunan, A. W. Pain, J. D. Pandian, C. I. A. Panelo, C. Papachristou, E.-K. Park, C. D. Parry, A. J. Paternina Caicedo, S. B. Patten, V. K. Paul, B. I. Pavlin, N. Pearce, L. S. Pedraza, A. Pedroza, L. Pejin Stokic, A. Pekericli, D. M. Pereira, R. Perez-Padilla, F. Perez-Ruiz, N. Perico, S. A. L. Perry, A. Pervaiz, K. Pesudovs, C. B. Peterson, M. Petzold, M. R. Phillips, H. P. Phua, D. Plass, D. Poenaru, G. V. Polanczyk, S. Polinder, C. D. Pond, C. A. Pope, D. Pope, S. Popova, F. Pourmalek, J. Powles, D. Prabhakaran, N. M. Prasad, D. M. Qato, A. D. Quezada, D. A. A. Quistberg, L. Racap, A. Rafay, K. Rahimi, V. Rahimi-Movaghar, S. U. Rahman, M. Raju, I. Rakovac, S. M. Rana, M. Rao, H. Razavi, K. S. Reddy, A. H. Refaat, J. Rehm, G. Remuzzi, A. L. Ribeiro, P. M. Riccio, L. Richardson, A. Riederer, M. Robinson, A. Roca, A. Rodriguez, D. Rojas-Rueda, I. Romieu, L. Ronfani, R. Room, N. Roy, G. M. Ruhago, L. Rushton, N. Sabin, R. L. Sacco, S. Saha, R. Sahathevan, M. A. Sahraian, J. A. Salomon, D. Salvo, U. K. Sampson, J. R. Sanabria, L. M. Sanchez, T. G. Snchez-Pimienta, L. Sanchez-Riera, L. Sandar, I. S. Santos, A. Sapkota, M. Satpathy, J. E. Saunders, M. Sawhney, M. I. Saylan, P. Scarborough, J. C. Schmidt, I. J. C. Schneider, B. Schttker, D. C. Schwebel,

J. G. Scott, S. Seedat, S. G. Sepanlou, B. Serdar, E. E. Servan-Mori, G. Shaddick, S. Shahraz, T. S. Levy, S. Shangguan, J. She, S. Sheikhbahaei, K. Shibuya, H. H. Shin, Y. Shinohara, R. Shiri, K. Shishani, I. Shiue, I. D. Sigfusdottir, D. H. Silberberg, E. P. Simard, S. Sindi, A. Singh, G. M. Singh, J. A. Singh, V. Skirbekk, K. Sliwa, M. Soljak, S. Soneji, K. Sreide, S. Soshnikov, L. A. Sposato, C. T. Sreeramareddy, N. J. C. Stapelberg, V. Stathopoulou, N. Steckling, D. J. Stein, M. B. Stein, N. Stephens, H. Stckl, K. Straif, K. Stroumpoulis, L. Sturua, B. F. Sunguya, S. Swaminathan, M. Swaroop, B. L. Sykes, K. M. Tabb, K. Takahashi, R. T. Talongwa, N. Tandon, D. Tanne, M. Tanner, M. Tavakkoli, B. J. Te Ao, C. M. Teixeira, M. M. Tllez Rojo, A. S. Terkawi, J. L. Texcalac-Sangrador, S. V. Thackway, B. Thomson, A. L. Thorne-Lyman, A. G. Thrift, G. D. Thurston, T. Tillmann, M. Tobollik, M. Tonelli, F. Topouzis, J. A. Towbin, H. Toyoshima, J. Traebert, B. X. Tran, L. Trasande, M. Trillini, U. Trujillo, Z. T. Dimbuene, M. Tsilimbaris, E. M. Tuzcu, U. S. Uchendu, K. N. Ukwaja, S. B. Uzun, S. van de Vijver, R. Van Dingenen, C. H. van Gool, J. van Os, Y. Y. Varakin, T. J. Vasankari, A. M. N. Vasconcelos, M. S. Vavilala, L. J. Veerman, G. Velasquez-Melendez, N. Venketasubramanian, L. Vijayakumar, S. Villalpando, F. S. Violante, V. V. Vlassov, S. E. Vollset, G. R. Wagner, S. G. Waller, M. T. Wallin, X. Wan, H. Wang, J. Wang, L. Wang, W. Wang, Y. Wang, T. S. Warouw, C. H. Watts, S. Weichenthal, E. Weiderpass, R. G. Weintraub, A. Werdecker, K. R. Wessells, R. Westerman, H. A. Whiteford, J. D. Wilkinson, H. C. Williams, T. N. Williams, S. M. Woldeyohannes, C. D. A. Wolfe, J. Q. Wong, A. D. Woolf, J. L. Wright, B. Wurtz, G. Xu, L. L. Yan, G. Yang, Y. Yano, P. Ye, M. Yenesew, G. K. Yentr, P. Yip, N. Yonemoto, S.-J. Yoon, M. Z. Younis, Z. Younoussi, C. Yu, M. E. Zaki, Y. Zhao, Y. Zheng, M. Zhou, J. Zhu, S. Zhu, X. Zou, J. R. Zunt, A. D. Lopez, T. Vos, and C. J. Murray, “Global,

regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013," *Lancet (London, England)*, vol. 386, pp. 2287–2323, Dec. 2015.

[2] D. W. Dockery, C. A. Pope, X. Xu, J. D. Spengler, J. H. Ware, M. E. Fay, B. G. J. Ferris, and F. E. Speizer, "An Association between Air Pollution and Mortality in Six U.S. Cities," *New England Journal of Medicine*, vol. 329, pp. 1753–1759, Dec. 1993.

[3] K. Donaldson, P. H. Beswick, and P. S. Gilmour, "Free radical activity associated with the surface of particles: a unifying factor in determining biological activity?," *Toxicology Letters*, vol. 88, pp. 293–298, Nov. 1996.

[4] Y.-F. Xing, Y.-H. Xu, M.-H. Shi, and Y.-X. Lian, "The impact of PM2.5 on the human respiratory system," *Journal of Thoracic Disease*, vol. 8, pp. E69–E74, Jan. 2016.

[5] B. Z. Simkhovich, M. T. Kleinman, and R. A. Kloner, "Air Pollution and Cardiovascular Injury: Epidemiology, Toxicology, and Mechanisms," *Journal of the American College of Cardiology*, vol. 52, pp. 719–726, Aug. 2008.

[6] K. Tsigaridis, M. Krol, F. J. Dentener, Y. Balkanski, J. Lathire, S. Metzger, D. A. Hauglustaine, and M. Kanakidou, "Change in global aerosol composition since preindustrial times," *Atmos. Chem. Phys.*, vol. 6, pp. 5143–5162, Nov. 2006.

[7] J. Lelieveld, J. S. Evans, M. Fnais, D. Giannadaki, and A. Pozzer, "The contribution of outdoor air pollution sources to premature mortality on a global scale," *Nature*, vol. 525, pp. 367–371, Sept. 2015.

[8] X. Hu, L. A. Waller, A. Lyapustin, Y. Wang, and Y. Liu, "10-year spatial and temporal trends of PM2.5 concentrations in the southeastern US estimated using high-resolution satellite data," *Atmos. Chem. Phys.*, vol. 14, pp. 6301–6314, June 2014.

[9] F. Laden, J. Schwartz, F. E. Speizer, and D. W. Dockery, "Reduction in Fine Particulate Air Pollution and Mortality," *American Journal of Respiratory and Critical Care Medicine*, vol. 173, pp. 667–672, Mar. 2006.

[10] R. W. Mutch, "Wildland Fires and EcosystemsA Hypothesis," *Ecology*, vol. 51, pp. 1046–1051, Nov. 1970.

[11] V. Trouet, A. H. Taylor, E. R. Wahl, C. N. Skinner, and S. L. Stephens, "Fireclimate interactions in the American West since 1400 CE," *Geophysical Research Letters*, vol. 37, Feb. 2010.

[12] A. L. Westerling, H. G. Hidalgo, D. R. Cayan, and T. W. Swetnam, "Warming and Earlier Spring Increase Western U.S. Forest Wildfire Activity," *Science*, vol. 313, pp. 940–943, Aug. 2006.

[13] D. V. Spracklen, L. J. Mickley, J. A. Logan, R. C. Hudman, R. Yevich, M. D. Flannigan, and A. L. Westerling, "Impacts of climate change from 2000 to 2050 on wildfire activity and carbonaceous aerosol concentrations in the western United States," *Journal of Geophysical Research: Atmospheres*, vol. 114, p. D20301, Oct. 2009.

[14] X. Yue, L. J. Mickley, J. A. Logan, and J. O. Kaplan, "Ensemble projections of wildfire activity and carbonaceous aerosol concentrations over the western United States in the mid-21st century," *Atmospheric environment (Oxford, England : 1994)*, vol. 77, pp. 767–780, Oct. 2013.

[15] O. US EPA, "NAAQS Table."

[16] M. Val Martin, C. L. Heald, B. Ford, A. J. Prenni, and C. Wiedinmyer, "A decadal satellite analysis of the origins and impacts of smoke in Colorado," *Atmos. Chem. Phys.*, vol. 13, pp. 7429–7439, Aug. 2013.

[17] P. R. Colarco, M. R. Schoeberl, B. G. Doddridge, L. T. Marufu, O. Torres, and E. J. Welton, "Transport of smoke from Canadian forest fires to the surface near Washington, D.C.: Injection height, entrainment, and optical properties," *Journal of Geophysical Research: Atmospheres*, vol. 109, p. D06203, Mar. 2004.

[18] L. J. DeBell, R. W. Talbot, J. E. Dibb, J. W. Munger, E. V. Fischer, and S. E. Frolking, "A major regional air pollution event in the northeastern United States caused by extensive forest fires in Quebec, Canada," *Journal of Geophysical Research: Atmospheres*, vol. 109, p. D19305, Oct. 2004.

[19] N. Knzli, E. Avol, J. Wu, W. J. Gauderman, E. Rappaport, J. Millstein, J. Bennion, R. McConnell, F. D. Gilliland, K. Berhane, F. Lurmann, A. Winer, and J. M. Peters, "Health Effects of the 2003 Southern California Wildfires on Children," *American Journal of Respiratory and Critical Care Medicine*, vol. 174, pp. 1221–1228, Dec. 2006.

[20] C. T. Elliott, S. B. Henderson, and V. Wan, "Time series analysis of fine particulate matter and asthma reliever dispensations in populations affected by forest fires," *Environmental Health: A Global Access Science Source*, vol. 12, p. 11, 2013.

[21] M. Paglione, S. Saarikoski, S. Carbone, R. Hillamo, M. C. Facchini, E. Finessi, L. Giulianelli, C. Carbone, S. Fuzzi, F. Moretti, E. Tagliavini, E. Swietlicki, K. Eriksson Stenstrm, A. S. H. Prvt, P. Massoli, M. Canaragatna, D. Worsnop, and S. Decesari, "Primary and secondary biomass burning aerosols determined by proton nuclear magnetic resonance (1h-NMR) spectroscopy during the 2008 EUCAARI campaign in the Po Valley (Italy)," *Atmos. Chem. Phys.*, vol. 14, pp. 5089–5110, May 2014.

[22] J. J. K. Jaakkola, "Case-crossover design in air pollution epidemiology," *The European Respiratory Journal. Supplement*, vol. 40, pp. 81s–85s, May 2003.

[23] A. G. Rappold, S. L. Stone, W. E. Cascio, L. M. Neas, V. J. Kilaru, M. S. Carraway, J. J. Szykman, A. Ising, W. E. Cleve, J. T. Meredith, H. Vaughan-Batten, L. Deyneka, and R. B. Devlin, "Peat Bog Wildfire Smoke Exposure in Rural North Carolina Is Associated with Cardiopulmonary Emergency Department Visits Assessed through Syndromic Surveillance," *Environmental Health Perspectives*, vol. 119, pp. 1415–1420, Oct. 2011.

[24] B. L. Alman, G. Pfister, H. Hao, J. Stowell, X. Hu, Y. Liu, and M. J. Strickland, "The association of wildfire smoke with respiratory and cardiovascular emergency department visits in Colorado in 2012: a case crossover study," *Environmental Health*, vol. 15, p. 64, 2016.

[25] A. van Donkelaar, R. V. Martin, and R. J. Park, "Estimating ground-level PM2.5 using aerosol optical depth determined from satellite remote sensing," *Journal of Geophysical Research: Atmospheres*, vol. 111, p. D21201, Nov. 2006.

[26] A. van Donkelaar, R. V. Martin, R. C. Levy, A. M. da Silva, M. Krzyzanowski, N. E. Chubarova, E. Semutnikova, and A. J. Cohen, "Satellite-based estimates of ground-level fine particulate matter during extreme events: A case study of the Moscow fires in 2010," *Atmospheric Environment*, vol. 45, pp. 6225–6232, Nov. 2011.

[27] M. Brauer, M. Amann, R. T. Burnett, A. Cohen, F. Dentener, M. Ezzati, S. B. Henderson, M. Krzyzanowski, R. V. Martin, R. Van Dingenen, A. van Donkelaar, and G. D. Thurston, "Exposure Assessment for Estimation of the Global Burden of Disease Attributable to Outdoor Air Pollution," *Environmental Science & Technology*, vol. 46, pp. 652–660, Jan. 2012.

[28] C. E. Reid, M. Jerrett, M. L. Petersen, G. G. Pfister, P. E. Morefield, I. B. Tager, S. M. Raffuse, and J. R. Balmes, "Spatiotemporal Prediction of Fine Particulate Matter

During the 2008 Northern California Wildfires Using Machine Learning," *Environmental Science & Technology*, vol. 49, pp. 3887–3896, Mar. 2015.

[29] J. Yao and S. B. Henderson, "An empirical model to estimate daily forest fire smoke exposure over a large geographic area using air quality, meteorological, and remote sensing data," *Journal of Exposure Science & Environmental Epidemiology*, vol. 24, pp. 328–335, May 2014.

[30] S. B. Henderson, M. Brauer, Y. C. MacNab, and S. M. Kennedy, "Three Measures of Forest Fire Smoke Exposure and Their Associations with Respiratory and Cardiovascular Health Outcomes in a Population-Based Cohort," *Environmental Health Perspectives*, vol. 119, pp. 1266–1271, Sept. 2011.

[31] J. C. Liu, G. Pereira, S. A. Uhl, M. A. Bravo, and M. L. Bell, "A systematic review of the physical health impacts from non-occupational exposure to wildfire smoke," *Environmental Research*, vol. 136, pp. 120–132, Jan. 2015.

[32] G. A. Grell, S. E. Peckham, R. Schmitz, S. A. McKeen, G. Frost, W. C. Skamarock, and B. Eder, "Fully coupled online chemistry within the WRF model," *Atmospheric Environment*, vol. 39, pp. 6957–6975, Dec. 2005.

[33] G. Grell, S. R. Freitas, M. Stuefer, and J. Fast, "Inclusion of biomass burning in WRF-Chem: impact of wildfires on weather forecasts," *Atmos. Chem. Phys.*, vol. 11, pp. 5289–5303, June 2011.

[34] L. K. Emmons, S. Walters, P. G. Hess, J.-F. Lamarque, G. G. Pfister, D. Fillmore, C. Granier, A. Guenther, D. Kinnison, T. Laepple, J. Orlando, X. Tie, G. Tyndall, C. Wiedinmyer, S. L. Baughcum, and S. Kloster, "Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4)," *Geosci. Model Dev.*, vol. 3, pp. 43–67, Jan. 2010.

[35] M. Chin, R. B. Rood, S.-J. Lin, J.-F. Mller, and A. M. Thompson, "Atmospheric sulfur cycle simulated in the global model GOCART: Model description and global properties," *Journal of Geophysical Research: Atmospheres*, vol. 105, pp. 24671–24687, Oct. 2000.

[36] X. Hu, L. A. Waller, M. Z. Al-Hamdan, W. L. Crosson, M. G. Estes Jr., S. M. Estes, D. A. Quattrochi, J. A. Sarnat, and Y. Liu, "Estimating ground-level PM2.5 concentrations in the southeastern U.S. using geographically weighted regression," *Environmental Research*, vol. 121, pp. 1–10, Feb. 2013.

[37] G. Thompson, R. M. Rasmussen, and K. Manning, "Explicit Forecasts of Winter Precipitation Using an Improved Bulk Microphysics Scheme. Part I: Description and Sensitivity Analysis," *Monthly Weather Review*, vol. 132, pp. 519–542, Feb. 2004.

[38] OAR US EPA, "2011 NEI Report."

[39] A. B. Guenther, X. Jiang, C. L. Heald, T. Sakulyanontvittaya, T. Duhl, L. K. Emmons, and X. Wang, "The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions," *Geosci. Model Dev.*, vol. 5, pp. 1471–1492, Nov. 2012.

[40] C. Wiedinmyer, S. K. Akagi, R. J. Yokelson, L. K. Emmons, J. A. Al-Saadi, J. J. Orlando, and A. J. Soja, "The Fire INventory from NCAR (FINN): a high resolution global model to estimate the emissions from open burning," *Geoscientific Model Development*, vol. 4, pp. 625–641, July 2011.

[41] S. R. Freitas, K. M. Longo, R. Chatfield, D. Latham, M. A. F. Silva Dias, M. O. Andreae, E. Prins, J. C. Santos, R. Gielow, and J. A. Carvalho Jr., "Including the sub-grid scale plume rise of vegetation fires in low resolution atmospheric transport models," *Atmos. Chem. Phys.*, vol. 7, pp. 3385–3398, July 2007.

[42] C. E. Reid, M. Brauer, F. Johnston, M. Jerrett, J. R. Balmes, and C. T. Elliott, "Critical Review of Health Impacts of Wildfire Smoke Exposure," *Environmental Health Perspectives*, Apr. 2016.

[43] D. Deligiorgi and K. Philippopoulos, "Spatial Interpolation Methodologies in Urban Air Pollution Modeling: Application for the Greater Area of Metropolitan Athens, Greece," in *Advanced Air Pollution* (F. Nejadkoorki, ed.), InTech, Aug. 2011.

[44] L. Li, X. Zhou, M. Kalo, and R. Piltner, "Spatiotemporal Interpolation Methods for the Application of Estimating Population Exposure to Fine Particulate Matter in the Contiguous U.S. and a Real-Time Web Application," *International Journal of Environmental Research and Public Health*, vol. 13, no. 8, 2016.

[45] E. H. Isaaks and R. M. Srivastava, *Applied geostatistics*. Oxford University Press, 1989.

[46] S. Janssen, G. Dumont, F. Fierens, and C. Mensink, "Spatial interpolation of air pollution measurements using CORINE land cover data," *Atmospheric Environment*, vol. 42, pp. 4884–4903, June 2008.

[47] "Spatial Analysis of Air Pollution and Mortality in Los Angeles:," *Epidemiology*, vol. 16, no. 6.

[48] S. R. Kethireddy, P. B. Tchounwou, H. A. Ahmad, A. Yerramilli, and J. H. Young, "Geospatial Interpolation and Mapping of Tropospheric Ozone Pollution Using Geostatistics," *International Journal of Environmental Research and Public Health*, vol. 11, pp. 983–1000, Jan. 2014.

[49] D. Shepard, "A two-dimensional interpolation function for irregularly-spaced data," pp. 517–524, ACM Press, 1968.

[50] B. S. Murphy, "PyKrige: Development of a Kriging Toolkit for Python," *AGU Fall Meeting Abstracts*, Dec. 2014.

[51] B. Ford and C. L. Heald, "Exploring the uncertainty associated with satellite-based estimates of premature mortality due to exposure to fine particulate matter," *Atmos. Chem. Phys.*, vol. 16, pp. 3499–3523, Mar. 2016.

[52] A. M. Sayer, L. A. Munchak, N. C. Hsu, R. C. Levy, C. Bettenhausen, and M.-J. Jeong, "MODIS Collection 6 aerosol products: Comparison between Aqua's e-Deep Blue, Dark Target, and merged data sets, and usage recommendations," *Journal of Geophysical Research: Atmospheres*, vol. 119, p. 2014JD022453, Dec. 2014.

[53] C. Brunsdon, S. Fotheringham, and M. Charlton, "Geographically Weighted Regression," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 47, pp. 431–443, Sept. 1998.

[54] W. Song, H. Jia, J. Huang, and Y. Zhang, "A satellite-based geographically weighted regression model for regional PM2.5 estimation over the Pearl River Delta region in China," *Remote Sensing of Environment*, vol. 154, pp. 1–7, Nov. 2014.

[55] W. You, Z. Zang, L. Zhang, Y. Li, and W. Wang, "Estimating national-scale ground-level PM25 concentration in China using geographically weighted regression based on MODIS and MISR AOD," *Environmental Science and Pollution Research*, vol. 23, pp. 8327–8338, Jan. 2016.

[56] A. S. Fotheringham, C. Brundson, and M. Charlton, *Geographically weighted regression the analysis of spatially varying relationships*. Wiley, 2002.

[57] B. Efron, *The Jackknife, the Bootstrap and Other Resampling Plans*. Society for Industrial and Applied Mathematics, Jan. 1982.

[58] M. Zhong, E. Saikawa, Y. Liu, V. Naik, L. W. Horowitz, M. Takigawa, Y. Zhao, N.-H. Lin, and E. A. Stone, "Air quality modeling with WRF-Chem v3.5 in East Asia:

sensitivity to emissions and evaluation of simulated air quality," *Geoscientific Model Development*, vol. 9, pp. 1201–1218, Apr. 2016.

[59] A. P. Tsimpidi, M. Trail, Y. Hu, A. Nenes, and A. G. Russell, "Modeling an air pollution episode in northwestern United States: Identifying the effect of nitrogen oxide and volatile organic compound emission changes on air pollutants formation using direct sensitivity analysis," *Journal of the Air & Waste Management Association*, vol. 62, pp. 1150–1165, Oct. 2012.

[60] P. Tuccella, G. Curci, G. Visconti, B. Bessagnet, L. Menut, and R. J. Park, "Modeling of gas and aerosol with WRF/Chem over Europe: Evaluation and sensitivity study: WRF/CHEM OVER EUROPE," *Journal of Geophysical Research: Atmospheres*, vol. 117, pp. n/a–n/a, Feb. 2012.

[61] J. He, R. He, and Y. Zhang, "Impacts of airsea interactions on regional air quality predictions using WRF/Chem v3.6.1 coupled with ROMS v3.7: southeastern US example," *Geoscientific Model Development Discussions*, vol. 8, pp. 9965–10009, Nov. 2015.

[62] L. Li, J. Tian, X. Zhang, J. B. Holt, and R. Piltner, "Estimating Population Exposure to Fine Particulate Matter in the Conterminous U.S. using Shape Function-based Spatiotemporal Interpolation Method: A County Level Analysis," *GSTF international journal on computing*, vol. 1, pp. 24–30, Jan. 2012.

[63] J. Wu, A. M Winer, and R. J Delfino, "Exposure assessment of particulate matter air pollution before, during, and after the 2003 Southern California wildfires," *Atmospheric Environment*, vol. 40, pp. 3333–3348, June 2006.

[64] C. Walter, S. R. Freitas, C. Kottmeier, I. Kraut, D. Rieger, H. Vogel, and B. Vogel, "The importance of plume rise on the concentrations and atmospheric impacts of biomass

burning aerosol," *Atmospheric Chemistry and Physics*, vol. 16, pp. 9201–9219, July 2016.

[65] M. L. Bell and HEI Health Review Committee, "Assessment of the health impacts of particulate matter characteristics," *Research Report (Health Effects Institute)*, pp. 5–38, Jan. 2012.

[66] J. L. Mauderly and J. M. Samet, "Is There Evidence for Synergy Among Air Pollutants in Causing Health Effects?," *Environmental Health Perspectives*, vol. 117, pp. 1–6, Jan. 2009.

[67] G. L. Mellor and T. Yamada, "A Hierarchy of Turbulence Closure Models for Planetary Boundary Layers," *Journal of the Atmospheric Sciences*, vol. 31, pp. 1791–1806, Oct. 1974.

[68] H. Morrison and A. Gettelman, "A New Two-Moment Bulk Stratiform Cloud Microphysics Scheme in the Community Atmosphere Model, Version 3 (CAM3). Part I: Description and Numerical Tests," *Journal of Climate*, vol. 21, pp. 3642–3659, Aug. 2008.

APPENDIX A

WRF-Chem Sensitivity Simulations

As mentioned in Chapter 2, we performed a sensitivity study of the WRF-Chem model to determine a) which combinations of parameters produced the best model performance and b) how sensitive the simulation results are to different simulation settings. We evaluated model performance by comparing the different simulated 48-hour surface concentrations in downtown Fort Collins with measured values (Figure A.1).

The simulation results vary widely. In many simulations, $PM_{2.5}$ concentrations are never predicted to be above a typical background value; this is due to the model transporting the smoke plume to the wrong location, erroneously concluding that Fort Collins was never impacted by smoke. In simulations that do show elevated $PM_{2.5}$, the concentrations exceed measured concentrations by 2-2.5 times. One simulation does accurately predict the 48-hour concentration at this surface site. However, this simulation used the coarsest model resolution, and a biomass-burning emissions inventory that we believe overpredicts $PM_{2.5}$ emissions (section A.4); therefore, we suspect that we are not getting the correct concentration for the right reasons.

Northern Colorado has a sparse network of in-situ measurements compared to the Pacific Northwest. Additionally, the smoke plumes generated during High Park were relatively isolated and narrow, with steep concentration gradients. Ultimately, it was difficult to evaluate the different simulations to determine which simulation performed the best with respect to surface concentrations.

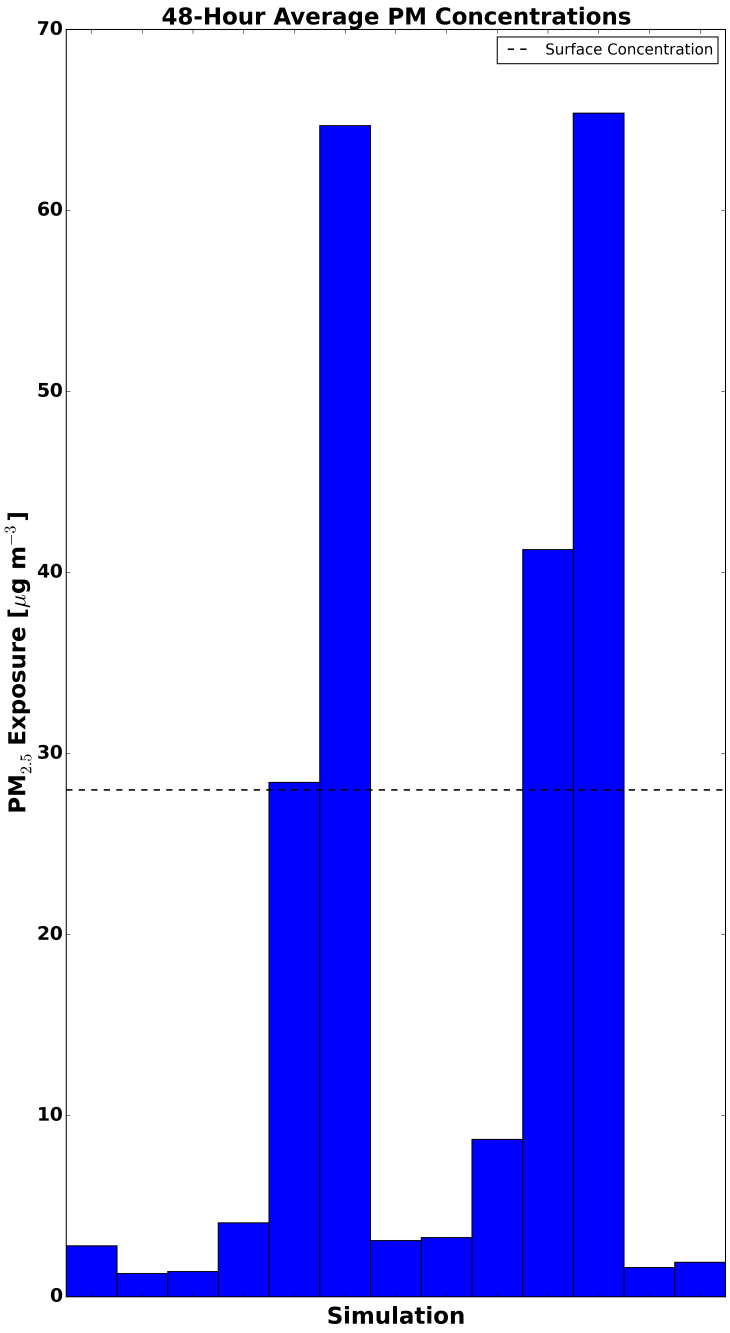

FIGURE A.1. WRF-Chem 48-hour average simulated $PM_{2.5}$ concentrations in downtown Fort Collins, CO during the High Park Fire. Different simulation parameter settings are plotted in the bars, with the dashed line representing measured concentrations.

A.1. METEOROLOGICAL INITIAL/BOUNDARY CONDITIONS

Figure A.2 shows the difference in surface $PM_{2.5}$ concentrations between WRF-Chem simulations run with GFS and NAM meteorology initial/boundary conditions, averaged over

a single day. We are only showing a single day because this is representative of what was seen on every day.

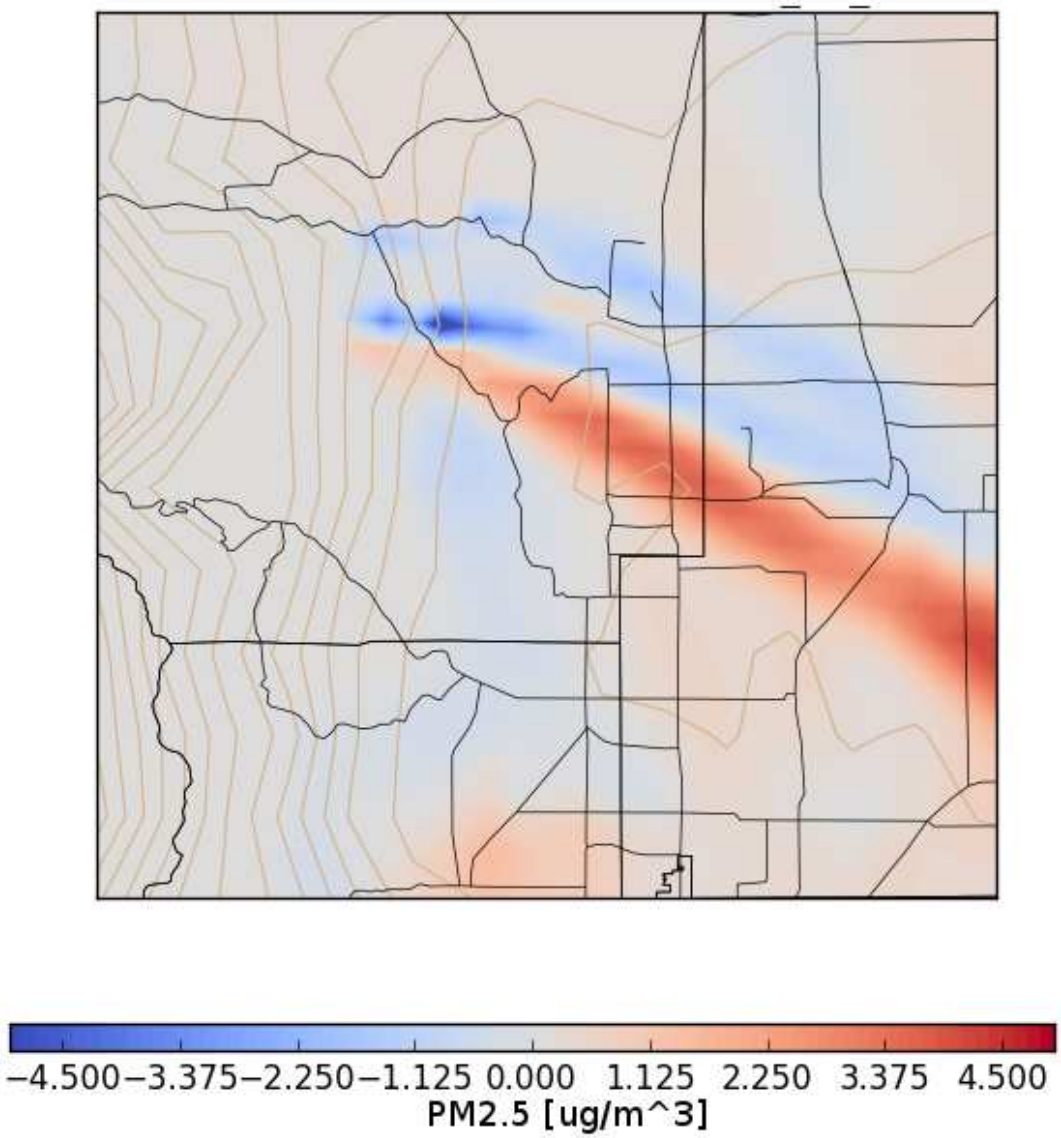

FIGURE A.2. 24-Hour average surface $PM_{2.5}$ concentration difference between GFS and NAM (GFS - NAM) meteorology reanalysis datasets for June 9th 2012.

We see that the simulation with Global Forecasting System (GFS) analysis initial conditions places the smoke plume further north than the North American Mesoscale Forecast System (NAM) analysis simulation, though the total amount of mass at the surface does not appear

to be different. We also note that the difference appears to be on the order of +/- 5 μg m^{-3}, which is a small difference considering concentrations exceed 100 μg m^{-3} in some places.

A.2. Model Paramaterizations

We also investigated the impact of WRF parameterizations on our simulation results.

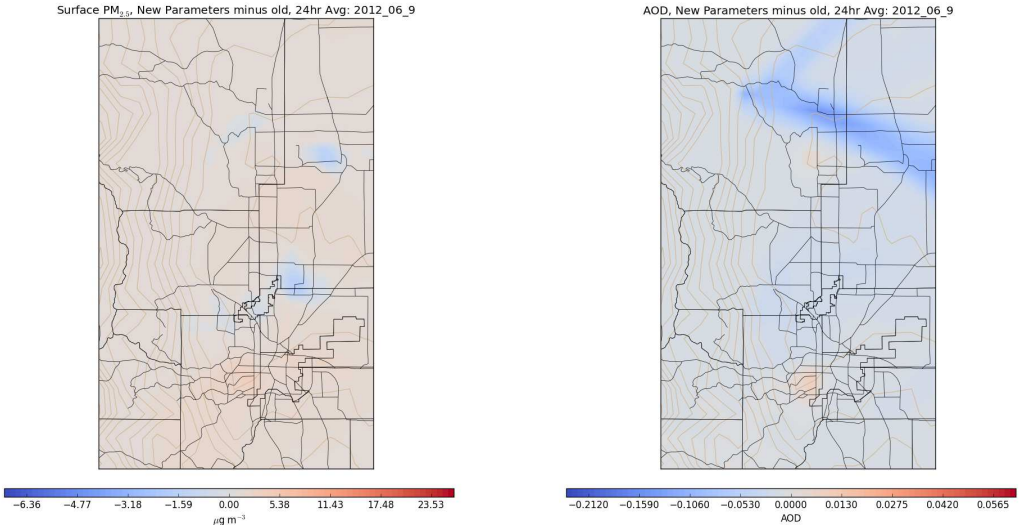

Figure A.3. 24-Hour average surface $PM_{2.5}$ concentration (left) and AOD (right) difference between modified parameterization set and base-case parameterization set.

Figure A.3 shows the differences when we changed our parametrization settings. Recall our base-case planetary boundary layer (PBL) scheme was the Yonsei University (YSU) scheme [36] and Thompson microphysics [37]. We ran an additional simulation with the Mellor-Yamada-Janjic (MYJ) PBL scheme [67] and Morrison microphysics [68]. With respect to surface concentrations, there does not appear to be any systematic difference in the surface $PM_{2.5}$ concentration. We do see a noticeable difference when comparing AOD, with negative values in the region where the smoke from the fire was located. This is likely because the new sets of parameters allowed for more efficient transport of fire smoke out of the domain, whereas the base case simulation kept the smoke closer to the surface, where it lingered.

A.3. Model Resolution

We ran WRF-Chem simulations at the following spatial resolutions: 15 km, 12 km, 4 km, 3 km, and 1 km. We observed that coarser resolutions led to smoke plumes that are diluted across larger grid boxes, producing broader areas of elevated $PM_{2.5}$ with lower concentrations.

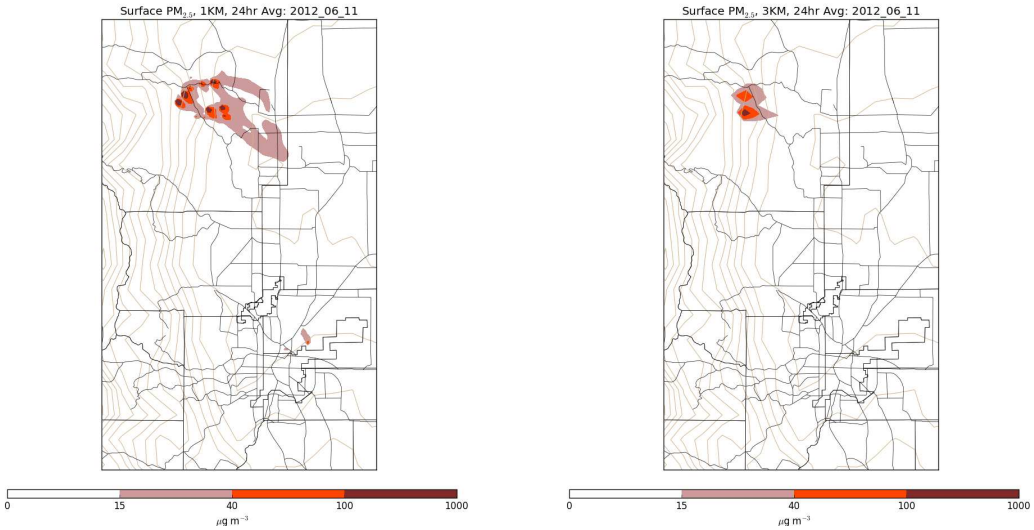

Figure A.4. 24-Hour average surface $PM_{2.5}$ concentration at 1 km resolution (left) and 3 km resolution (right) for June 11, 2012.

Figure A.4 shows the 24-hour average surface $PM_{2.5}$ concentrations for one day during the High Park fire. We see for a constant emission mass, smoke is diluted in coarser resolution simulations, producing broader impacts of lower concentrations, when compared to higher resolution.

We expected that higher resolution model runs would produce more accurate concentrations because they better resolve the complex mountainous terrain. However, many of the canyons that smoke drains through are sufficiently small enough, finer resolution is required to resolve them. Therefore, at 1 km resolution, we did not see improvements in performance of the model.

A.4. Biomass Burning Emissions Inventory

We found that the choice of biomass burning emissions inventory makes a big difference in the resulting simulated $PM_{2.5}$ fields. We conducted a simulation using the FINN emission factors, but substituting Smartfire (http://www.airfire.org/smartfire/) area burned for FINN.

Figure A.5 shows an example of the concentration differences in surface $PM_{2.5}$ that this change produced. Smartfire area burned produced 24-hour averages over 200 μg m^{-3} above those produced by FINN. We believe that the Smartfire burn area produces unrealistically high $PM_{2.5}$ emission rates when paired with FINN emission factors. However, the takeaway from this is that how emissions are calculated, and what inventory is used makes a critical difference when predicting population-level exposure.

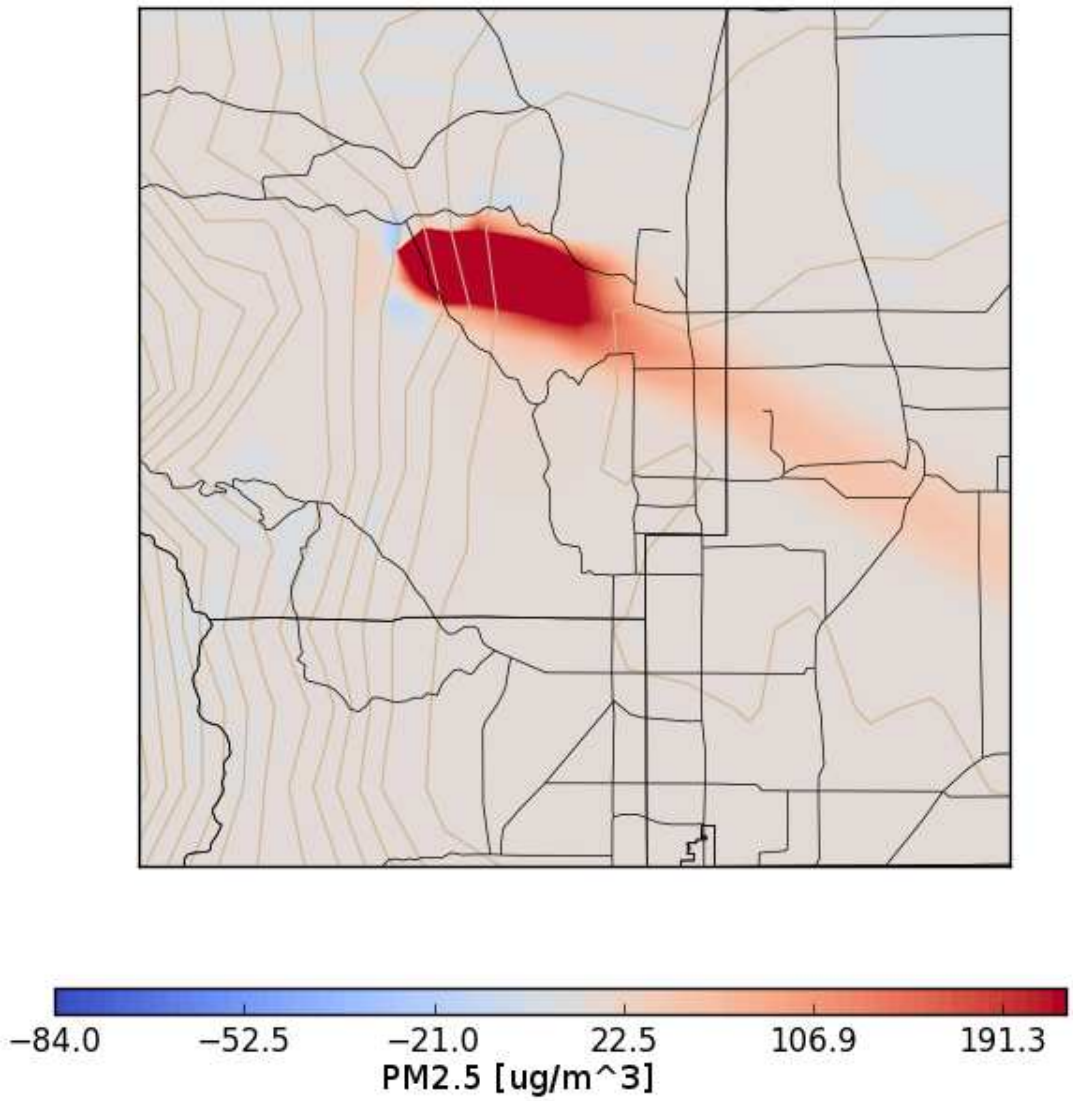

FIGURE A.5. 24-Hour average surface $PM_{2.5}$ concentration difference between FINN/Smartfire inventory and FINN inventory for June 9th, 2012.

APPENDIX B

Other Blend Techniques

Here we will describe all of the different physical (i.e. non-regression) blends we tried, and the results. Many of these were done with the first iteration of data, which consisted of fewer in-situ monitors, and only spanned the month of September.

B.1. Simple Average

The first blend we tried was a proof-of-concept of averaging WRF-Chem and kriging estimates together, with no input from satellite-based observations.

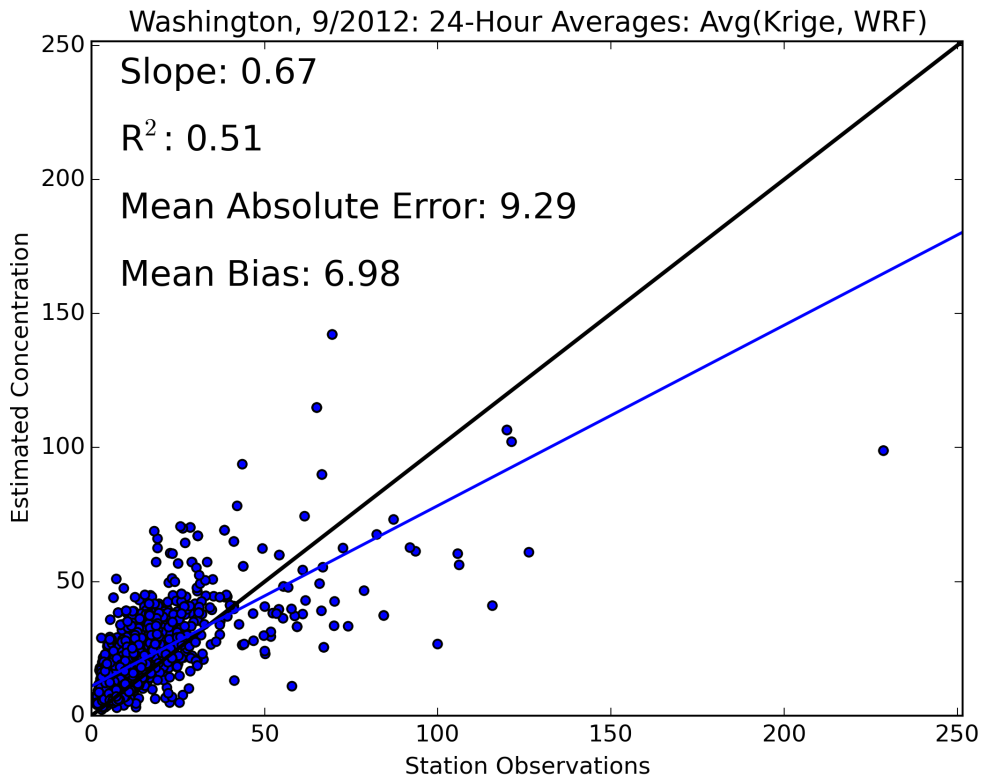

Figure B.1. 24-Hour average surface $PM_{2.5}$ concentration predicted by averaging WRF-Chem and kriging estimates vs. in-situ observations.

We see that this strategy produces summary statistics that are roughly a compromise between the two. We are able to increase the slope closer to one relative to the kriging-only estimates, but introduce considerable error and bias, and lower the R^2 value considerably in the process.

B.2. Weighted Average with Decaying Kriging Influence

One hypothesis we had was that kriging performs better closer to surface monitors and worse further away. We therefore tried to use a weighted average between WRF-Chem and kriging estimates. Our weighting was done with the distance kernel in Equation 9:

$$K = \frac{1}{1 + \left(\frac{D}{D_s}\right)^{\eta}} \tag{9}$$

where η and D_s are parameters that were selected to optimize performance. We use $\eta = 3.0$ and $D_s = 30$km, which produced the best statistics. We then weighted WRF-Chem such that it would be weighted the equivalent of 2 scale distances from a surface monitor in every grid box. Therefore, far from any monitor, the kriging weight kernel will decrease, and the blend will rely more heavily on WRF-Chem. The results of this approach can be found in Figure B.2.

As we see, the results are very similar to the kriging model, with some reduction in performance where the model is introduced. The hope was that in parts of the domain close to multiple surface monitors, the model would entirely use kriging, while further from monitors, it would use WRF-Chem output. However, the parameter set that performed the best (30km scale distance, 3rd power decay rate) leave very little opportunity for the model to contribute.

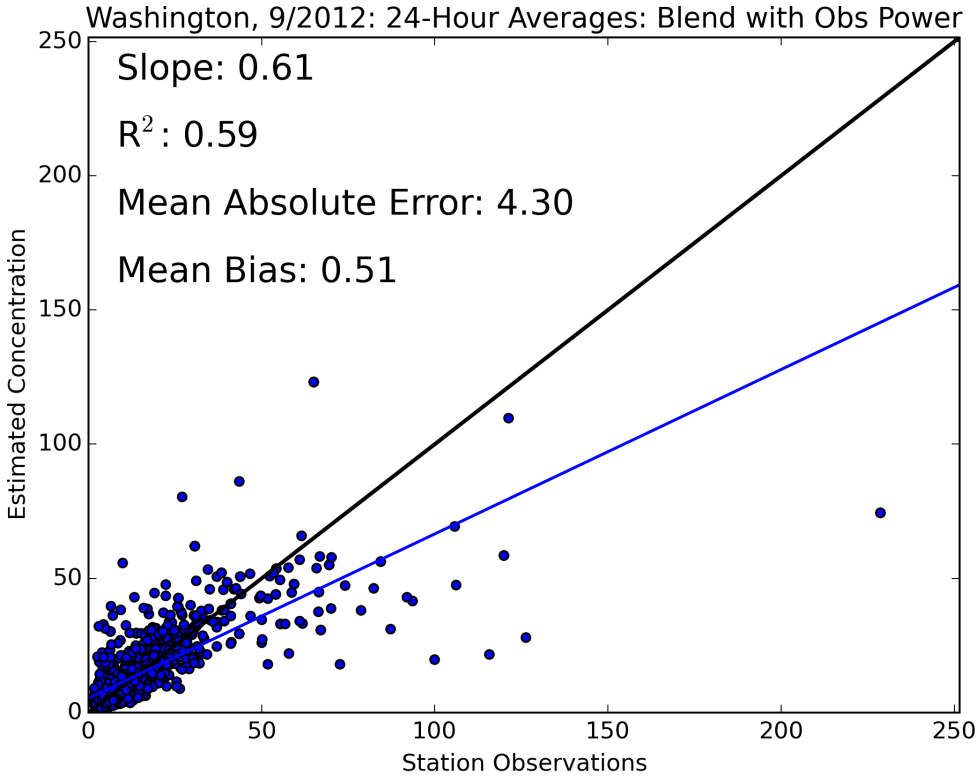

FIGURE B.2. 24-Hour average surface $PM_{2.5}$ concentration predicted by weighting kriging contribution to the blend by the value of the distance kernel given in Equation 9, vs in-situ observations.

B.3. Weighted Average with Decaying Kriging Influence and FMS Modulation of WRF-Chem

We attempted to improve performance by modulating the WRF-Chem contribution to the blend according to the WRF-Chem agreement with MODIS AOD. For every day, we calculate the Figure of Merit in Space (FMS), as defined in Equation 10

$$FMS = \frac{W_{AOD} \cap M_{AOD}}{W_{AOD} \cup M_{AOD}} \tag{10}$$

where $FMS \in [0, 1]$. A value of 1 implies that WRF-Chem and MODIS AOD match exactly, while a value of 0 implies no overlap between the two fields. We used the calculated FMS to scale the WRF-Chem contribution to the weighted average and produced the prediction in Figure B In order to use FMS, a binary threshold must be applied to both input fields. The value of this threshold was tweaked to try to produce the best predictive skill.

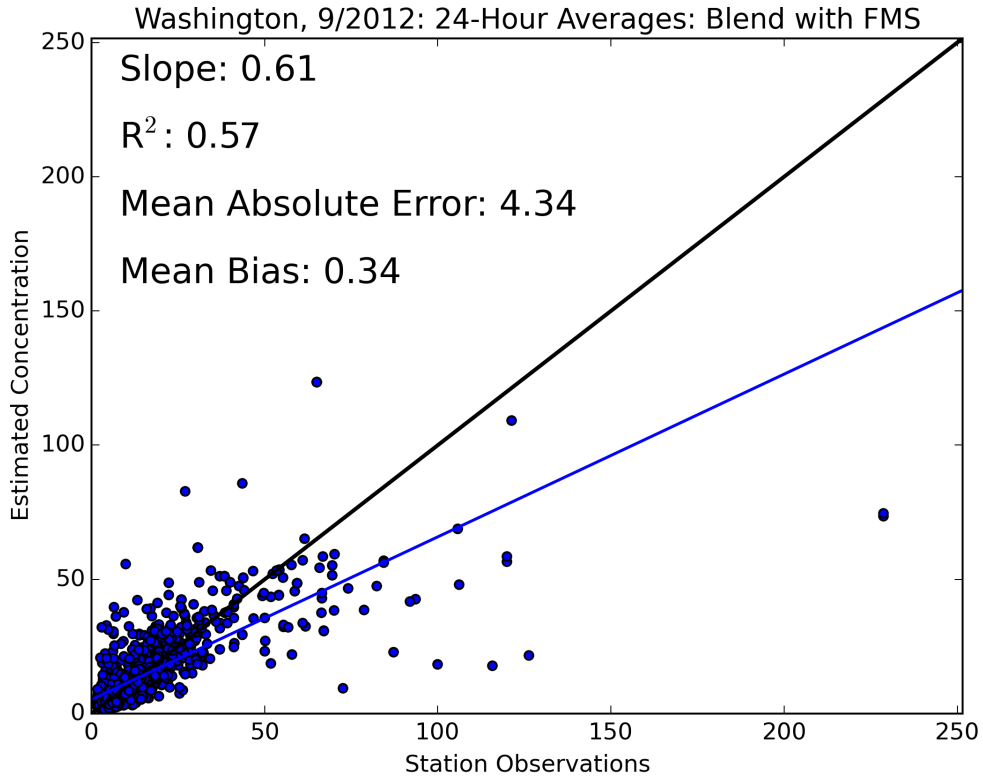

FIGURE B.3. 24-Hour average surface $PM_{2.5}$ concentration predicted by weighting kriging contribution to the blend by the value of the distance kernel given in Equation 9 and modulating WRF-Chem weighting by FMS, vs in-situ observations.

As we see in Figure B.3, this strategy also fails to improve performance over the kriged $PM_{2.5}$ estimates.

B.4. Weighted Average with Decaying Kriging Influence and WRF-Chem/MODIS AOD Correlation

As we saw in Section 3.2.1, the skill of WRF-Chem has a strong regional dependence on its skill. To try to exploit this, we attempted to modulate the influence WRF-Chem on the blending model based on its agreement with MODIS observations.

Figure B.4 shows the correlation coefficient between MODIS and WRF-Chem AOD in each gridbox within our analysis domain.

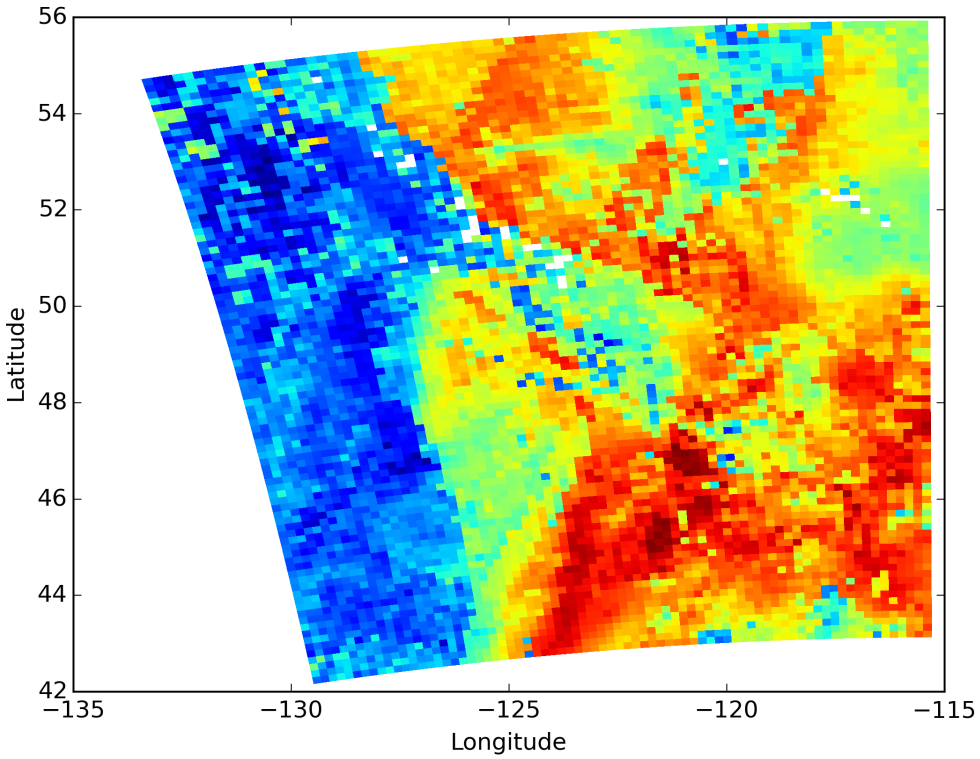

Figure B.4. Correlation coefficient between MODS and WRF-Chem AOD in every gridbox.

As we saw in Sections 3.2.1 and 3.2.3, MODIS and WRF-Chem both perform better in the central and eastern part of the state, and this is reflected in their correlation coefficients with each other. Therefore, we combine the distance kernel weighting of kriging (Equation

9) with WRF-Chem estimates, weighted by the strength of the correlation between MODIS and AOD. The results are shown in Figure B.5.

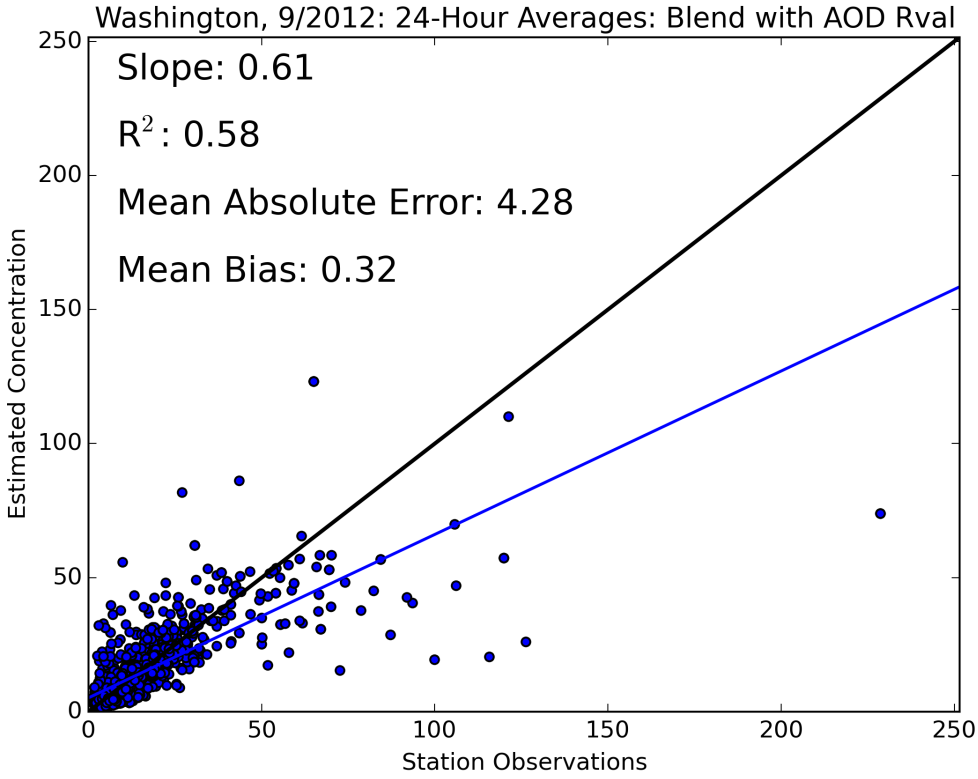

FIGURE B.5. 24-Hour average surface $PM_{2.5}$ concentration predicted by weighting kriging contribution to the blend by the value of the distance kernel given in Equation 9 and modulating WRF-Chem weighting by correlation coefficient between WRF-Chem AOD and MODIS AOD, vs in-situ observations.

As we see in Figure B.5, this approach of using MODIS to module WRF-Chem influence on the blend does not produce any improvement over the kriging-only estimate.

B.5. Weighted Average with Decaying Kriging Influence and WRF-Chem/MODIS Pixel-By-Pixel Comparisons

The final blend we present here tries to take into account spatial and temporal variability in WRF-Chem skill. First, we calculated the distance kernel from Equation 9. Next,

as a baseline, we set the WRF-Chem power equal to kriging power two scale distances from a monitor. We then adjusted the WRF-Chem contribution according to the following procedure.

On each day, for each pixel, we computed a score of agreement between MODIS AOD and WRF-Chem AOD as shown in Equation 11:

$$S = 1 - \left| \log_{10} \left(\frac{M_{AOD}}{W_{AOD}} \right) \right| \tag{11}$$

Once a score is computed, the following algorithm was implemented to determine weighting of WRF-Chem and kriging $PM_{2.5}$ for that location and time:

(1) If $S > 0.7$: Good agreement with satellite: Increase WRF-Chem contribution by factor of 2

(2) If $S < 0.2$: Bad agreement with satellite: remove WRF-Chem from blend

(3) If $0.2 < S < 0.7$: Rescale score such that $S = 0.2; \rightarrow S = 0$, and $S = 0.7; \rightarrow S = 1$. Scale WRF-Chem contribution to blend by this value.

The algorithm was optimized by tweaking both the upper and lower limits for the score, as well as the multiplicative modulation of WRF-Chem influence. The parameters shown in the algorithm produced the best agreement with surface observations; the results are shown in Figure B.6.

We see that performance of this blend is on par with all of the other blends that were tried, slightly worse than kriging by itself.

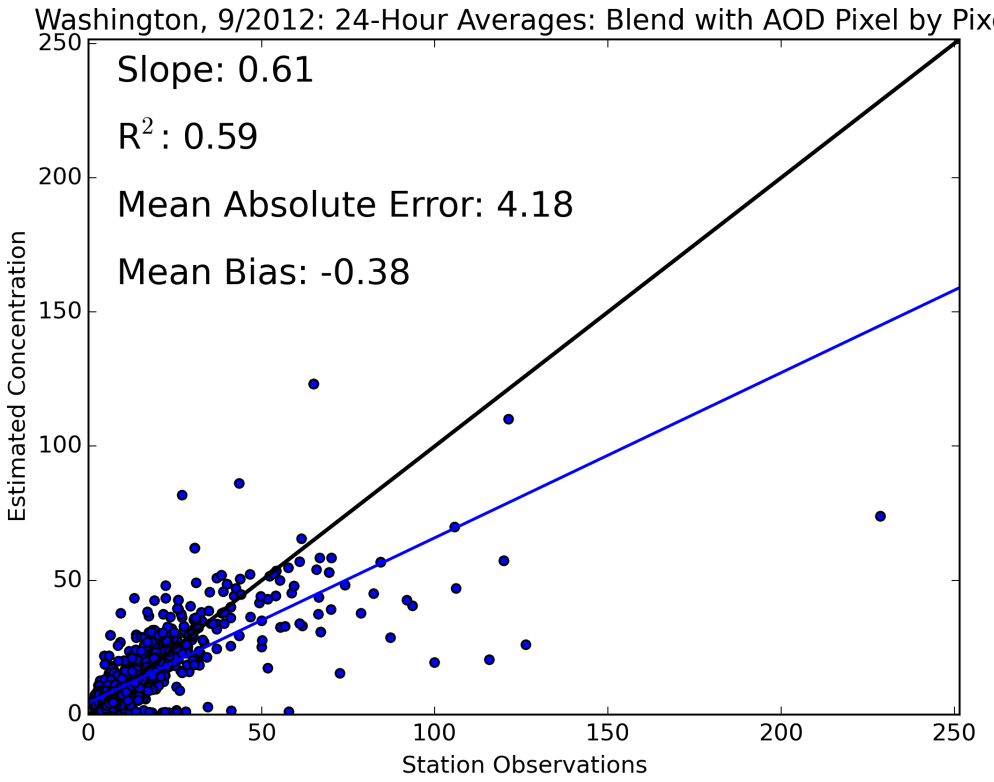

FIGURE B.6. 24-Hour average surface $PM_{2.5}$ concentration predicted by weighting kriging contribution to the blend by the value of the distance kernel given in Equation 9 and modulating WRF-Chem by Equation 11, vs in-situ observations.

Regression is a mathematical tool for optimizing contributions of each dataset for this exact purpose. As we showed in this appendix, we were unable to find a way to combine WRF-Chem, interpolation, and MODIS AOD in an intuitive algorithm that produced dramatic improvements over ordinary kriging. Therefore, we proceeded with the regression approaches outlined in Sections 2.3.1 and 2.3.2.

APPENDIX C

Non WRF-Chem Background

To use these elements in a health study, we also need to provide an estimate of the background $PM_{2.5}$ to look at the health response of exposure to wildfire smoke only. When running WRF-Chem, this is straightforward; we can run the model with biomass burning emissions turned off. However, for these other techniques, we would like to be able to estimate the background without WRF-Chem, especially given our hypothesis that WRF-Chem overestimates the anthropogenic emissions in much of Washington.

To make this non-WRF background estimate, we made the following assumptions:

(1) During the time-period, variability in the background is small relative to wildfire smoke

(2) It is better to underestimate background concentrations than overestimate, to prevent negative concentrations

To estimate the background concentration, we took in-situ measurements of 24-hour average $PM_{2.5}$ concentrations at every surface monitor in and around our domain. We then removed days where NOAA HMS indicated that there was smoke in the vicinity of this surface monitor. Finally, we plotted a series of histograms of the 20th, 40th, 60th, and 80th percentile of $PM_{2.5}$ concentrations, in addition to the mean and median.

These histograms showed that from monitor to monitor, even the 60th percentile was usually not above 10 μg m^{-3}, and that the variability in $PM_{2.5}$ concentrations was small until around the 80th percentile. To avoid overestimating the background, we therefore selected the median value at each in-situ monitor as the input to our model.

We then kriged these median $PM_{2.5}$ values to produce a $PM_{2.5}$ concentration surface throughout our domain. This was assumed to be our background $PM_{2.5}$ concentration for the health effects calculations that made use of kriged or regression $PM_{2.5}$ estimates.

Figure C.1 shows the background $PM_{2.5}$ concentrations in our analysis domain calculated with this procedure. While there are certainly regional variations, there are no values above 12 μg m^{-3} anywhere in the domain, and spatial variability is low.

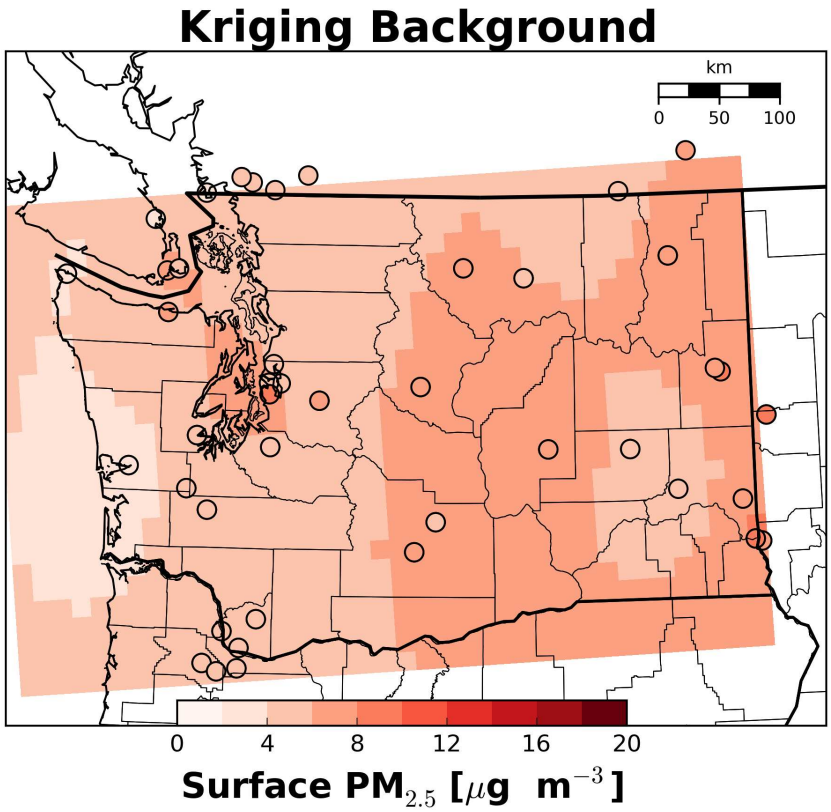

FIGURE C.1. Non-WRF-Chem estimate of background $PM_{2.5}$ concentrations.