THESIS

NATIONWIDE DECADAL SOURCE APPORTIONMENT OF $PM_{2.5}$ WITH A FOCUS ON IRON

Submitted by

Lance Niño

Department of Atmospheric Science

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2021

Master's Committee:

Advisor: Sonia Kreidenweis

Elizabeth Barnes
Tami Bond

ABSTRACT

NATIONWIDE DECADAL SOURCE APPORTIONMENT OF $PM_{2.5}$ WITH A FOCUS ON IRON

Fine particulate matter pollution ($PM_{2.5}$) has detrimental effects on human health, visibility, and the environment. One component of $PM_{2.5}$, aerosol-phase iron, also has a multi-faceted effect on climate. In its largely insoluble iron oxide form, found in dust aerosol, it absorbs shortwave radiation. Emissions from anthropogenic processes, primarily industry and coal combustion, also contain iron, with most of that iron in soluble forms. Soluble iron is an important phytoplankton nutrient and thus its atmospheric abundance is intertwined with carbon sequestration. To ascertain the various sources of $PM_{2.5}$ as well as aerosol-phase iron across the contiguous United States, we used the ME-2 version of PMF to obtain a 10-factor source apportionment solution using IMPROVE data from 2011-2019. The percentage of anthropogenic iron at various sites during that time span varied from nearly none in the inter-mountain West to over 50% over the eastern half of the US. The percentage of total iron detected that was classified as soluble iron reached over 20% along coastal sites but was only around 3% of the total iron emitted. Trends in $PM_{2.5}$ component factors showed a pronounced decrease in $PM_{2.5}$ from coal combustion and various industrial sources during the time period, but trends were mixed and not significant for other sources. Further research is needed applying source apportionment to nationwide speciated datasets like IMPROVE, and a more comprehensive global $PM_{2.5}$ observation network would enable source apportionment on a global scale.

ACKNOWLEDGEMENTS

I acknowledge and thank Colorado State University, my advisor Sonia Kreidenweis, Sagar Rathod and Tami Bond for suggesting this avenue of research, Emily Bian for introducing me to PMF, Phil Hopke and Pentti Paatero for answering my numerous questions on PMF, Nicole Hyslop and Warren H. White for guidance on IMPROVE data, my Master's Committee—Libby Barnes, Tami Bond, and Sonia Kreidenweis—and all the staff and faculty at Colorado State University for supporting me through both a Master's program and a pandemic.

This material is based upon research supported by the U. S. Office of Naval Research under award number N000l4-16-1-2040. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the Office of Naval Research.

TABLE OF CONTENTS

ABSTRACT ii
ACKNOWLEDGEMENTS iii
Chapter 1 - Introduction 1
1.1 Effects of $PM_{2.5}$ 1
1.2 Particulate Iron 2
1.3 Source Apportionment 3
Chapter 2 - Methods 4
2.1 Positive Matrix Factorization 4
2.2 IMPROVE Data and Preprocessing 5
2.3 PMF Setup 9
2.4 Trend Analysis 10
Chapter 3 - Results and Discussion 12
3.1 Factor Composition and Source Interpretation 12
3.1.1 Bromides factor 13
3.1.2 Combustion factor 15
3.1.3 K-rich factor 15
3.1.4 Secondary nitrate factor 16
3.1.5 Coal Combustion factor 17
3.1.6 Oil combustion (shipping) factor 18
3.1.7 Industrial emissions factor 19
3.1.8 Sea salt factor 20
3.1.9 Dust factors 21
3.2 Factor Residual Analysis 23
3.3 Apportionment of $PM_{2.5}$ Fe 27
3.4 Comparison with Measured and Modeled Soluble Fe 31
3.5 Trend Analyses 34
Chapter 4 - Summary, Conclusions, and Future Work 39
REFERENCES 41
APPENDIX 46

CHAPTER 1: INTRODUCTION

1.1 Effects of $PM_{2.5}$

Particulate matter, specifically that which has a diameter of less than 2.5 μm ($PM_{2.5}$), has numerous health and environmental impacts. In sufficient concentrations, $PM_{2.5}$ has been associated with numerous respiratory and cardiovascular problems, including premature death in those with underlying heart or lung disease (Wang et al., 2015). In fact, ambient $PM_{2.5}$ is estimated to cause approximately 4.2 million deaths per year; globally, 9 out of 10 people breathe air containing unhealthy $PM_{2.5}$ levels, mostly in developing countries (Forouzanfar, 2016). $PM_{2.5}$ is also efficient at reducing visibility, because many of these particles have diameters of the same order of magnitude as visible sunlight. The visibility reduction process, Mie scattering, is very efficient with respect to mass in the accumulation mode (100 nm < D_p < 1 μm), so a small amount of $PM_{2.5}$ can greatly degrade visibility (Hand et al., 2014). In addition, when $PM_{2.5}$ settles out due to dry or wet deposition, depending on its chemical composition, it can have wide-ranging environmental effects: acid rain, soil nutrient imbalance, and loss of ecosystem diversity, to name a few (Rai, 2016). For these reasons, the U.S., and the field of atmospheric chemistry in general, have devoted many years of research in order to understand and characterize $PM_{2.5}$ to provide the scientific basis for development of regulations of $PM_{2.5}$ emissions. As a result, atmospheric concentrations of $PM_{2.5}$ have generally decreased in the U.S. over the last couple of decades (Hand et al., 2012; Hand et al., 2013). However, both understanding of sources and policies for active management lag in some developing nations. Now and for the foreseeable future, $PM_{2.5}$ remains one of the most pressing global issues.

1.2 Particulate Iron

Climate change is another pressing global issue that scientists are tasked with understanding. One important and still uncertain climate change forcing is that of short-wave absorbing $PM_{2.5}$. Most $PM_{2.5}$ scatters visible sunlight rather than absorbing it, causing a net cooling effect on Earth's climate (Hartmann, 2016). A minority of $PM_{2.5}$, such as elemental (black) carbon and particulate iron in crustal matter, absorbs most incident sunlight (Derimian et al, 2008). This absorption contributes to Earth's warming, mainly through positive feedbacks like the snow-albedo effect (Qu and Hall, 2007).

We focus on iron in fine atmospheric particulate matter in this study because iron is of interest to climate forcing for an additional reason beyond its optical properties: some atmospheric iron is soluble. In certain emissions, like those associated with oil combustion and wood smoke, the iron is in a chemical form that can be readily dissolved in seawater (Mahowald et al., 2009; Fu et al., 2014; Li et al., 2017; Conway et al., 2019). There, it serves as an important nutrient for phytoplankton (Mahowald et al., 2009; Li et al., 2017). The prevalence and distribution of phytoplankton is, again, relevant to climate change. Phytoplankton play a critical role in sequestering atmospheric CO_2 and transporting it to the deep ocean (Falkowski, 2012). This multi-faceted nature of atmospheric iron means that although there is very little of it relative to total $PM_{2.5}$, iron remains integral to the deeper understanding of Earth's climate, and by extension, climate change. For this reason, it is important to understand the spatial variability of both anthropogenically-emitted and soluble atmospheric iron.

1.3 Source Apportionment

In order to tease out sources of $PM_{2.5}$ that have health and visibility effects—and atmospheric iron for climate effects—we used the technique of source apportionment. Source apportionment consists of using a method or algorithm to find several physical "factors" that can explain observations of speciated composition, and based on the nature of those factors, they can be tentatively assigned to likely sources. Most published source apportionment methods are linear matrix decomposition models like principal component analysis (PCA) or Positive Matrix Factorization (PMF), but there is potential for source apportionment using some other approach, even a non-linear, machine learning approach (Liu et al., 2019; Li et al., 2019; Toms et al., 2020). Most source apportionment approaches are on the site-specific or regional scale, and the history of nationwide, long-term source apportionment is fairly limited. The evolution of this type of source apportionment has mainly been tied to advances in computational power. Blifford and Meeker used a traditional rotated factor analysis (FA) in the first nationwide $PM_{2.5}$ source apportionment study (Blifford and Meeker, 1967). They were able to identify some factors (e.g., traffic emissions) but could not attribute mass to factors because of their method. In the mid-1980s, studies began to emerge using PCA to ascertain source apportionment nationwide—this time using the results in a comprehensive mortality analysis (Thurston and Spengler, 1985; Ozkaynak and Thurston, 1987). In 2011, Thurston et al. iterated on this approach by combining FA and PCA techniques to create a more detailed look at U.S. $PM_{2.5}$ pollution (Thurston et al., 2011). What all three of these studies lacked, however, was a way to incorporate measurement uncertainty into source apportionment; Thurston himself laments this in his 2011 work. In this study, we attempt to solve these aforementioned science questions by incorporating measurement uncertainty and using more sophisticated source apportionment techniques on a nationwide scale.

CHAPTER 2: METHODS

2.1 Positive Matrix Factorization

We used the Positive Matrix Factorization (PMF) technique to identify and apportion mass concentrations to sources. At its core, PMF is a more refined version of non-negative matrix factorization (NMF), a quick and well-understood method of matrix factorization. Both PMF and NMF constrain their input and output to positive, non-zero values, as required since chemistry data are physically positive and non-zero. PMF uses an algorithm to solve the equation

$$x_{ij} = \sum_{k=1}^{p} g_{ik} f_{kj} + e_{ij}$$

where x_{ij} is the concentration dataset with j number of variables (constituents in this analysis) and i number of data points. The right-hand side of the equation shows that PMF decomposes the data into an array of p factors, with the number of factors chosen by the user. The process by which p is chosen will be discussed later. The f_{kj} matrix shows how much of each constituent j is in each factor k, while the g_{ik} matrix determines how much of each factor k is in each data point i. The residual e_{ij} is any data in x that could not be explained by the p factors. The specific algorithm in PMF minimizes the objective function Q through gradient descent:

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \left[\frac{x_{ij} - \sum_{k=1}^{p} g_{ik} f_{kj}}{\sigma_{ij}} \right]$$

The specifics of PMF are discussed in greater detail in Paatero and Tappert (1994) and Paatero (1997). Importantly, both x_{ij} and σ_{ij}, the uncertainty of every data point, are required to run a PMF analysis. This differs from most other source apportionment methods, including its predecessor, NMF. For this analysis, we used the Multilinear Engine 2 (ME-2) version of PMF

(Paatero, 1999) running on the EPA PMF 5.0 wrapper (Paatero et al., 2014). This package available from the EPA has been used in many prior $PM_{2.5}$ source apportionment studies (Viana et al., 2007; Ulbrich et al., 2009; S. Lee et al., 2018).

2.2 IMPROVE Data and Preprocessing

The chemical composition of atmospheric aerosols can provide clues to the sources impacting air quality at a measurement site. The IMPROVE network (Figure 2.1) represents a long-term, speciated aerosol composition dataset with a large number of sites across the U.S.

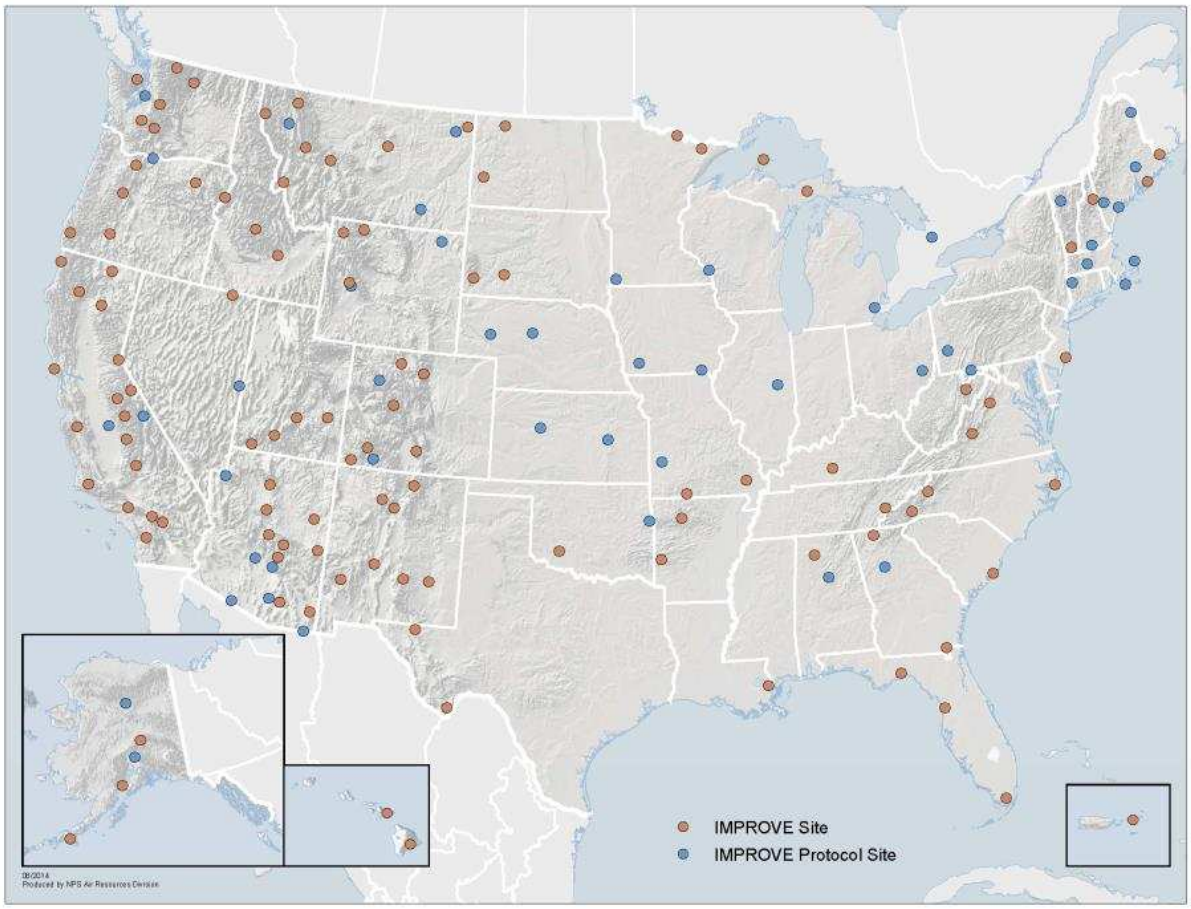

Figure 2.1. IMPROVE and IMPROVE Protocol sites as of 2016. This figure is from the IMPROVE website: http://vista.cira.colostate.edu/Improve/improve-program/

IMPROVE sites monitor in federal Class I areas, usually national parks. IMPROVE protocol sites are separately sponsored by state, regional, tribal, and national organizations, and may be located in urban regions. They employ the same instrumentation, monitoring, and analysis protocols as IMPROVE, expanding the coverage of the sampling. Data from the network have been used in a large number of studies of trends in visibility and emissions (e.g., Hand et al., 2012, 2013, 2019; Sickles II and Shadwick, 2015; Prenni et al., 2016, Malm et al., 2017). Iron is among the elements analyzed in the fine particle fraction ($PM_{2.5}$), and thus the IMPROVE dataset is ideal for purposes of further understanding atmospheric concentrations and sources of iron. Although the network was started in 1988, we used data only from the period 2011-2019, avoiding data collected across changes in analytical protocols that were implemented starting in 2011 (Indresand and Dillner, 2012).

IMPROVE data advisories are issued to alert users to changes in protocols, sampling and analysis anomalies, and other data quality issues. They are published at http://vista.cira.colostate.edu/Improve/data-advisories/. We reviewed all advisories issued in the 2011-2019 time frame for applicable adjustments to the dataset before using it as input to PMF. In accordance with the recommendation from a published IMPROVE advisory, we reduced the reported vanadium concentrations from January 2011 to October 2017 by 23% (http://vista.cira.colostate.edu/improve/Data/QA_QC/Advisory/da0038/da0038_V_advisory.pdf) . Total reported elemental and organic carbon concentrations did not have corresponding uncertainties in the database before 2017. Only the fractions themselves (EC1, EC2, EC3, OP, OC1, OC2, OC3, OC4) had uncertainties and minimum detection limits (MDL). The IMPROVE Standard Operating Procedure (SOP) had some of the information required to back out pre-2017 uncertainties (Hyslop and White, 2009; IMPROVE SOP #351,

http://vista.cira.colostate.edu/improve/wp-content/uploads/2019/06/IMPROVE-SOP-351_Data-Processing-and-Validation_06.2019.pdf). The uncertainty and MDL of the fractions and post-2017 total EC and OC concentrations are given in the SOP as:

$$\sigma = 1000\frac{ng}{\mu g} * \frac{\sqrt{\left(Max(\sigma_{dfb},t)\right)^2 + (f*(A-B))^2}}{V_{C\,module}} \quad (1)$$

$$mdl = 1000\frac{ng}{\mu g} * \frac{Max((95th\ percentile\ of\ FB-B),t)}{V_{C\,module}} \quad (2)$$

These equations reduce down to a simple form:

$$\sigma = \sqrt{\left(\delta * MDL_{EC,OC}\right)^2 + (f * C)^2} \quad (3)$$

where

$$MDL_{OC} = \sqrt{{MDL_{OC1}}^2 + {MDL_{OC2}}^2 + {MDL_{OC3}}^2 + {MDL_{OC4}}^2} \quad (4)$$

$$MDL_{EC} = \sqrt{{MDL_{EC1}}^2 + {MDL_{EC2}}^2 + {MDL_{EC3}}^2} \quad (5)$$

and C is the concentration of EC or OC in $\frac{\mu g}{m^3}$, f is a fractional uncertainty parameter given by the IMPROVE SOP, and δ is a parameter that must exist because $mdl \neq \sigma_{dfb}$, the standard deviation for the field blank measurements. The MDL is held to a higher standard than one standard deviation, but its precise value is not reported in the SOP. We backed out δ by using the EC and OC uncertainties reported after 2017 and manipulating the above equation to solve for δ. These values were $\delta \approx 0.481$ for EC and $\delta \approx 0.642$ for OC. Notice that the MDL for organic pyrolized carbon (OP) is not accounted for in the total EC or OC uncertainty calculation. However, the mass from OP carbon is added to OC and subtracted from EC.

Table 2.1 lists the species reported for each sample, which are obtained every three days. The bolded species were used for source apportionment in this study. These elements and

molecules were chosen both to paint a clearer picture of source apportionment and to account for a large portion of total $PM_{2.5}$ mass. The choice of chlorine vs. chloride and 3*sulfur vs. sulfate followed recommendations from the IMPROVE team as well as those in recent published work (e.g., Correal et al., 2020). We included every site in CONUS that had continuous measurements from 2011 through 2019. Zero or negative concentrations were set to a random number between zero and the MDL. The uncertainties of those concentrations were recalculated according to the IMPROVE SOP. We removed any data points that had one or more missing values for any of our selected components, either in concentration or uncertainty.

Table 2.1. IMPROVE species considered for source apportionment. **Bolded** species were used in the PMF analysis.

Aluminum	Arsenic	**Bromine**	**Calcium**	**Elemental Carbon**
Organic Carbon	**Chlorides**	Chlorine	**Chromium**	**Copper**
Iron	**Lead**	**Magnesium**	**Manganese**	**Total $PM_{2.5}$**
Total PM_{10}	Coarse mass	**Nickel**	**Nitrates**	Phosphorous
Potassium	Rubidium	**Selenium**	**Silicon**	**Sodium**
Sulfur	**Sulfate**	**Titanium**	**Vanadium**	**Zinc**

Ultimately, this approach included 148 sites and yielded approximately 176,000 samples. The ME-2 engine can only calculate 100,000 samples at a time (Paatero, *personal communication*), so we split the data in half using the Python pandas "sample" method with random seed 125. Additionally, we created a separate bootstrapped version of our data with 8,000 samples and the same pandas method. This was used for the EPA PMF 5.0 software's

uncertainty calculations and was necessary to accommodate its input limitations. The results of those uncertainty calculations were robust (Figure A.1).

2.3 PMF Setup

The first crucial component of PMF setup is to determine which variables are "weak" and "bad"; that is, which elements should have their uncertainties tripled or be excluded entirely, respectively. We consulted the EPA PMF 5.0 user's manual for this information (https://www.epa.gov/sites/production/files/2015-02/documents/pmf_5.0_user_guide.pdf). We set variables whose signal-to-noise ratio (S/N), calculated by the EPA PMF 5.0 software, was between 0.5 and 1.0 to "weak", and those less than 0.5 to "bad". Uncertainties were therefore tripled for Cr, Ni, and Se. As and Rb were ultimately removed entirely using this method, as noted in Table 2.1. Additionally, we set total $PM_{2.5}$ as a weak total variable, tripling its uncertainty as well.

The second crucial component of PMF setup is to determine how many factors to choose. The choice is largely subjective, but there are some objective methods. As suggested by Ulbrich et al. (2009), we chose the objective method of finding the "elbowing point", or second derivative maximum, of a plot of number of factors vs. Q/Q_{exp}. Q is the same objective function that PMF tries to minimize, while Q_{exp} is the expected Q if all PMF factors were perfect at explaining the data. Both Q and Q_{exp} are given by the PMF output and the second derivative was calculated with a finite centered difference:

$$f''(x) \approx \frac{f(x+h)-2f(x)+f(x-h)}{h^2} \quad (6)$$

where $f(x)$ is Q/Q_{exp} at some factor number x and h is a factor difference of one. Essentially, this is a way to normalize Q to the number of factors. Using this analysis, we found that either 9 or 10 factors is the optimal number (Figure 2.2). We chose 10 factors for this analysis because all

10 made physical sense and were intuitive. Once the 10-factor solution was deemed optimal, we ran 20 PMF runs initialized with random seed 125 and let the software choose the optimal run based on lowest Q. The results presented herein are the factors produced in that optimal run.

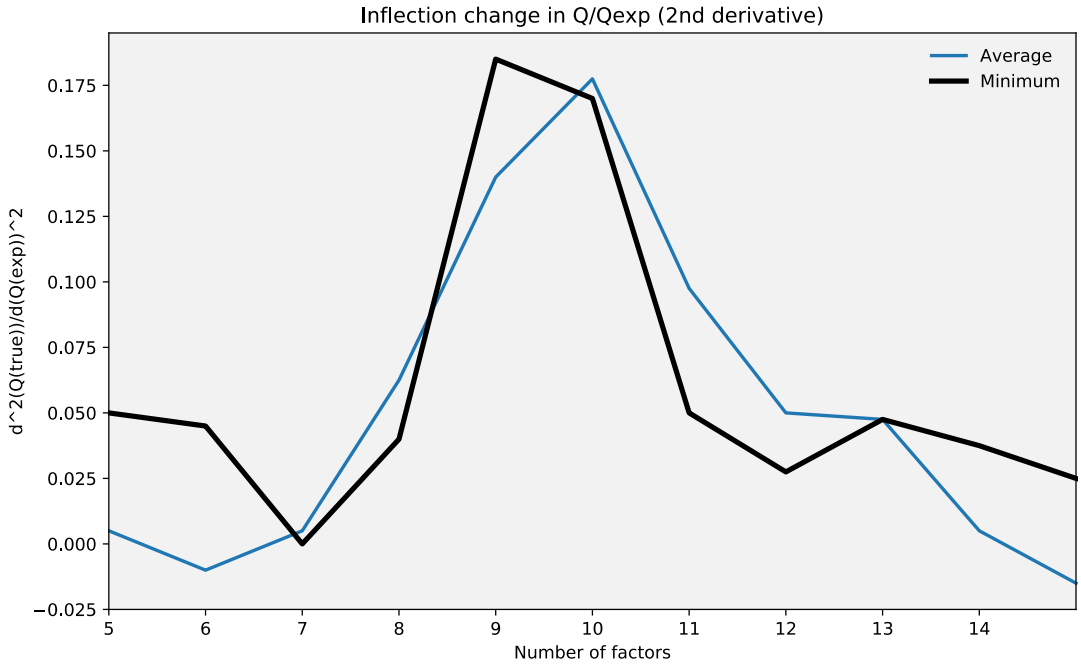

Figure 2.2. "Elbowing point", or second derivative maximum of Q/Q_{exp} when plotted against number of factors. The greatest drop in fitness parameter with respect to factor increase is shown to be either 9 or 10 factors. The blue line indicates the average values among 20 runs, each having a slight perturbation, while the black line indicates just the minimum Q value among the 20 runs, which was the run used for analysis.

2.4 Trend Analysis

To investigate trends in the factors, we employed the Theil-Sen Regressor. We chose to do this firstly for consistency between trend analyses compared to prior work with IMPROVE data (Hand et al., 2012), and secondly because this method is insensitive to outliers compared to a simple ordinary least squares regression (Sen, 1968). With interannual variability quite high for multiple factors and only a 9-year dataset, we deemed a non-parametric estimator best. The

Theil-Sen estimator is defined as the median of the slopes of all pairs of sample points (Theil, 1992):

$$median\left\{\frac{y_j - y_i}{x_j - x_i}\right\} \quad (7)$$

Our sample points are annual mean $PM_{2.5}$ contribution ($y_{i,j}$) from the factor k versus year ($x_{i,j}$) for every IMPROVE site. The Theil-Sen regressor was computed using Python's SciPy library.

CHAPTER 3: RESULTS AND DISCUSSION

3.1 Factor Composition and Source Interpretation

The distribution of the included species, including Fe (labeled FEf in the figure), into the ten factors is best viewed through a fingerprint plot (Figure 3.1) which is closely linked to the factor matrix f_{kj}. The compositions (factor profiles) of the ten factors are provided in Table 3.1 and Table A.1. Based on both the dominant species included in each factor and prior work on typical source profiles (Thurston and Spengler, 1985; Ozkaynak and Thurston, 1987, Thurston et al., 2011) we assigned the indicated source types. Each is discussed further below.

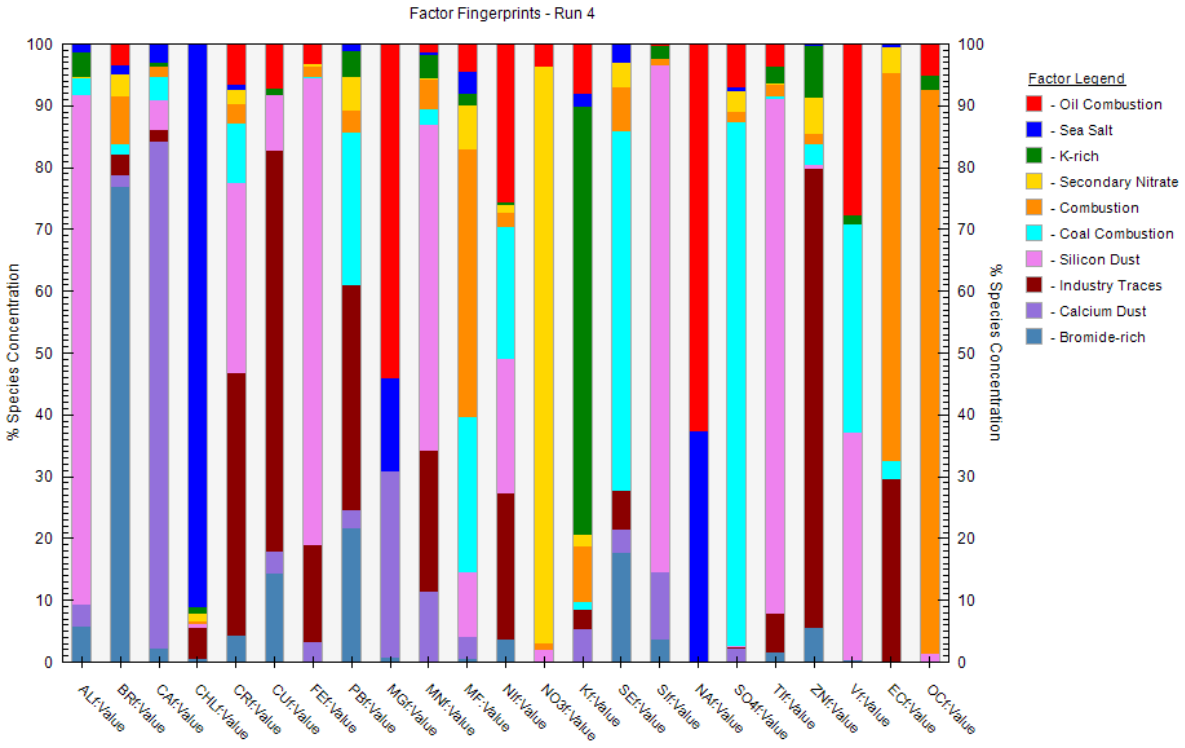

Figure 3.1. Factor Fingerprints, 10 factors: For each of the 23 variables, 22 of which nearly sum up to total $PM_{2.5}$, this plot shows the percentage of each element associated with each factor.

Table 3.1. Factor profiles (f_{kj} matrix) for the ten-factor solution. The units are μg/m^3 per standard deviation of the factor. Species which were key in identifying each factor are given in **bold**. This table is shown with alternate units of percentage of species sum in Table A.1.

	Comb.	Second. Nitrates	Coal Comb.	Silicon-Dust	K-rich	Br-rich	Oil Comb.	Industry	Calcium-dust	Sea Salt
Al	0	1.98E-05	1.36E-03	**4.16E-02**	1.96E-03	2.65E-03	0	0	1.52E-03	6.62E-04
Br	9.20E-05	4.32E-05	1.80E-05	0	0	**1.28E-03**	0	5.03E-05	0	4.51E-05
Ca	9.44E-04	1.27E-04	1.06E-03	2.20E-03	8.27E-05	0	0	1.27E-03	**3.35E-02**	1.21E-03
Chl.	1.31E-04	1.06E-03	0	1.43E-04	8.77E-04	3.09E-03	0	2.21E-03	0	**7.00E-02**
Cr	1.54E-06	2.32E-06	9.06E-06	**3.13E-05**	0	8.25E-06	5.50E-06	**4.19E-05**	0	1.04E-06
Cu	0	0	0	3.94E-05	0	6.73E-05	2.77E-05	**2.66E-04**	6.10E-06	0
Fe	6.29E-04	2.45E-04	0	**2.78E-02**	0	0	1.08E-03	5.61E-03	1.09E-03	0
Pb	0	2.39E-05	1.33E-04	0	2.84E-05	**1.47E-04**	0	**2.52E-04**	1.62E-05	5.67E-06
Mg	0	0	0	0	0	1.28E-04	7.96E-03	0	4.29E-03	2.54E-03
Mn	4.32E-05	3.40E-06	2.28E-05	**5.15E-04**	3.21E-05	0	1.17E-05	**2.23E-04**	1.07E-04	4.11E-06
$PM_{2.5}$	**1.76E+00**	2.77E-01	1.10E+00	4.68E-01	5.99E-02	5.97E-02	1.93E-01	0	1.44E-01	1.51E-01
Ni	1.40E-06	9.12E-07	2.19E-05	1.91E-05	2.53E-07	5.43E-06	**2.10E-05**	2.06E-05	0	6.86E-07
NO_3	5.13E-03	**3.19E-01**	0	2.11E-03	1.40E-03	0	1.26E-02	3.07E-03	0	0
K	3.14E-03	0	0	4.53E-06	**2.62E-02**	0	3.59E-03	8.20E-04	2.67E-03	5.58E-04
Se	7.79E-06	5.53E-06	**9.33E-05**	0	0	2.87E-05	9.76E-08	9.89E-06	5.47E-06	4.51E-06
Si	1.30E-03	0	0	**9.54E-02**	2.26E-03	3.68E-03	5.55E-05	1.15E-04	**1.19E-02**	2.01E-04
Na	0	0	0	0	0	0	3.78E-02	0	0	**2.50E-02**
SO_4	0	1.79E-02	**6.49E-01**	5.61E-03	0	0	5.64E-02	0	7.41E-03	5.89E-03
Ti	6.08E-05	1.95E-06	1.72E-05	**2.57E-03**	7.76E-05	5.19E-05	1.12E-04	1.88E-04	0	0
Zn	3.28E-06	1.27E-04	4.80E-05	0	2.63E-04	4.64E-05	0	**1.99E-03**	0	1.17E-05
V	0	0	4.63E-05	4.87E-05	0	1.75E-06	**3.83E-05**	6.05E-07	0	0
EC	**8.80E-02**	5.72E-03	3.69E-03	0	0	0	0	**4.38E-02**	2.60E-04	5.27E-04
OC	**8.36E-01**	0	0	1.89E-02	7.98E-03	1.84E-02	4.65E-02	0	0	0

3.1.1. Bromides Factor

The Br-rich (bromides) factor, shown in Figure 3.2, seems to be spatially uniform throughout the contiguous U.S. The majority of particulate bromine was present in this factor, along with some trace metals. Of the species loaded into the factor in this work, Ca and Pb were also loaded into a "bromide" factor in the results reported by Lee et al. (1999). Those authors conducted a PMF analysis on speciated PM_{10} samples collected from 1992-1994 in Hong Kong,

and they identified a bromine-rich factor associated with vehicle/road dust emissions. This source definition would be consistent with the widespread and uniform occurrence of this factor in our analysis. However, we note that vehicular and road dust emissions would be expected to differ between the U.S. and Hong Kong based on differences in vehicle fleet and fuel composition, and those emissions would have also varied between the early 1990's and the time period of our analysis. However, more recently, Jain et al. (2017) also found Br to be associated with vehicular emissions. In our study, very little mass was encapsulated in this factor, with average daily $PM_{2.5}$ ranging from 20-50 ng/m^3, depending on the site.

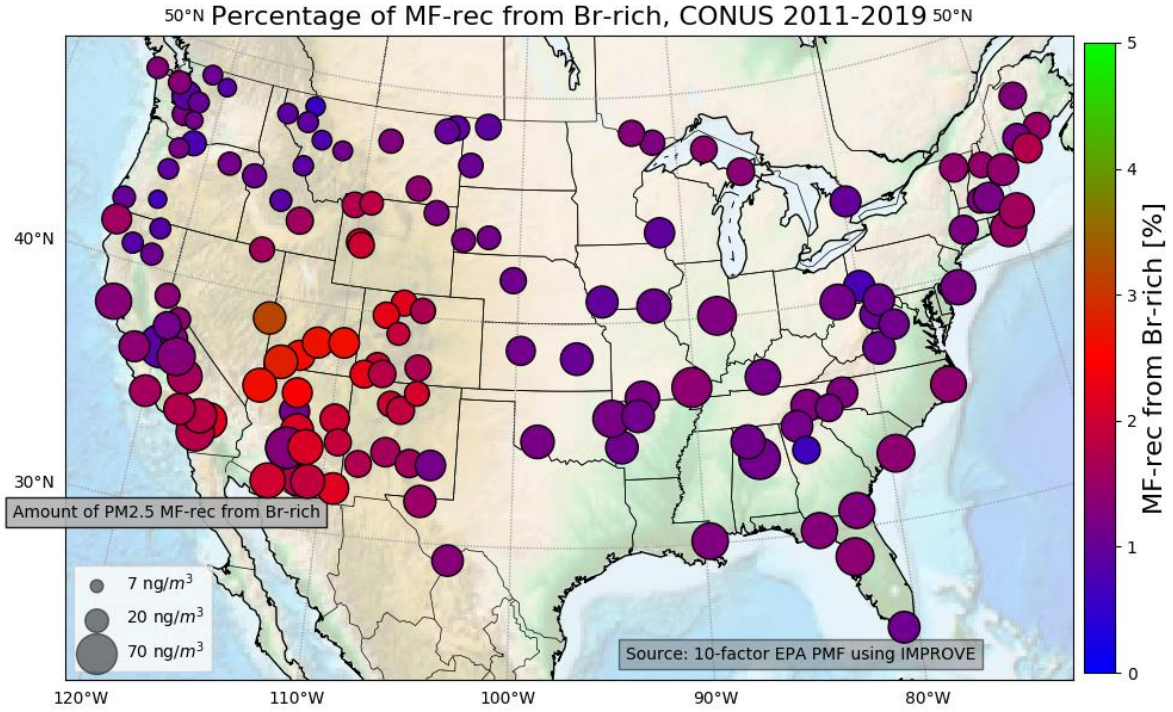

Figure 3.2. Percentage of $PM_{2.5}$ from the bromides factor. The size of the circle indicates average concentration of $PM_{2.5}$ from this factor at each site, while the color indicates the percentage of the total fine mass concentration attributed to this factor. This factor has low concentration everywhere, so note the color scale of 0-5%, which differs from the other plots in this section.

3.1.2 Combustion (elemental- and organic-carbon-rich) factor

This factor is associated with the most gravimetrically-measured $PM_{2.5}$ in the PMF model (average over all sites of 41%; Figure 3.3), and contains mainly EC and OC. There are some trace metals alongside carbon, but not nearly as much as we might expect for biomass burning or other combustion processes. A possible mechanism for this is that particle-phase OC can be formed by secondary chemical processes occurring in biogenic and anthropogenic emissions. Thus, although this factor was labeled as "combustion" it also likely represents contributions from secondary organic aerosol (SOA), including biogenic emissions and their respective SOA. There is also likely some contribution from traffic given the high elemental carbon fraction in the factor. "Combustion" therefore is used as a shorthand, recognizing that a number of sources contribute to this EC- and OC-rich factor.

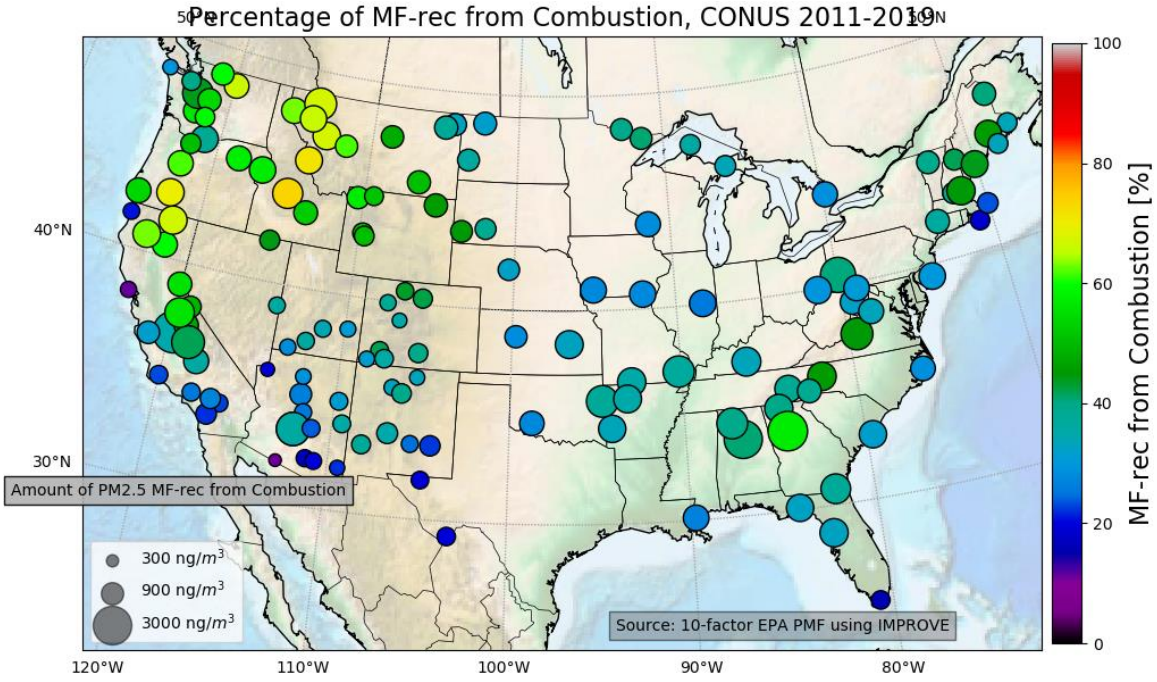

Figure 3.3. Percentage of $PM_{2.5}$ from the combustion factor. The size of the circle indicates average concentration of $PM_{2.5}$ from this factor at each site, while the color indicates the percentage of the total fine mass concentration attributed to this factor.

3.1.3 K-rich factor

The K-rich factor contains most of the potassium, a typical tracer for primary emissions from biomass burning, but K is also found in soils (Pachon et al., 2013). K emissions from biomass burning are typically associated with flaming rather than smoldering combustion of wood and plant fuels (Lee et al., 2010) and are not expected to necessarily correlate with secondary organic aerosol (here, as OC). The contributions from this factor, shown in Figure 3.4, could be summed with those from the combustion factor to estimate total contributions from biomass burning and other OC sources. Both of these factors show similar spatial variability, with the highest percentage of total $PM_{2.5}$ from combustion occurring in the Pacific NW. There is high potassium in the desert Southwest as well, though that may be associated with soil.

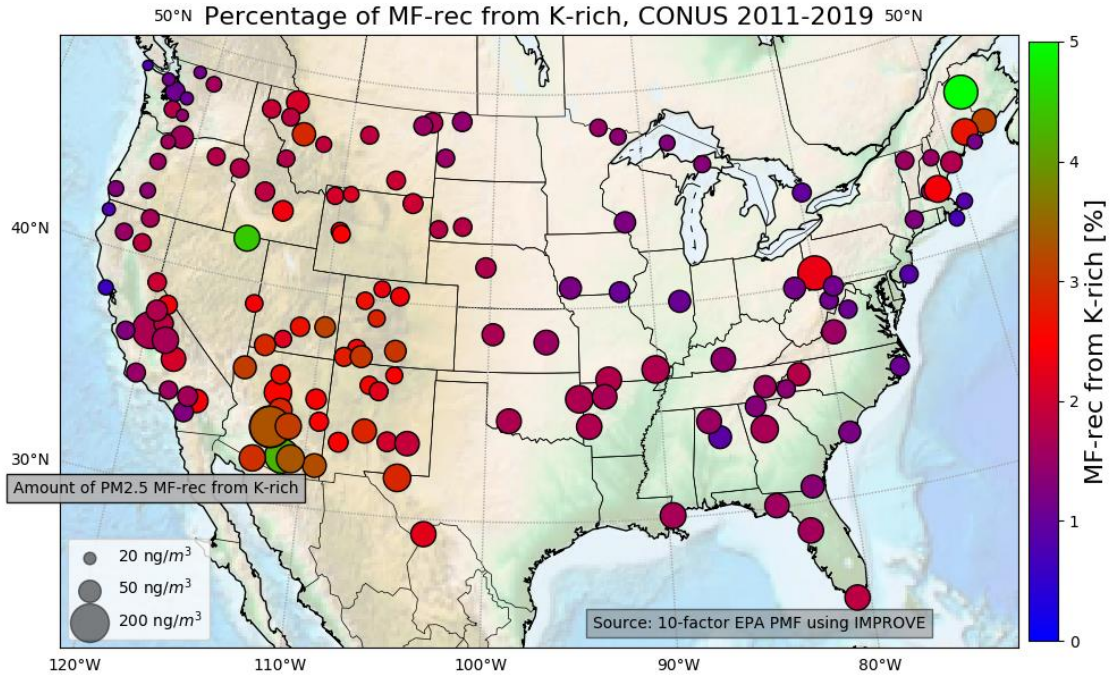

Figure 3.4. Percentage of $PM_{2.5}$ from the K-rich factor. The size of the circle indicates average concentration of $PM_{2.5}$ from this factor at each site, while the color indicates the percentage of the total fine mass concentration attributed to this factor. This factor has low concentration everywhere, so note the color scale of 0-5%, which differs from the other plots in this section.

3.1.4 Secondary nitrate factor

Particulate NO_3 is a secondary aerosol formed via NOx reactions, and it also requires the presence of ammonia (or another cation; see T. Lee et al., 2008) to form. This factor contributed significantly to $PM_{2.5}$ mass concentrations in regions of intensive agriculture, such as the Midwest and San Joaquin Valley where NO_3 accounts for almost 40% of $PM_{2.5}$ (Figure 3.5). In the desert Southwest, nitrate may be associated primarily with soil in the coarse mode (Lee et al., 2008), but is measured here only in the fine mode.

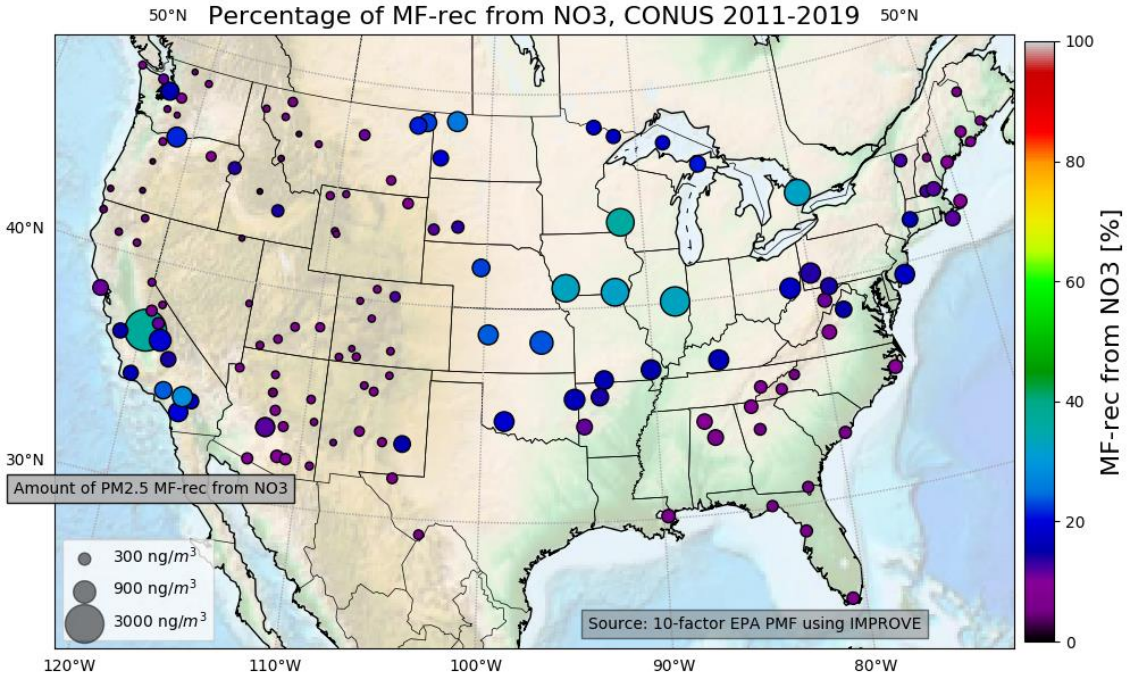

Figure 3.5. Percentage of $PM_{2.5}$ from the secondary nitrate factor. The size of the circle indicates average concentration of $PM_{2.5}$ from this factor at each site, while the color indicates the percentage of the total fine mass concentration attributed to this factor.

3.1.5 Coal combustion factor

The coal combustion factor had a unique and easily-identifiable signature (Thurston et al., 2011), namely the presence of SO_4 and Se. The east-west gradient for this factor (Figure 3.6), in both total mass contributed and the percentage of fine mass that it represents, was expected, with increases toward the eastern seaboard in regions where energy generation relied on coal

combustion during the time period. There, coal combustion accounted for approximately 50%, or 2 µg/m^3, of modeled mass. West of the Rockies, those numbers dip to around 20% and 0.5 µg/m^3, respectively.

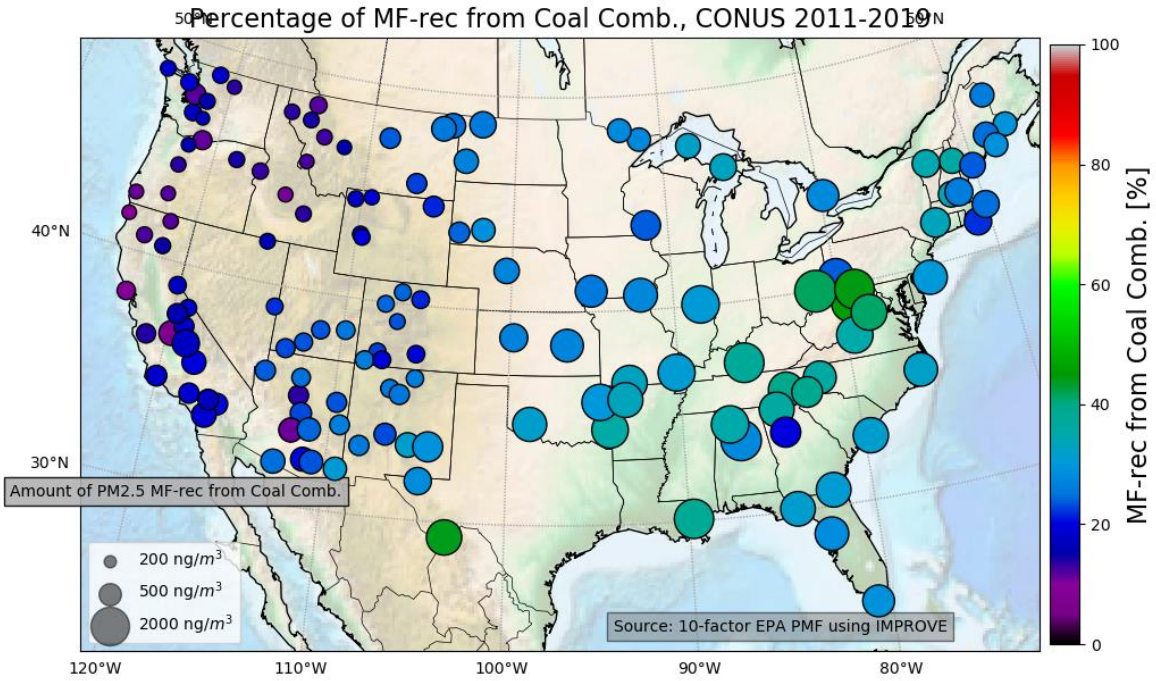

Figure 3.6. Percentage of $PM_{2.5}$ from the coal combustion factor. The size of the circle indicates average concentration of $PM_{2.5}$ from this factor at each site, while the color indicates the percentage of the total fine mass concentration attributed to this factor.

3.1.6 Oil combustion (shipping) factor

The oil combustion factor relied on primarily nickel and vanadium as identifiers, pointing toward the shipping industry as the major source (Corral et al., 2020). Some Na and Mg that likely originated from sea salt, as well as spatial variability with the highest mass contributions at the coasts (Figure 3.7), confirmed the likely contributions from shipping. Although both Ni and V are trace metals and are measured and predicted with lower S/N, PMF generated this factor with surprising consistency. Along the coasts, up to 20% of the $PM_{2.5}$ modeled could be predicted by this factor.

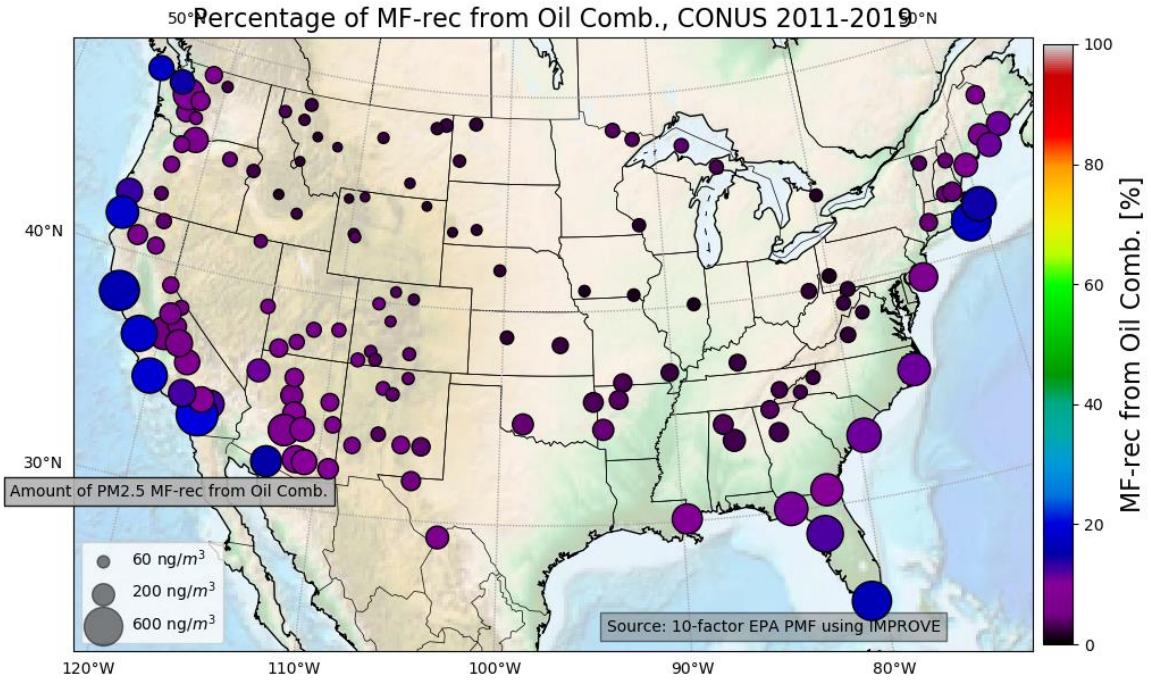

Figure 3.7. Percentage of $PM_{2.5}$ from the oil combustion factor. The size of the circle indicates average concentration of $PM_{2.5}$ from this factor at each site, while the color indicates the percentage of the total fine mass concentration attributed to this factor.

3.1.7 Industrial emissions factor

The industrial emissions factor contained very little mass but many important trace metals associated with anthropogenic activity. High Cu and EC indicate traffic emissions; Zn and Pb, metal smelting; and Fe and Mn, steel production (Thurston et al., 2011). Apart from a couple of outlier sites near large industrial operations, most sites had only a small contribution from this factor (Figure 3.8).

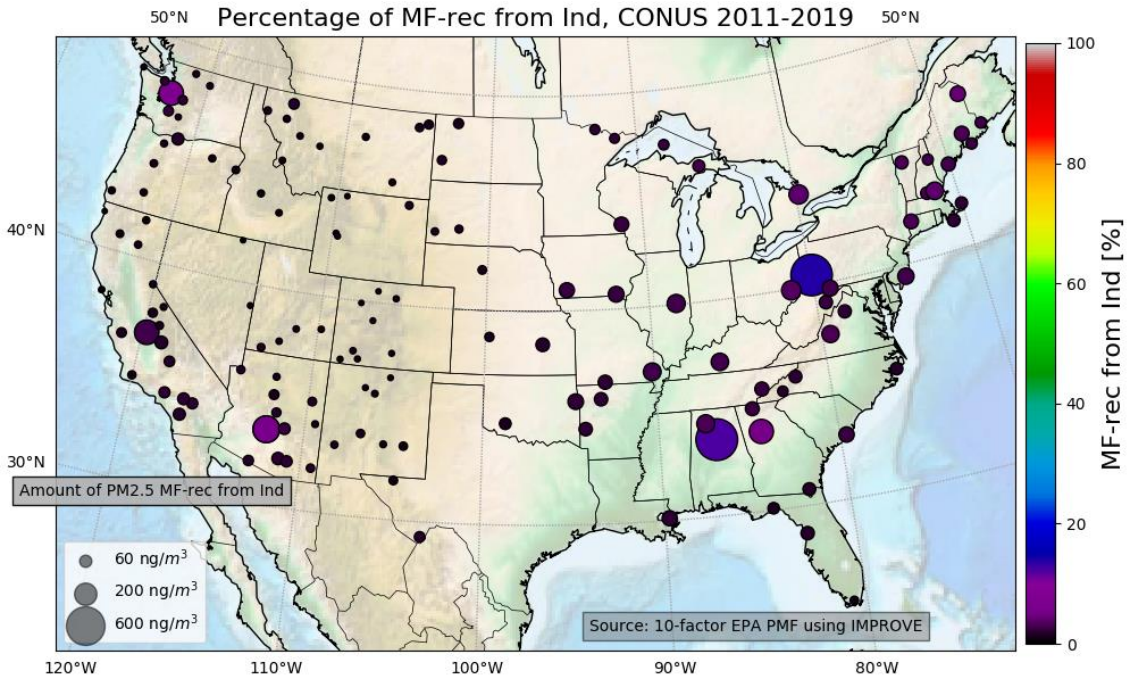

Figure 3.8. Percentage of $PM_{2.5}$ from the industrial emissions factor. The size of the circle indicates average concentration of $PM_{2.5}$ from this factor at each site, while the color indicates the percentage of the total fine mass concentration attributed to this factor.

3.1.8 Sea salt factor

The sea salt factor consisted mainly of sodium and chloride, and represented both fresh and reacted sea salt. As one might expect, this factor was found primarily near the coasts (Figure 3.9). There, as low as 10% and as high as 50% (Point Reyes, CA) of $PM_{2.5}$ mass was attributed to this source.

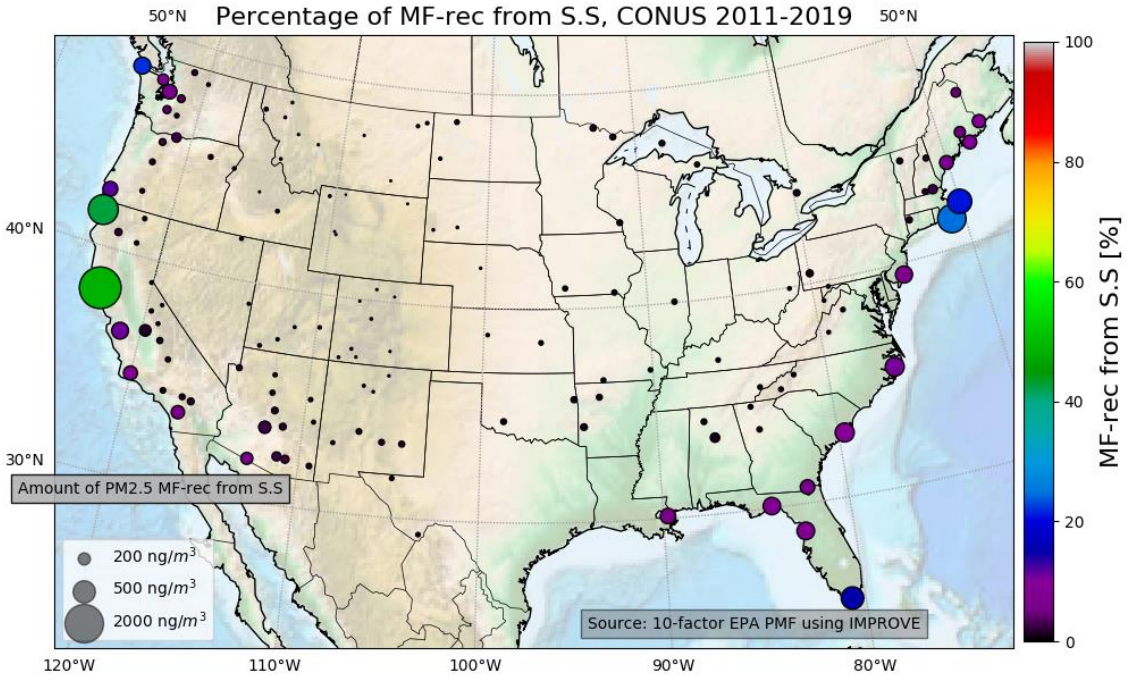

Figure 3.9. Percentage of $PM_{2.5}$ from the sea salt factor. The size of the circle indicates average concentration of $PM_{2.5}$ from this factor at each site, while the color indicates the percentage of the total fine mass concentration attributed to this factor.

3.1.9 Dust factors

Two dust factors were identified: silicon-rich and calcium-rich dusts (Figures 3.10 and 3.11). The highest concentrations of Si-based dust were found mainly in the desert Southwest and comprised, on average, 20% of the $PM_{2.5}$ mass concentrations. Crustal elements, including Al, Fe, Si, and Ti, were the primary components loaded into this factor. The Ca-rich dust factor also had its highest concentrations in the desert Southwest, with some other locations of relatively high concentrations in the Midwest likely associated with agricultural activity. In the Southwest, Ca-based dust represented a smaller faction (5%) of $PM_{2.5}$ concentrations than did Si-based dust (note the scale for the bubble sizes is about a factor of three smaller for the Ca-rich dust factor).

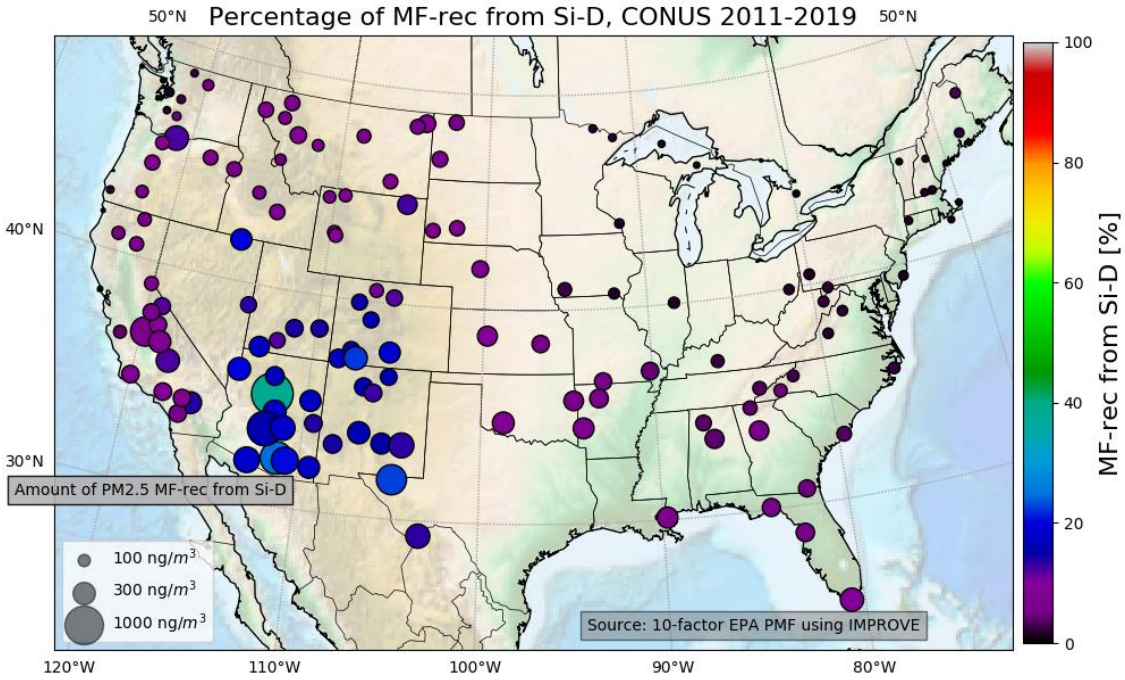

Figure 3.10. Percentage of $PM_{2.5}$ from the silicon-rich dust factor. The size of the circle indicates average concentration of $PM_{2.5}$ from this factor at each site, while the color indicates the percentage of the total fine mass concentration attributed to this factor.

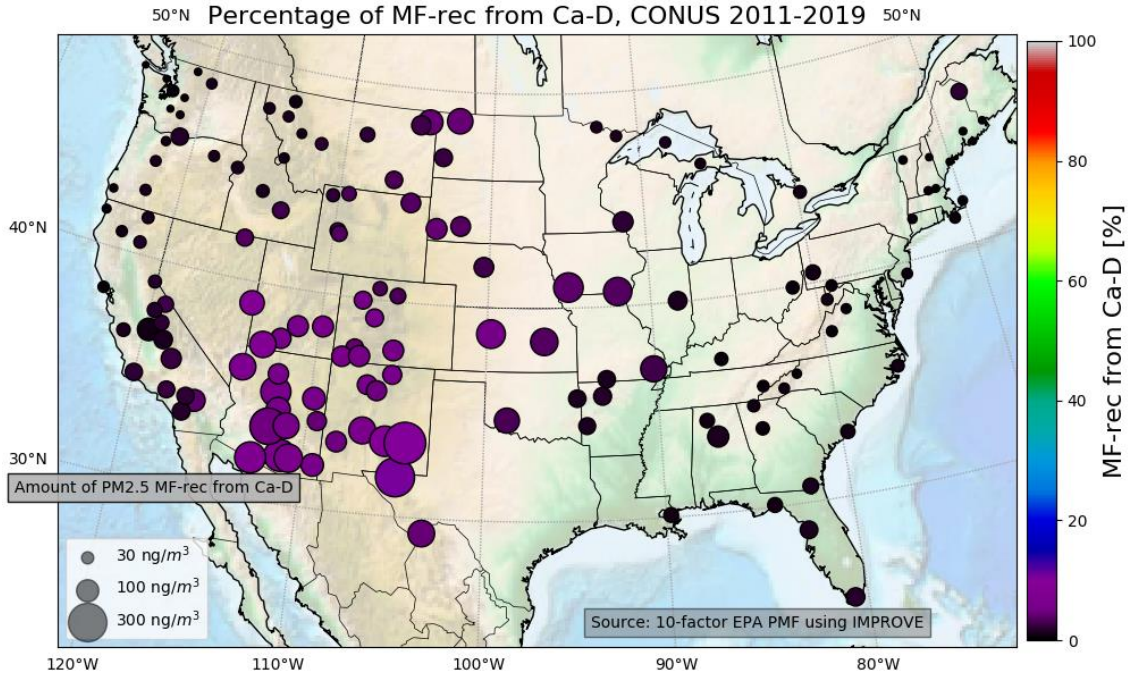

Figure 3.11. Percentage of $PM_{2.5}$ from the calcium-rich dust factor. The size of the circle indicates average concentration of $PM_{2.5}$ from this factor at each site, while the color indicates

the percentage of the total fine mass concentration attributed to this factor. Note the concentration scale (circle size) is different from that in Figure 3.10.

3.2 Factor Residual Analysis

To assess how well PMF accounted for individual data point variability, we tested predicted values $g_{ik} * f_{kj}$ against the data x_{ij}, similar to the PMF definition equation in section 2.1, isolating the residual e_{ij}. We made scatter plots of these variables, tested goodness-of-fit with both ordinary least squares (OLS) and RANSAC (Fischler and Bolles, 1981, in order to reduce the weight of outliers) regressors, and computed an r^2 for the OLS fit. These results are shown in Figure 3.12. Most elements were predicted quite well by PMF. We were able to reconstruct gravitationally-measured $PM_{2.5}$ to an r^2 of 0.934. Though most of the speciated elements measured by IMPROVE are included in the PMF model, the calculated mass is an underestimation by nearly 50% on average. This is expected, since the reconstruction did not account for ammonium ions that are certainly present in association with sulfate and nitrate; this species is not measured with its associated uncertainty by IMPROVE. Also, assumptions about mineral oxygen and other elemental constituents are used to rebuild the dust mass concentration from key elements, which are reliable on average but may not be applicable at a particular site and were not included here. Similarly, the organic carbon (OC) concentrations provided are only representative of the C mass concentrations, and the hydrogen and oxygen contents of the OC have not been included. Finally, the gravimetric measurements are not made at a consistently low relative humidity (http://vista.cira.colostate.edu/improve/Data/QA_QC/Advisory/da0035/da0035_IncreasedRH.pdf). Since the atmospheric aerosol is hygroscopic, to varying degrees depending on composition,

the gravimetric mass concentration includes some aerosol water that is not included in the reconstruction.

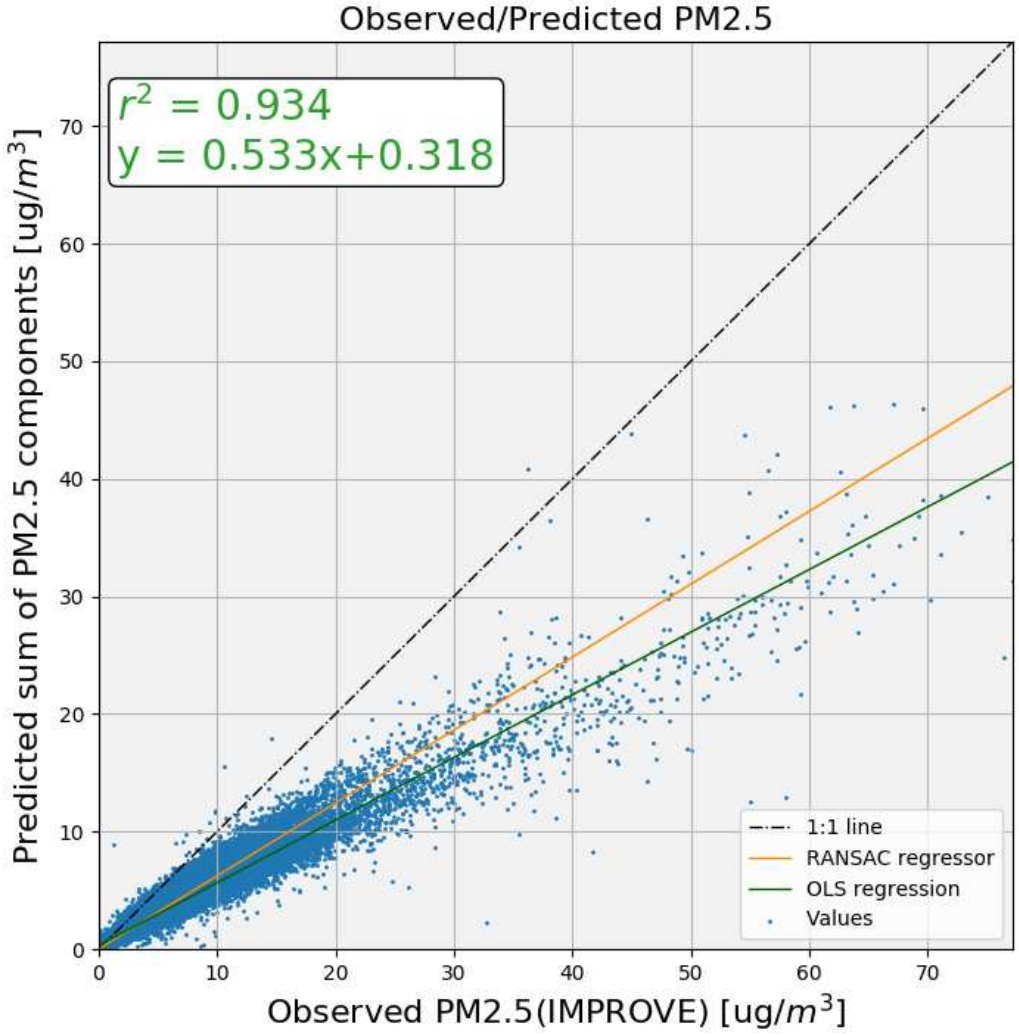

Figure 3.12. Predicted (sum of all speciated components of PMF factors) vs. observed (IMPROVE) for gravitationally-measured $PM_{2.5}$.

We also created spatial variability maps to assess where and with what elements PMF clustered either well or poorly. In general, most elements with high S/N (>2), were well-predicted ($r^2 > 0.7$). Some noisy elements, especially those with S/N < 1, which we set to "weak" for the PMF runs, had poor predictive capabilities—this was true for Cr, Cu, Pb, Ni, Se, and V. Figure 3.13 shows the site-specific analyses for iron (Fe), since apportionment of Fe to various

sources is a focus of this work. Overall, ~94% of the variability was predicted (Figure 3.14); r^2 > 0.8 were achieved at nearly all sites, with a few clear exceptions. We note that r^2 in this sense only predicts the ability of the PMF algorithm to cluster the input species and does not imply the accuracy of these factors. The sites in Alabama and Georgia are IMPROVE protocol sites, operated as urban air quality sites, with measurements that likely include local sources that are not well represented by the nationally-averaged analysis performed in this work.

Reasons for the lower r^2 at the more rural sites in Idaho, Virginia, and Minnesota are less clear. As shown in the Appendix, the Idaho site is located close to a major molybdenum mine, the Minnesota sites are located close to iron mines (Figure A.2), and Minnesota and Virginia are among the nation's larger producers of non-fuels metals (Figure A.3). Depending on the IMPROVE site's location, meteorology, and siting and nature of the particular metals operations, the measurements may be influenced by sources that are also not well represented by the nationally-averaged source factors.

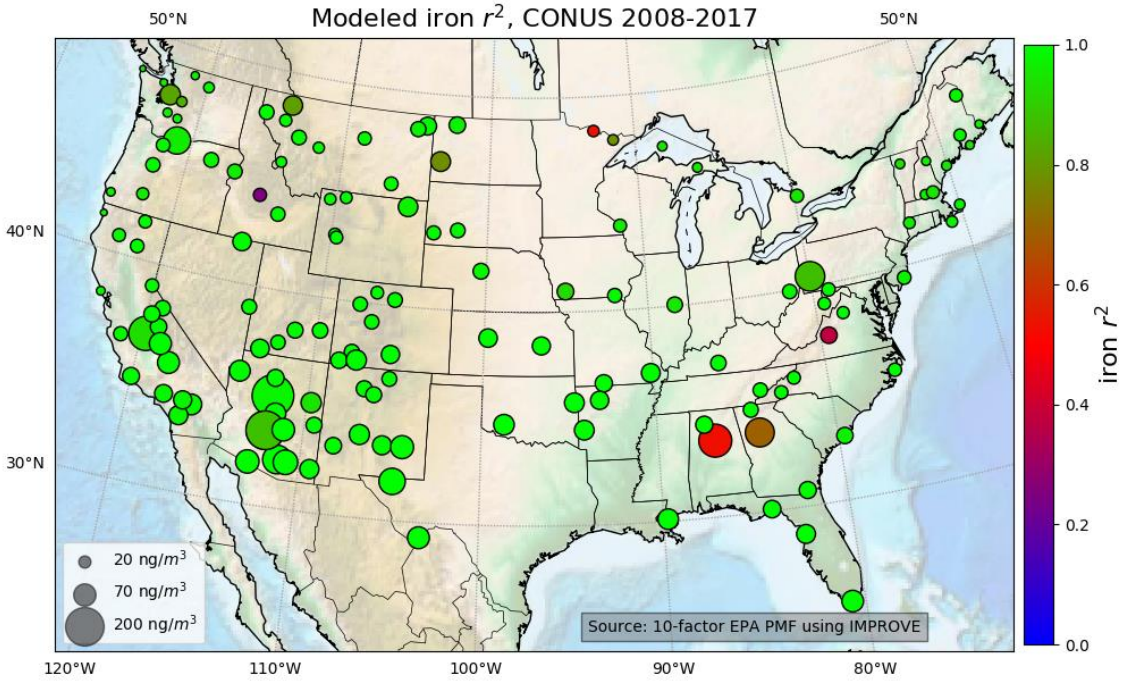

Figure 3.13. Site-specific iron r^2, comparing Fe mass concentrations for observed (IMPROVE) vs. predicted (reconstruction with PMF factors). The color of the circle indicates the r^2 value. The size of the circle corresponds with average total particulate iron mass concentration measured at the site.

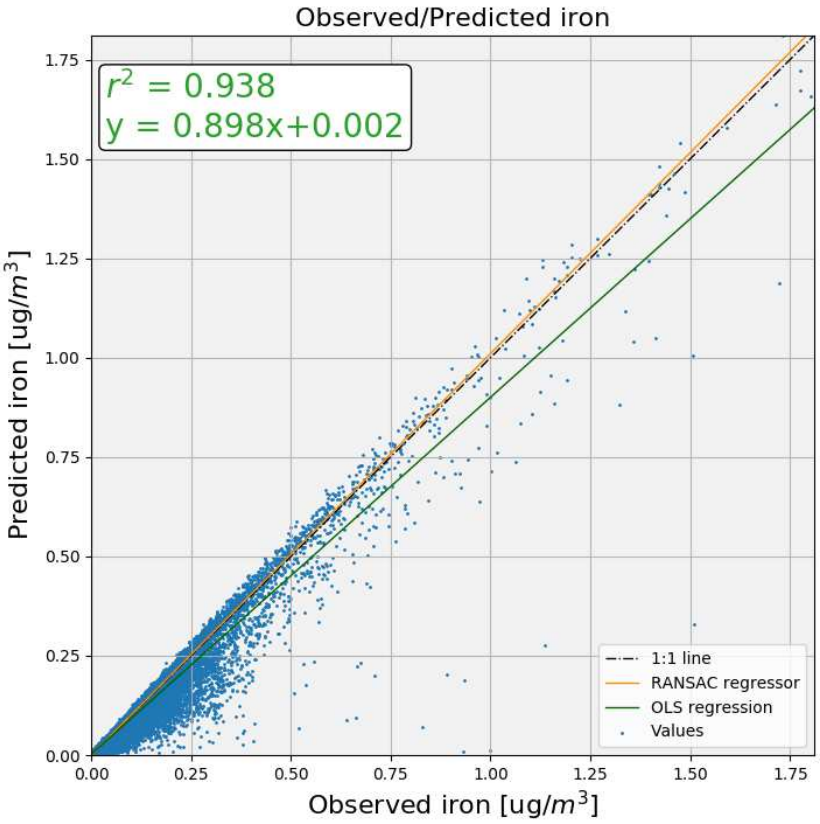

Figure 3.14. Predicted (sum of all PMF factors' iron) vs. observed (IMPROVE) for iron in $PM_{2.5}$. The RANSAC regressor, which is insensitive to outliers, lies nearly on the 1:1 line.

3.3 Apportionment of $PM_{2.5}$ Fe

In this section, we use the PMF results, shown in Figure 3.15, to explore the sources of atmospheric fine particulate Fe in more detail. Iron was associated with 6 factors, with varying degrees of prevalence: combustion (0-5%), Ca-dust (0-5%), and NO_3 (0-5% of total iron) were all relatively minor sources. Of note is combustion Fe, which is usually soluble (Mahowald et al., 2009; Fu et al., 2014; Ito et al., 2019). The site- and time-averaged concentration associated with that source was around 1 ng/m^3. An important form of soluble Fe often comes from oil combustion, which comprised 10%-50% of the total Fe concentrations at coastal sites. Iron concentrations for this factor ranged from 1-4 ng/m^3. The Fe associated with industry sources varied greatly across the U.S. Some urban sites, like the ones in Pittsburgh, Birmingham, and

Seattle, had the vast majority of observed Fe associated with this source, and the specific industry was likely steel production (Luo et al., 2008). However, most Eastern sites had around half of the observed Fe concentrations attributed to this factor, likely representing a variety of industries. West of the Rockies, save for a few urban sites, Fe attributed to the industry factor was nearly non-existent. The final and most dominant Fe factor was that of Si-dust, which is the source of most of the dust in the global atmosphere. Most of the West had nearly all of the observed Fe apportioned to dust, while in the East, a minority of Fe came from this factor. Interestingly, Miami, FL, which sees many Saharan dust plumes, looks more like the West in this regard.

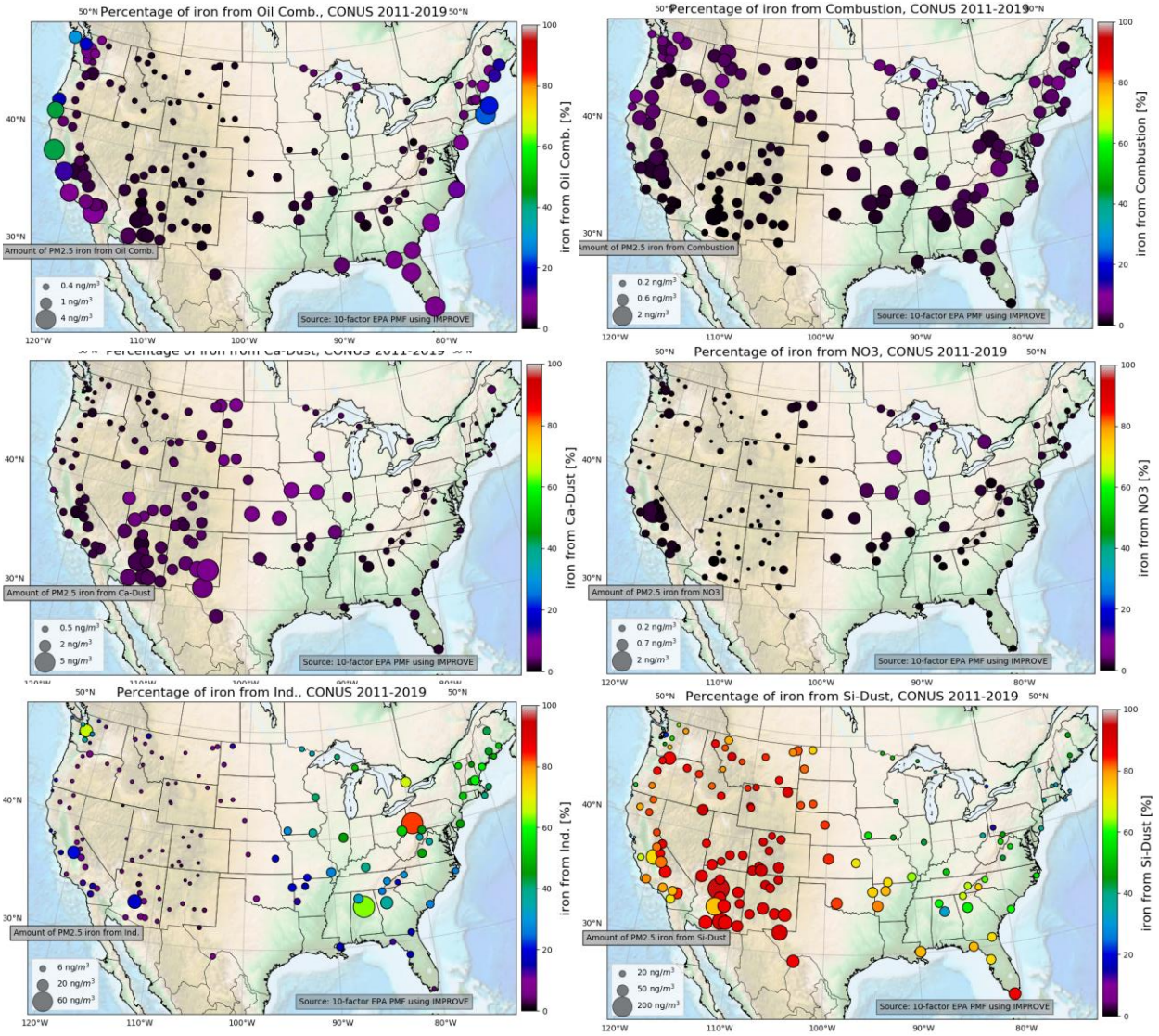

Figure 3.15. Percentage of iron from the six indicated factors: (top row) oil combustion, combustion; (middle row) Ca-Dust, NO_3; (bottom row) industry, Si-Dust. The size of the circle indicates average mass concentration of Fe in $PM_{2.5}$ at each site from each factor, while the color indicates how much of the total measured Fe was attributed to that factor.

To compute the total anthropogenically-derived Fe, we summed the Fe associated with the following factors: nitrate, oil combustion, coal combustion, and industry (Figure 3.16). With these assumptions, most of the $PM_{2.5}$ iron in the Northeast and along the West Coast was apportioned as coming from an anthropogenic source. Apart from the coast, most of the West

and Florida had little to no anthropogenic influence, reflecting the dominant contribution of dust sources to measured fine particulate Fe at those locations.

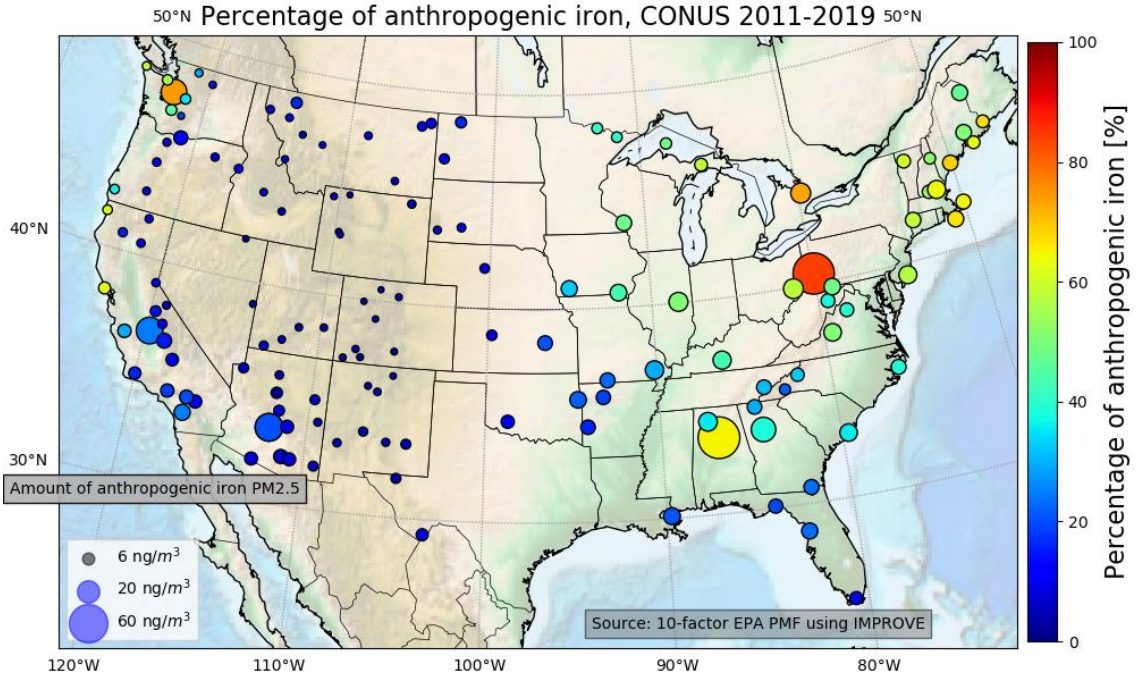

Figure 3.16. Percentage of measured fine particulate Fe from anthropogenic sources. The size of the circle indicates average amount of total anthropogenic $PM_{2.5}$ Fe at each site, while the color indicates the percentage of the total fine particulate Fe attributed by PMF to anthropogenic sources.

Soluble iron is a specific interest of this work. We made the approximation that potentially-soluble iron was equal to Fe from the combustion and oil combustion factors (Mahowald et al., 2009; Ito et al., 2019); further, since coarse mass is largely contributed by dust, we assume very little soluble Fe is missed by the lack of speciation for coarse-mode aerosols (Gao et al., 2019). Figure 3.17 shows the spatial distribution of soluble Fe under these assumptions. The highest mass concentrations are now clearly associated with coastal regions, pointing to the strong contributions from shipping (e.g., Corral et al., 2020). Along coastal regions, some sites have up to 40% of Fe in a soluble form. The majority of soluble Fe from this analysis was in the form of oil combustion iron, comprising around 3% of the total Fe across all

sites. The combustion (EC- and OC-rich) factor, which includes biomass burning, only captured 1.7% of the total Fe across all sites (Table A.1). Note that most IMPROVE sites are in the Intermountain West, which sees little oil combustion emissions. The Fe available for deposition over phytoplankton areas is likely closer to the values along coastal sites (e.g. 2-4 ng/m^3).

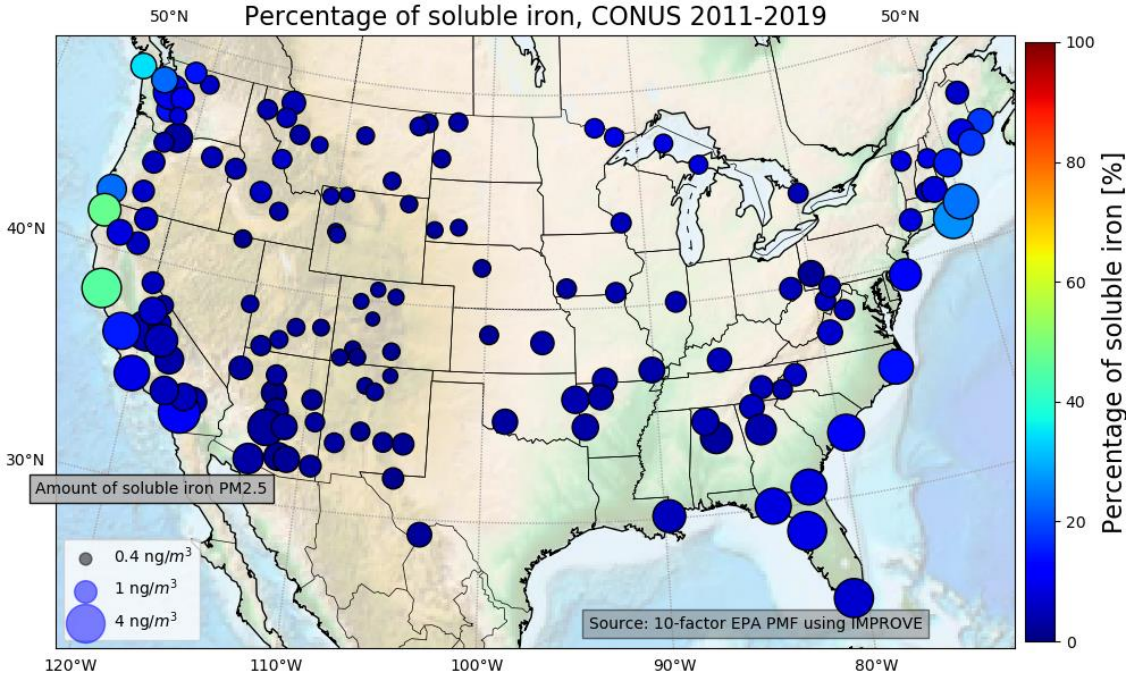

Figure 3.17. Percentage of measured fine particulate Fe from potentially soluble sources (oil combustion and "combustion"). The size of the circle indicates average amount of potentially soluble measured $PM_{2.5}$ Fe at each site, while the color indicates the percentage of the total fine particulate Fe attributed by PMF to soluble sources.

3.4 Comparison with Measured and Modeled Anthropogenic and Soluble Fe

Rathod et al. (2020) present an emissions inventory for anthropogenic combustion-derived Fe, with a specific aim to represent soluble Fe to assess deposition fluxes. In Rathod et al. (2021, *in preparation*), they expand on this work to simulate source-resolved atmospheric concentrations of Fe using global meteorology for the year 2010. Although outside our time period of 2011-2019, here we provide comparisons between our averaged observations and their model predictions. Figure 3.18 presents their predictions of the fraction of observed $PM_{2.5}$ Fe that

was attributed to anthropogenic sources. Comparing with Figure 3.16, fractions are highest in the northeastern U.S. for both studies, although Rathod et al. predicted >80% due to anthropogenic sources. The PMF predictions are generally lower, but still >50%. Figure 3.19 shows a comparison of the mass concentrations, which are within a factor of 2 over 2 orders of magnitude, again with Rathod's model having larger values than the PMF estimates. Figure 3.20 shows the mass concentrations of $PM_{2.5}$ Fe from one source, oil combustion, as estimated from the PMF results, and compares these with the CAM6 transport model simulations of Rathod et al. (2021). This source of atmospheric Fe is generally soluble but represents a minority of anthropogenic Fe.

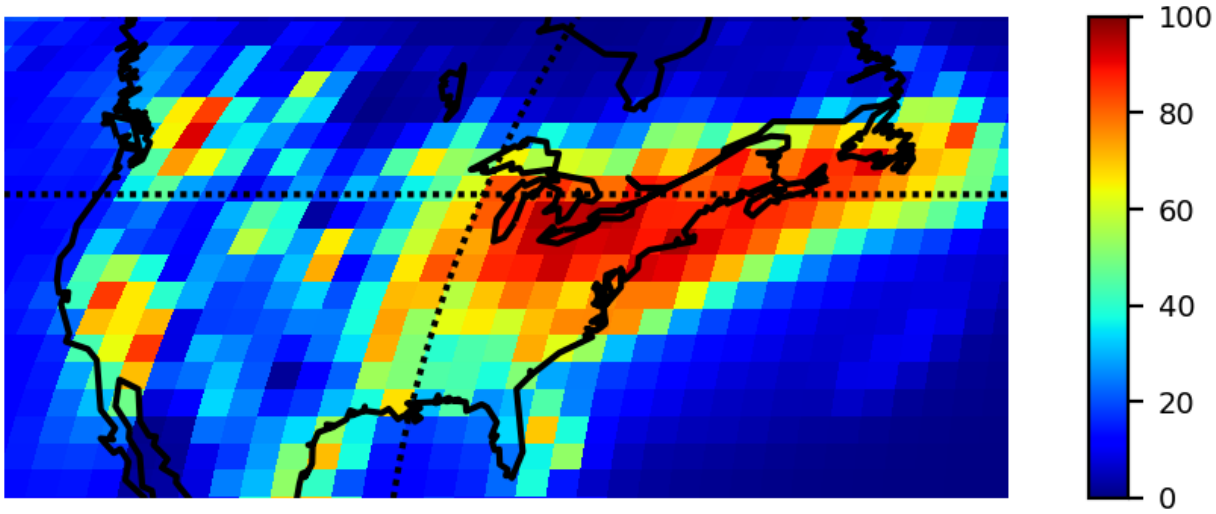

Figure 3.18. Modeled (Rathod et al., 2021) percentage of anthropogenic (defined as Total Fe – Fire Fe – Dust Fe) relative to total $PM_{2.5}$ iron. Compare with Figure 3.16, which has the same color scale.

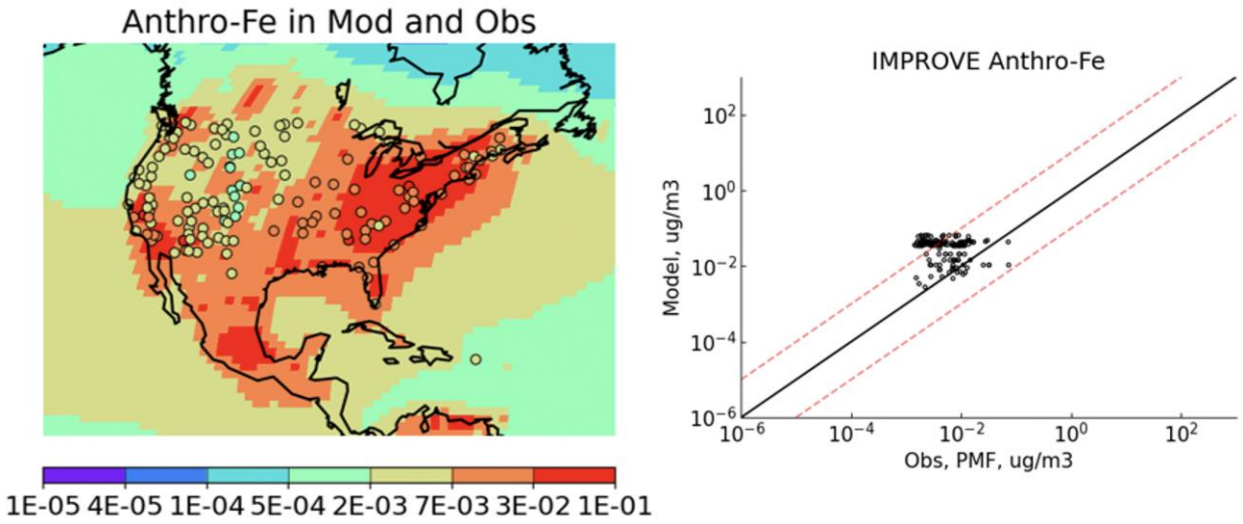

Figure 3.19a,b. Comparison between anthropogenic iron predicted by Rathod et al. (2021) and our PMF estimates (see Figure 3.17). The units are in µg/m^3.

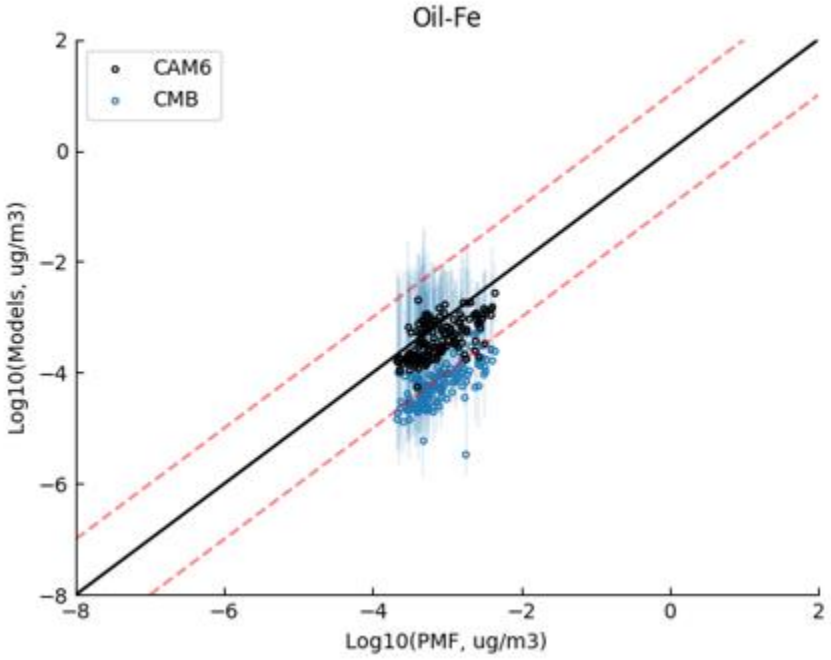

Figure 3.20. Comparison of oil combustion iron concentrations predicted by Rathod et al. (2021) (y-axis) and this PMF analysis (x-axis). The black line is a 1 to 1 comparison, and the dashed red lines are an order of magnitude difference. The black points use CAM6 atmospheric transport while the blue points use a Rathod et al's (2021) chemical mass balance (CMB) inferred oil combustion iron; CMB methods and results are not discussed herein.

3.5 Trend Analyses

Because we have 9 years of data, it was possible to perform trend analyses on some factors to gauge how certain sources of $PM_{2.5}$ pollution might be changing regionally across the U.S. We chose a Theil-Sen regressor to estimate linear trends and significance at each site, following Hand et al. (2012). Results of some of the trend analyses are shown in Figures 3.21 through 3.24, with some of the other trend figures shown in the Appendix. In these figures, the magnitude of the trend (as % change per year) is shown by the symbol color; symbols edged in white denote that the trend is not statistically significant, whereas symbols edged in black denote statistically-significant trends (p-value < 0.05).

Between 2011-2019, trends in the West for the biomass burning-related K-rich factor were positive, which would be expected due to increased wildfire activity (Figure A.4). However, as shown in Figure 3.21, no site had a statistically significant trend of an increase or decrease of combustion-derived $PM_{2.5}$, likely due to high interannual and spatial variability in fire seasons (Figure 3.22) as well as the multiple sources that the combustion factor reflects.

The bromides factor had interesting trend characteristics (Figure A.5), with decreasing values on the order of 5%/year over the Southern Plains at a 95% confidence level. Bromine is used in some agricultural pesticides, and these trends may reflect changes in application of such compounds. The trend for the industry factor (Figure A.6) was decreasing in most locations except for in the Sierras, where emissions (presumably from the San Joaquin Valley and Bay Area) caused a positive anthropogenic trace metal trend. Conversely, for the nearby Fresno site, NO_3 emissions were decreasing at a relatively high and significant rate of change. Similar downward trends in the nitrate factor were seen over the Great Plains where the highest nitrate

concentrations typically occurred (Figure A.7). There is a likely linkage to agricultural practices, but these are not explored further here.

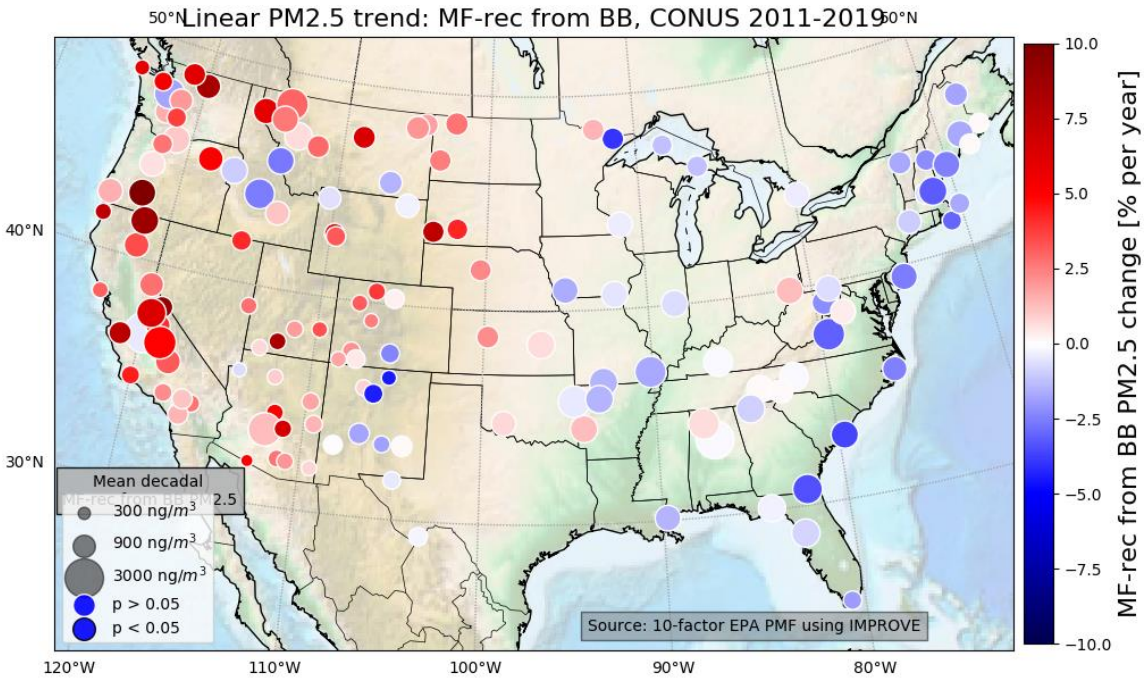

Fig 3.21. Trends in the combustion factor: $PM_{2.5}$ from combustion does not show a significant trend over any site, likely due to high inter-annual variability. Color of the circle shows linearly-modeled mass change per year from this factor, circle size indicates decadal-average mass per site from coal combustion, and the outline of the circle (white or black) indicates significance at the 95th percentile.

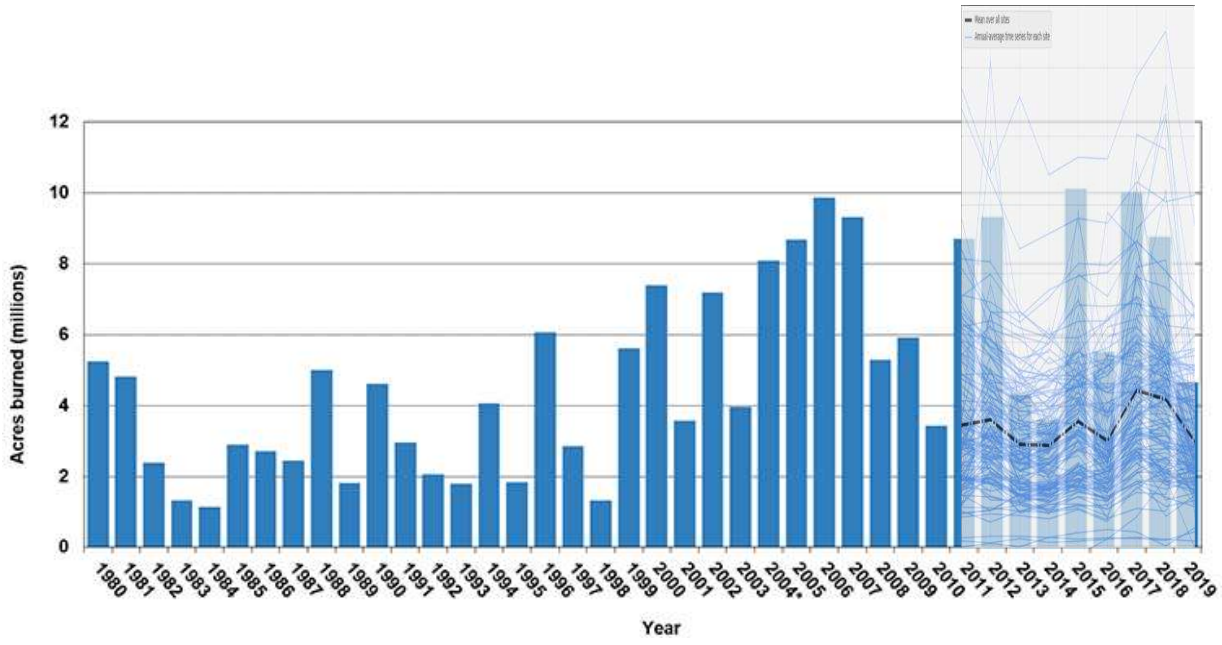

Figure 3.22. Wildfire burned acreage per year (source: National Interagency Fire Center) overlaid with $PM_{2.5}$ from the combustion factor (arbitrary scale) over all sites. The thin blue lines show the interannual variability for each site, and the black line is the mean over all sites.

The most clear and statistically-significant trends were the decreasing ones in concentrations associated with the coal combustion factor (Figure 3.23). With a decrease of nearly 10% per year over much of the Eastern U.S., every site east of 100 deg W had a decreasing trend significant at the 95% confidence level. Even many sites west of 100 deg W had a statistically significant decreasing trend in this factor. These decreases can also be seen in a plot of the annually-averaged coal combustion $PM_{2.5}$ concentration time series, constructed for all sites (Figure A.8). Hand et al. (2012) showed a similar trend in sulfates from the early 1990s through 2010; Chan et al. (2018), from 2000 through 2015; and Hand et al. (2019), from 2005 through 2016. These findings also mirror trends in coal-fired electricity generation, showing around a 3.5% nationally-averaged decrease per year (Figure A.8 and Table A.2).

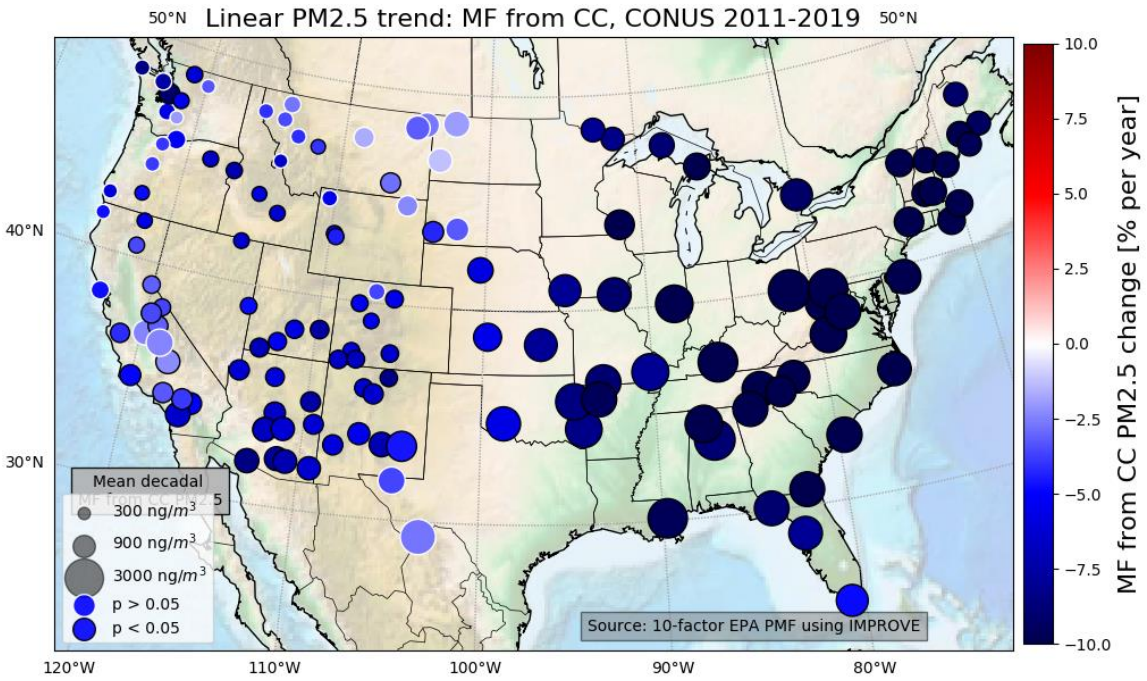

Fig 3.23. Trends in coal combustion factor. Color of the symbol shows the linearly-modeled percentage mass change per year from this factor, circle size indicates decadal-average mass per site from coal combustion, and the outline of the circle (white or black) indicates significance at the 95th percentile.

Figure 3.24 shows the trends computed for the mass concentrations associated with the oil combustion factor, which is a good proxy for soluble iron along the coasts as mentioned in Section 3.3. There were only a few sites that had statistically-significant trends, most notably, the Martha's Vineyard site in Massachusetts. This site had a trend of -2.5%/year as well as some of the largest contributions to fine mass concentrations from this factor. However, overall, coastal sites did not show statistically significant increases or decreases in fine mass from oil combustion, suggesting that soluble iron from this source also remained relatively unchanged over the time period. As discussed in Spada et al. (2018), major changes in regulations covering the types of fuels burned in shipping operations occurred in 2010 and 2015. They found large decreases in coastal V concentrations between 2011 and 2015, suggesting these were attributable to changes in bunker fuel emissions. It is interesting that our study did not find a statistically-

significant linear trend in the oil combustion factor, which constituted about 30% of the observed V concentrations (Figure 3.1). One possible explanation is that increases in shipping activity over this period, if correctly captured in our oil combustion factor, may dominate over changes in fuel composition. However, to isolate trends in shipping-related Fe, similar to those found for V, additional work is needed.

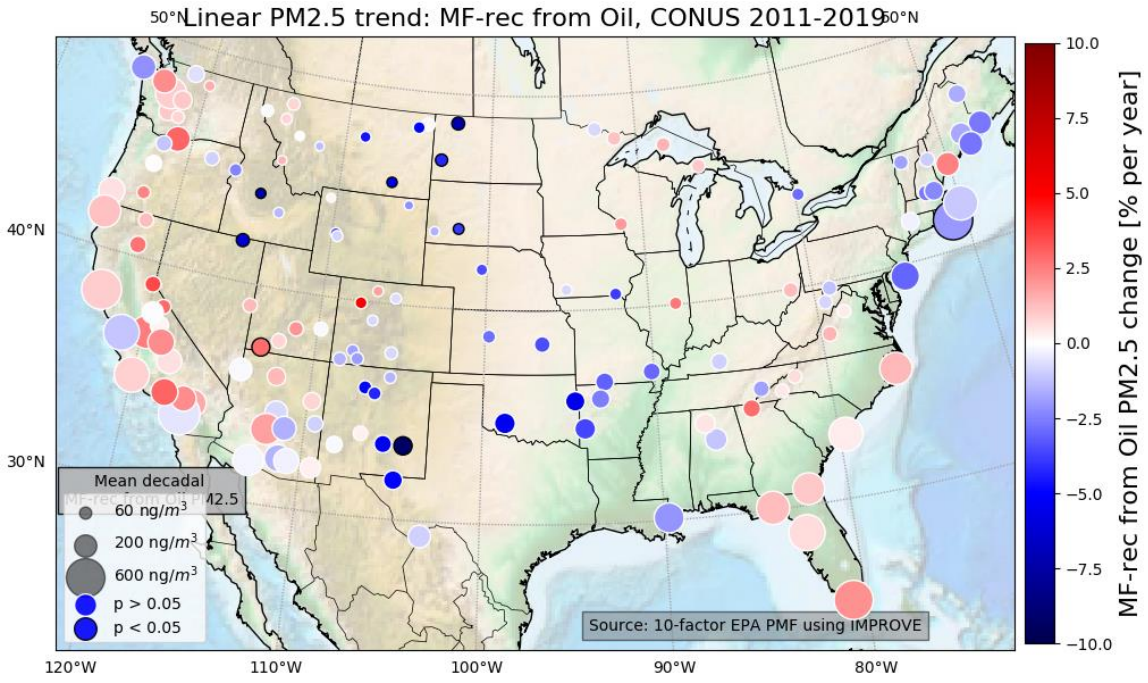

Fig 3.24. Trends in oil combustion factor. Color of the symbol shows linearly-modeled mass change per year from this factor, circle size indicates decadal-average mass concentration per site from oil combustion, and the outline of the circle (white or black) indicates significance at the 95th percentile.

CHAPTER 4: SUMMARY, CONCLUSIONS, AND FUTURE WORK

In this study, we used 2011-2019 IMPROVE data from multiple sites across the U.S. to identify a 10-factor PMF solution with physically-interpretable factors. Though some factors had mixed interpretations (e.g. the combustion factor comprising anthropogenic primary emissions, biomass burning primary emissions, and multiple secondary OC sources), most were easily identifiable as coming from a single source. Factors that included substantial amounts of Fe were primarily reflective of a single source type, except for the industrial factor, in which many anthropogenic sources (steel production, smelting, traffic, etc.) were rolled into one. An analysis with more factors could separate these anthropogenic factors further at the potential risk of overfitting.

Results were presented herein as averages over the entire study period. Additional information that may shed light on sources contributing to each factor may be gained through seasonal analyses. For example, wildfire and prescribed burns generally occur during specific seasons that vary with location across the U.S., while industry and traffic-related sources are expected to be constant. Agricultural operations that may contribute to dust emissions also follow specific seasonal patterns and might be identifiable in seasonal variability of dust factors. Further, restricting sites to non-urban locations and performing PMF on a regional rather than a national scale may refine some of the factors.

This study had a particular focus on determining sources and concentrations of soluble iron because of its effects on biogeochemistry and the carbon cycle. We were able to quantitatively show the spatial and temporal variability of soluble and anthropogenic iron through interpretation of sources associated with each factor. In addition, performing a trend

analysis on these factors proved possible; and there are other analyses that one can perform on this solution.

An interesting extension of this work is to apply a different source apportionment solution, to explore the next evolution of source apportionment. Whether it is a simple clustering algorithm such as factor analysis or principal component analysis, k-means clustering, or a non-linear model, the problem of source apportionment should not be confined to one algorithm. During this project, there were many computational limitations that had to be overcome that could be mitigated through the use of other approaches.

Extending this study worldwide would require long-term, speciated $PM_{2.5}$ datasets with point uncertainties in other regions of the world outside the United States. There are some long-term datasets in Europe and Asia that are speciated that may be applicable. An approach for source apportionment would be either to estimate uncertainties using an equation similar to that in the IMPROVE Standard Operating Procedure, if uncertainties are not provided, or to use a source apportionment technique other than PMF.

REFERENCES

Blifford, I. H., & Meeker, G. O. (1967). A factor analysis model of large scale pollution. *Atmospheric Environment (1967), 1*(2), 147-157. doi:10.1016/0004-6981(67)90042-x

Chan, E. A., Gantt, B., & McDow, S. (2018). The reduction of summer sulfate and switch from summertime to Wintertime PM2.5 concentration maxima in the United States. *Atmospheric Environment, 175*, 25-32. doi:10.1016/j.atmosenv.2017.11.055

Conway, T. M., Hamilton, D. S., Shelley, R. U., Aguilar-Islas, A. M., Landing, W. M., Mahowald, N. M., & John, S. G. (2019). Tracing and constraining anthropogenic aerosol iron fluxes to the North Atlantic Ocean using iron isotopes. *Nature Communications, 10*(1). doi:10.1038/s41467-019-10457-w

Corral, A. F., Dadashazar, H., Stahl, C., Edwards, E., Zuidema, P., & Sorooshian, A. (2020). Source apportionment of aerosol at a coastal site and relationships with precipitation chemistry: A case study over the southeast united states. *Atmosphere, 11*(11), 1212. doi:10.3390/atmos11111212

Derimian, Y., Karnieli, A., Kaufman, Y. J., Andreae, M. O., Andreae, T. W., Dubovik, O., . . . Koren, I. (2008). The role of iron and black carbon in aerosol light absorption. *Atmospheric Chemistry and Physics, 8*(13), 3623-3637. doi:10.5194/acp-8-3623-2008

Falkowski, P. (2012). Ocean science: The power of plankton. *Nature, 483*(7387). doi:10.1038/483s17a

Feigin, V. L., Roth, G. A., Naghavi, M., Parmar, P., Krishnamurthi, R., Chugh, S., . . . Forouzanfar, M. H. (2016). Global burden of stroke and risk factors in 188 countries, during 1990–2013: A systematic analysis for the Global burden of Disease Study 2013. *The Lancet Neurology, 15*(9), 913-924. doi:10.1016/s1474-4422(16)30073-4

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus. *Communications of the ACM, 24*(6), 381-395. doi:10.1145/358669.358692

Fu, H., Shang, G., Lin, J., Hu, Y., Hu, Q., Guo, L., . . . Chen, J. (2014). Fractional iron solubility of aerosol particles enhanced by biomass burning and Ship emission in Shanghai, East China. *Science of The Total Environment, 481*, 377-391. doi:10.1016/j.scitotenv.2014.01.118

Gao, Y., Marsay, C. M., Yu, S., Fan, S., Mukherjee, P., Buck, C. S., & Landing, W. M. (2019). Particle-Size variability of aerosol iron and impact on iron solubility and Dry Deposition fluxes to the Arctic Ocean. *Scientific Reports, 9*(1). doi:10.1038/s41598-019-52468-z

Hand, J. L., Schichtel, B. A., Malm, W. C., & Frank, N. H. (2013). Spatial and temporal trends in PM2.5 Organic and Elemental carbon across the United States. *Advances in Meteorology, 2013*, 1-13. doi:10.1155/2013/367674

Hand, J. L., Schichtel, B. A., Malm, W. C., & Pitchford, M. L. (2012). Particulate sulfate ion concentration And SO_2 emission trends in the United States from the early 1990s through 2010. *Atmospheric Chemistry and Physics, 12*(21), 10353-10365. doi:10.5194/acp-12-10353-2012

Hand, J., Prenni, A., Schichtel, B., Malm, W., & Chow, J. (2019). Trends in Remote PM2.5 Residual mass across the United States: Implications for Aerosol mass reconstruction in the improve network. *Atmospheric Environment, 203*, 141-152. doi:10.1016/j.atmosenv.2019.01.049

Hand, J., Schichtel, B., Malm, W., Copeland, S., Molenar, J., Frank, N., & Pitchford, M. (2014). Widespread reductions in haze across the United States from the early 1990s through 2011. *Atmospheric Environment, 94*, 671-679. doi:10.1016/j.atmosenv.2014.05.062

Hartmann, D. L. (2016). *Global physical climatology*. Amsterdam: Elsevier.

Hyslop, N. P., & White, W. H. (2009). Estimating precision using duplicate measurements. *Journal of the Air & Waste Management Association, 59*(9), 1032-1039. doi:10.3155/1047-3289.59.9.1032

Indresand, H., & Dillner, A. M. (2012). Experimental characterization of sulfur interference in IMPROVE aluminum and silicon XRF data. *Atmospheric Environment, 61*, 140-147. doi:10.1016/j.atmosenv.2012.06.079

Ito, A., Myriokefalitakis, S., Kanakidou, M., Mahowald, N. M., Scanza, R. A., Hamilton, D. S., . . . Duce, R. A. (2019). Pyrogenic iron: The missing link to high iron solubility in aerosols. *Science Advances, 5*(5). doi:10.1126/sciadv.aau7671

Jain, S., Sharma, S. K., Choudhary, N., Masiwal, R., Saxena, M., Sharma, A., . . . Sharma, C. (2017). Chemical characteristics and source apportionment of PM2.5 Using PCA/ACPS, UNMIX, and PMF at an urban site of Delhi, India. *Environmental Science and Pollution Research, 24*(17), 14637-14656. doi:10.1007/s11356-017-8925-5

Lee, E., Chan, C. K., & Paatero, P. (1999). Application of Positive matrix factorization in source apportionment of particulate pollutants in Hong Kong. *Atmospheric Environment, 33*(19), 3201-3212. doi:10.1016/s1352-2310(99)00113-2

Lee, S., Liu, W., Wang, Y., Russell, A. G., & Edgerton, E. S. (2008). Source apportionment of PM2.5: Comparing PMF and CMB results for four ambient monitoring sites in the southeastern United States. *Atmospheric Environment, 42*(18), 4126-4137. doi:10.1016/j.atmosenv.2008.01.025

Lee, T., Yu, X., Ayres, B., Kreidenweis, S., Malm, W., & Collett Jr, J. (2008). Observations of fine and coarse particle nitrate at several rural locations in the United States. *Atmospheric Environment, 42*(11), 2720-2732. doi:10.1016/j.atmosenv.2007.05.016

Li, L., Girguis, M., Lurmann, F., Pavlovic, N., McClure, C., Franklin, M., . . . Habre, R. (2020). Ensemble-based deep learning for estimating PM2.5 over California With Multisource big data including wildfire smoke. *Environment International, 145*, 106143. doi:10.1016/j.envint.2020.106143

Li, W., Xu, L., Liu, X., Zhang, J., Lin, Y., Yao, X., . . . Shi, Z. (2017). Air pollution–aerosol interactions produce more bioavailable iron for ocean ecosystems. *Science Advances, 3*(3). doi:10.1126/sciadv.1601749

Liu, J., Weng, F., & Li, Z. (2019). Satellite-based PM2.5 estimation directly from Reflectance at the top of the atmosphere using a machine learning algorithm. *Atmospheric Environment, 208*, 113-122. doi:10.1016/j.atmosenv.2019.04.002

Luo, C., Mahowald, N., Bond, T., Chuang, P. Y., Artaxo, P., Siefert, R., . . . Schauer, J. (2008). Combustion iron distribution and deposition. *Global Biogeochemical Cycles, 22*(1). doi:10.1029/2007gb002964

Mahowald, N. M., Engelstaedter, S., Luo, C., Sealy, A., Artaxo, P., Benitez-Nelson, C., . . . Siefert, R. L. (2009). Atmospheric iron Deposition: Global Distribution, variability, and Human perturbations. *Annual Review of Marine Science, 1*(1), 245-278. doi:10.1146/annurev.marine.010908.163727

Malm, W. C., Schichtel, B. A., Hand, J. L., & Collett, J. L. (2017). Concurrent temporal and spatial trends in sulfate and organic mass concentrations measured in the improve monitoring program. *Journal of Geophysical Research: Atmospheres, 122*(19). doi:10.1002/2017jd026865

Ozkaynak, H., & Thurston, G. D. (1987). Associations between 1980 u.s. mortality rates and alternative measures of Airborne Particle Concentration. *Risk Analysis, 7*(4), 449-461. doi:10.1111/j.1539-6924.1987.tb00482.x

Paatero, P. (1997). Least squares formulation of robust non-negative factor analysis. *Chemometrics and Intelligent Laboratory Systems, 37*(1), 23-35. doi:10.1016/s0169-7439(96)00044-5

Paatero, P. (1999). The multilinear engine—a table-driven, least squares program for solving multilinear problems, including then-way parallel factor analysis model. *Journal of Computational and Graphical Statistics, 8*(4), 854-888. doi:10.1080/10618600.1999.10474853

Paatero, P., & Tapper, U. (1994). Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics, 5*(2), 111-126. doi:10.1002/env.3170050203

Paatero, P., Eberly, S., Brown, S. G., & Norris, G. A. (2014). Methods for estimating uncertainty in factor analytic solutions. *Atmospheric Measurement Techniques, 7*(3), 781-797. doi:10.5194/amt-7-781-2014

Pachon, J. E., Weber, R. J., Zhang, X., Mulholland, J. A., & Russell, A. G. (2013). Revising the use of potassium (K) in the source apportionment of PM2.5. *Atmospheric Pollution Research, 4*(1), 14-21. doi:10.5094/apr.2013.002

Prenni, A. J., Day, D. E., Evanoski-Cole, A. R., Sive, B. C., Hecobian, A., Zhou, Y., . . . Schichtel, B. A. (2016). Oil and gas impacts on air quality in federal lands in the Bakken region: An overview of the Bakken air quality study and first results. *Atmospheric Chemistry and Physics, 16*(3), 1401-1416. doi:10.5194/acp-16-1401-2016

Qu, X., & Hall, A. (2007). What controls the strength of snow-albedo feedback? *Journal of Climate, 20*(15), 3971-3981. doi:10.1175/jcli4186.1

Rai, P. K. (2016). Impacts of particulate matter pollution on plants: Implications for environmental biomonitoring. *Ecotoxicology and Environmental Safety, 129*, 120-136. doi:10.1016/j.ecoenv.2016.03.012

Rathod, S. D., Hamilton, D. S., Mahowald, N. M., Klimont, Z., Corbett, J. J., & Bond, T. C. (2020). A mineralogy-based anthropogenic combustion-iron emission inventory. *Journal of Geophysical Research: Atmospheres, 125*(17). doi:10.1029/2019jd032114

Rathod, S. D., Bond, T., Pierce, J., Niño, L. W., Kreidenweis, S., ., . . . Mahowald, N. M. (2021). Constraining anthropogenic total iron emissions using models and observations. *In preparation for Journal of Geophysical Research: Atmospheres*.

Sen, P. K. (1968). Estimates of the regression coefficient based On Kendall's Tau. *Journal of the American Statistical Association, 63*(324), 1379-1389. doi:10.1080/01621459.1968.10480934

Sickles II, J. E., & Shadwick, D. S. (2015). Air quality and atmospheric deposition in the eastern Us: 20 years of change. *Atmospheric Chemistry and Physics, 15*(1), 173-197. doi:10.5194/acp-15-173-2015

Spada, N., Cheng, X., White, W., & Hyslop, N. (2018). Decreasing Vanadium Footprint of Bunker Fuel Emissions. *Environmental Science & Technology*, *52*(20), 11528–11534. doi:10.1021/acs.est.8b02942

Theil, H. (1992). A rank-invariant method of linear and polynomial regression analysis. *Advanced Studies in Theoretical and Applied Econometrics,* 345-381. doi:10.1007/97894-011-2546-8_20

Thurston, G. D., & Spengler, J. D. (1985). A quantitative assessment of source contributions to inhalable particulate matter pollution in metropolitan Boston. *Atmospheric Environment (1967), 19*(1), 9-25. doi:10.1016/0004-6981(85)90132-5

Thurston, G. D., Ito, K., & Lall, R. (2011). A source apportionment of U.S. fine particulate matter air pollution. *Atmospheric Environment, 45*(24), 3924-3936. doi:10.1016/j.atmosenv.2011.04.070

Toms, B. A., Barnes, E. A., & Ebert-Uphoff, I. (2020). Physically interpretable neural networks for the geosciences: Applications to earth system variability. *Journal of Advances in Modeling Earth Systems, 12*(9). doi:10.1029/2019ms002002

Ulbrich, I. M., Canagaratna, M. R., Zhang, Q., Worsnop, D. R., & Jimenez, J. L. (2009). Interpretation of organic components from Positive matrix factorization of Aerosol mass spectrometric data. *Atmospheric Chemistry and Physics, 9*(9), 2891-2918. doi:10.5194/acp-9-2891-2009

Viana, M., Kuhlbusch, T., Querol, X., Alastuey, A., Harrison, R., Hopke, P., . . . Hitzenberger, R. (2008). Source apportionment of particulate matter in Europe: A review of methods and results. *Journal of Aerosol Science, 39*(10), 827-849. doi:10.1016/j.jaerosci.2008.05.007

Wang, C., Tu, Y., Yu, Z., & Lu, R. (2015). PM2.5 and cardiovascular diseases in the Elderly: An overview. *International Journal of Environmental Research and Public Health, 12*(7), 8187-8197. doi:10.3390/ijerph120708187

APPENDIX

Table A.1. Factor profiles (f_{kj} matrix) for the ten-factor solution. The units are percentages of the total element. Species which were key in identifying each factor are given in **bold**. This table is shown with units of μg/m^3 in Table 3.2.

	Comb.	Second. Nitrates	Coal Comb.	Silicon-Dust	K-rich	Br-rich	Oil Comb.	Industry	Calcium-dust	Sea Salt
Al	0	0.04	2.74	**83.58**	3.94	5.32	0	0	3.06	1.33
Br	6.04	2.83	1.18	0	0	**83.69**	0	3.3	0	2.96
Ca	2.34	0.31	2.64	5.44	0.2	0	0	3.14	**82.92**	3
Chl.	0.17	1.37	0	0.18	1.13	3.98	0	2.86	0	**90.31**
Cr	1.53	2.3	8.98	**30.98**	0	8.18	5.45	**41.56**	0	1.03
Cu	0	0	0	9.71	0	16.57	6.82	**65.41**	1.5	0
Fe	1.73	0.67	0	**76.26**	0	0	2.97	15.39	2.99	0
Pb	0	3.94	21.95	0	4.68	**24.23**	0	**41.59**	2.68	0.94
Mg	0	0	0	0	0	0.86	53.39	0	28.73	17.02
Mn	4.48	0.35	2.37	**53.49**	3.34	0	1.22	23.2	11.13	0.43
$PM_{2.5}$	**41.76**	6.58	26.11	11.13	1.42	1.42	4.57	0	3.42	3.58
Ni	1.53	1	24	20.94	0.28	5.95	**22.97**	22.58	0	0.75
NO_3	1.5	**92.92**	0	0.62	0.41	0	3.66	0.9	0	0
K	8.51	0	0	0.01	**70.83**	0	9.7	2.22	7.22	1.51
Se	5.02	3.56	**60.09**	0	0	18.47	0.06	6.37	3.52	2.91
Si	1.13	0	0	**83.03**	1.97	3.2	0.05	0.1	**10.34**	0.17
Na	0	0	0	0	0	0	60.24	0	0	**39.76**
SO_4	0	2.41	**87.44**	0.76	0	0	7.6	0	1	0.79
Ti	1.98	0.06	0.56	**83.44**	2.52	1.69	3.63	6.12	0	0
Zn	0.13	5.09	1.93	0	10.57	1.86	0	**79.95**	0	0.47
V	0	0	34.1	35.91	0	1.29	**28.26**	0.45	0	0
EC	**61.96**	4.03	2.6	0	0	0	0	30.86	0.18	0.37
OC	**90.11**	0	0	2.03	0.86	1.99	5.01	0	0	0

Table A.2. Coal production and consumption statistics from the U.S. Energy Information Administration (https://www.brookings.edu/wp-content/uploads/2019/01/H.Gruenspecht_U.S.-Coal-Sector_Final_Jan_20191.pdf) showing an approximately 3.5% decrease per year in coal consumption.

Year	Coal production (million short tons)	Met coal share of production*	Export share of production	Coal consumption, total (million short tons)	Electric power sector share of consumption	Industrial sector share of consumption
2005	1,131	4.6%	4.4%	1,126	92.1%	7.4%
2006	1,163	4.3%	4.3%	1,112	92.3%	7.4%
2007	1,147	4.8%	5.2%	1,128	92.7%	7.0%
2008	1,172	5.5%	7.0%	1,121	92.9%	6.8%
2009	1,075	4.9%	5.5%	997	93.6%	6.1%
2010	1,084	7.1%	7.5%	1,049	93.0%	6.7%
2011	1,096	8.3%	9.8%	1,003	93.0%	6.7%
2012	1,016	8.9%	12.4%	889	92.6%	7.2%
2013	985	8.8%	11.9%	924	92.8%	7.0%
2014	1,000	8.1%	9.7%	918	92.8%	7.0%
2015	897	7.3%	8.2%	798	92.5%	7.3%
2016	728	7.9%	8.3%	731	92.8%	7.0%
2017	774	9.4%	12.5%	717	92.7%	7.1%

Note: *Met coal production estimated as sum of exports and domestic use.

Sources: U.S. Energy Information Administration, "Monthly Energy Review," Tables 6.1 and 6.2; U.S. Energy Information Administration, "Quarterly Coal Report October-December 2017," export data.

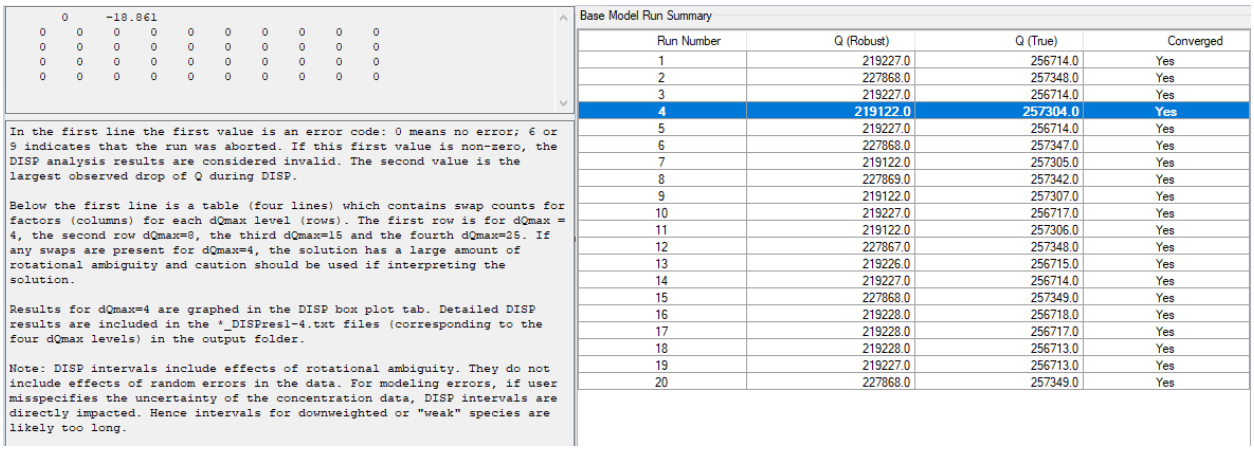

Figure A.1. (left) Displacement error estimation results (see (https://www.epa.gov/sites/production/files/2015-02/documents/pmf_5.0_user_guide.pdf) and (right) base model results for the bootstrapped 8,000-sample subset. All 20 runs converged and no factor swaps occurred during error estimation.

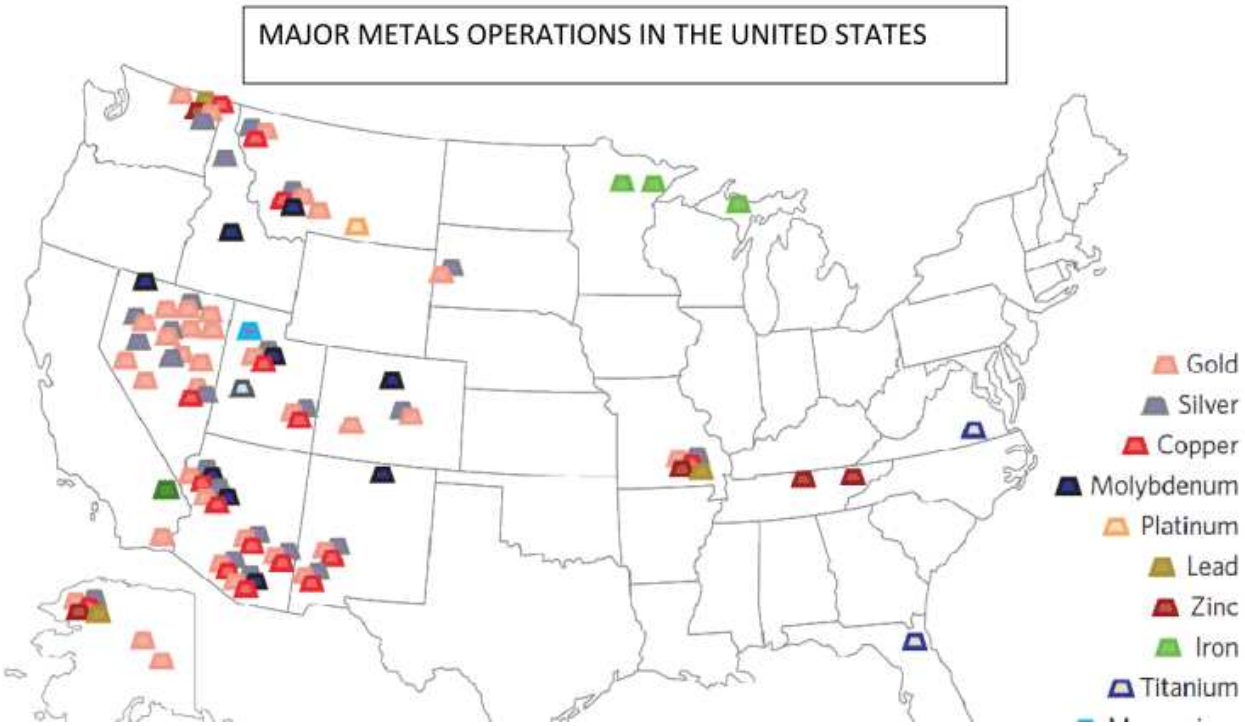

Figure A.2. Major metals operations in the United States (National Mining Association, https://nma.org/2018/09/02/map-major-metals-operations-in-the-united-states/). The Idaho molybdenum mine in question is near the IMPROVE site and may affect the iron variability. There are also two iron mine sites near the IMPROVE sites with low predictability in Minnesota.

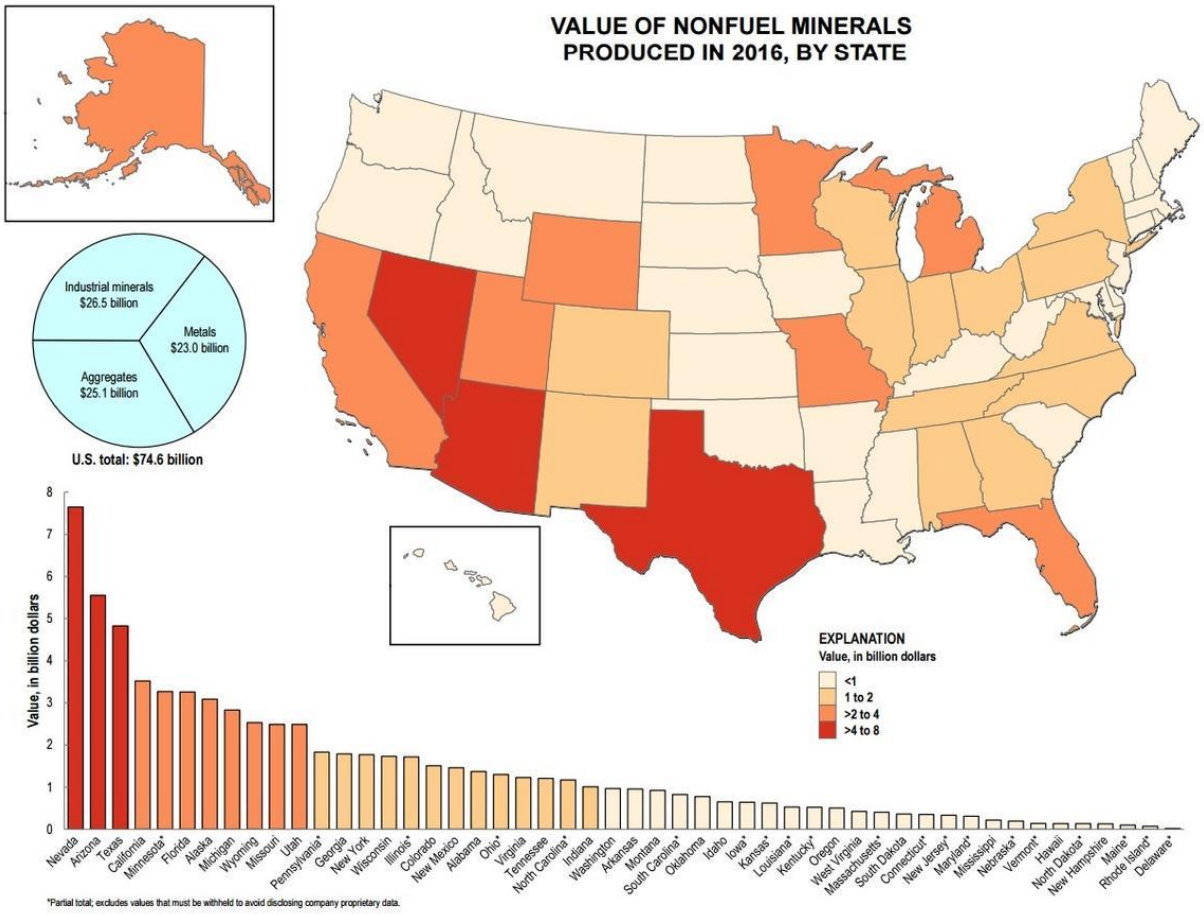

Figure A.3. Value of nonfuel minerals produced by state in 2016 (USGS, https://www.usgs.gov/news/top-5-mineral-producing-states). Minnesota and Virginia both have higher-than-average values, though the relationship between this and iron variability seems tenuous.

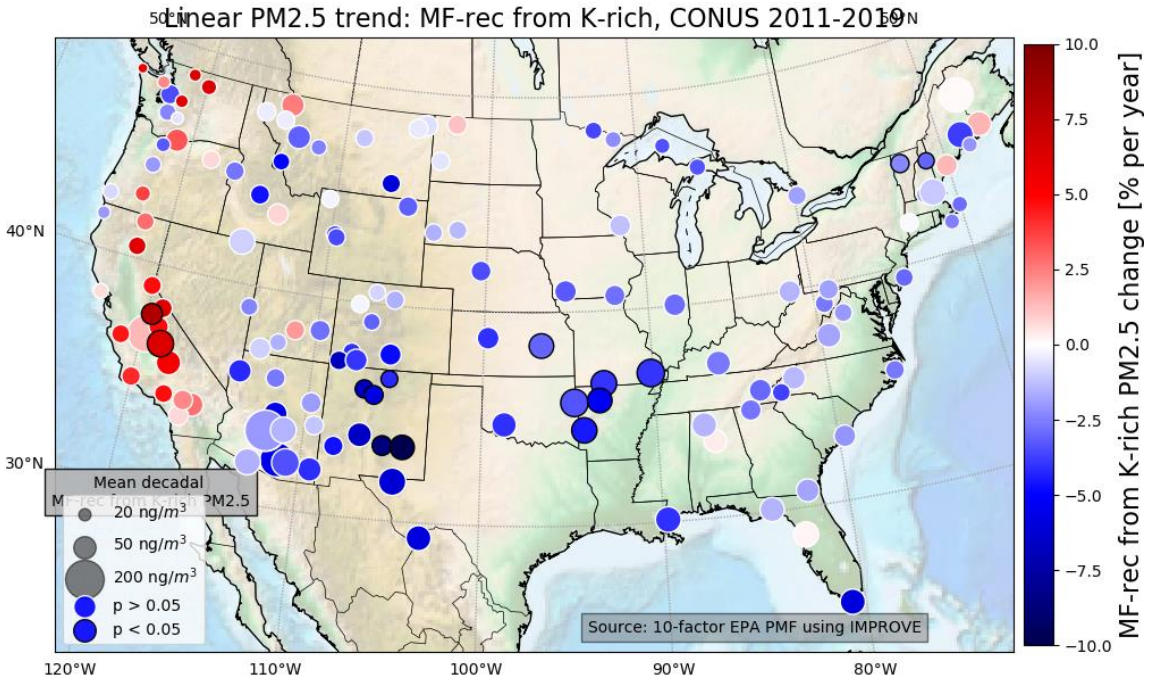

Fig A.4. Trends in the K-rich factor: $PM_{2.5}$ from this factor shows an increase over the Sierras, possibly due to increases in biomass burning, and a decrease over New Mexico, possibly due to a high fire year in 2011 (http://www.emnrd.state.nm.us/SFD/FireMgt/Historical.html). The color of the circle shows linearly-modeled mass change per year from this factor, circle size indicates decadal-average mass per site from coal combustion, and the outline of the circle (white or black) indicates significance at the 95th percentile.

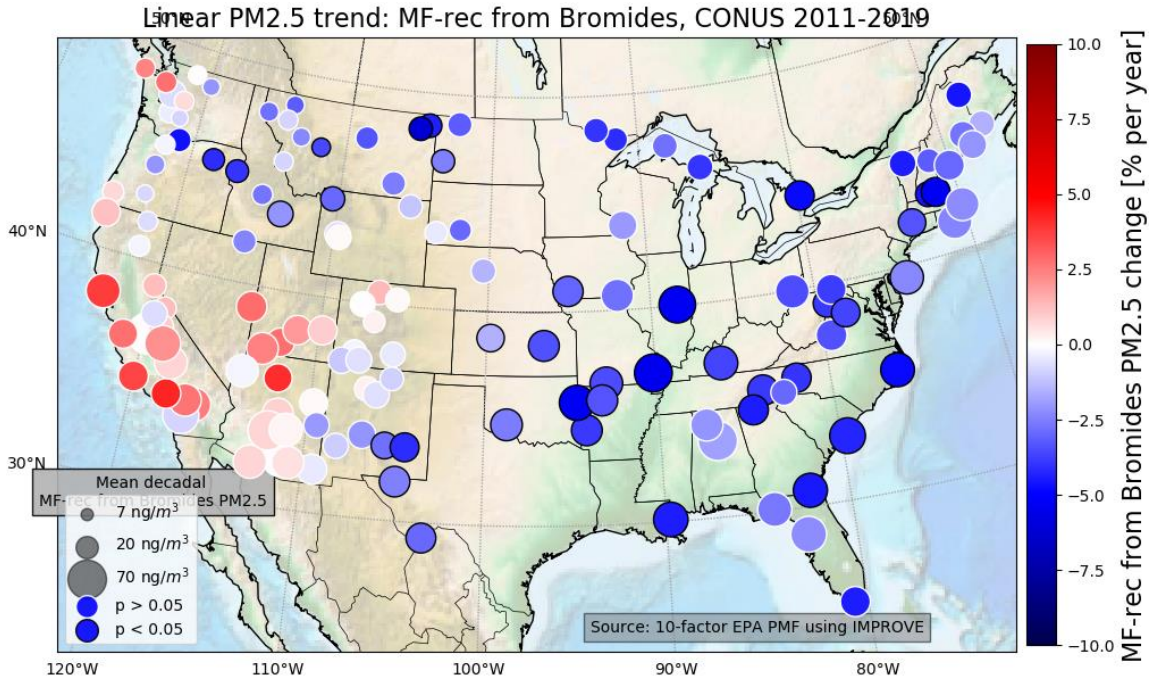

Fig A.5. Trends in the Br-rich factor: $PM_{2.5}$ from this factor shows a decrease over some of the Midwest and Plains. The color of the circle shows linearly-modeled mass change per year from this factor, circle size indicates decadal-average mass per site from coal combustion, and the outline of the circle (white or black) indicates significance at the 95th percentile.

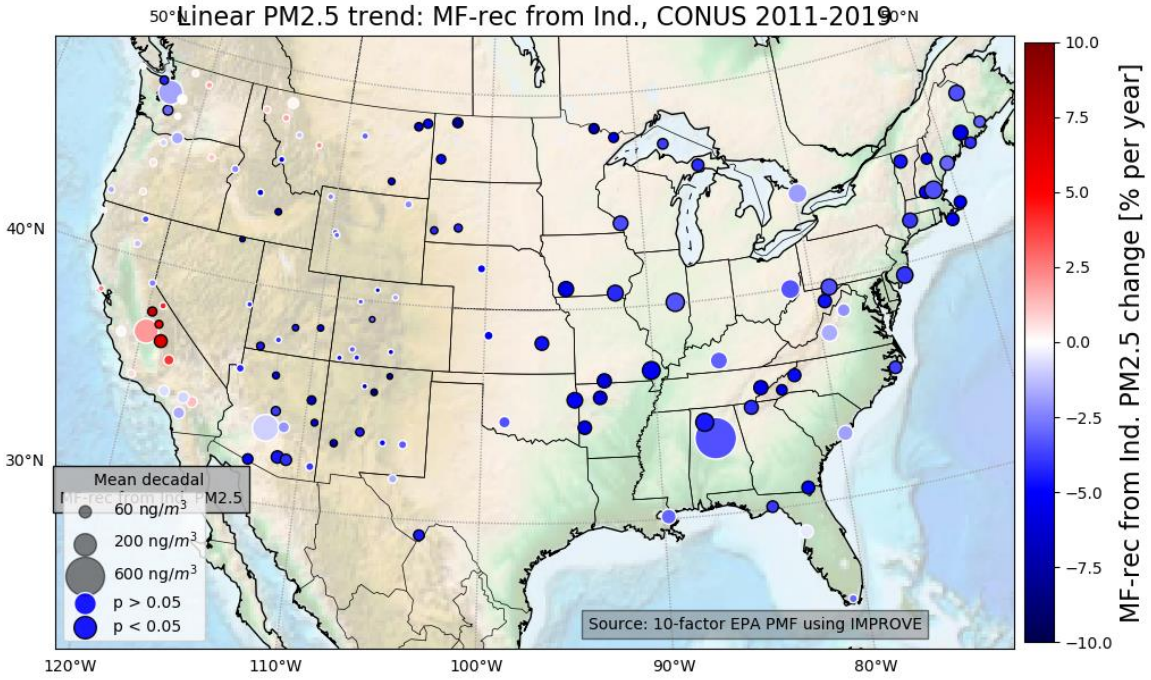

Fig A.6. Trends in the industry factor: $PM_{2.5}$ from this factor shows an increase over the Sierras and a decrease over much of the U.S. The color of the circle shows linearly-modeled mass change per year from this factor, circle size indicates decadal-average mass per site, and the outline of the circle (white or black) indicates significance at the 95th percentile.

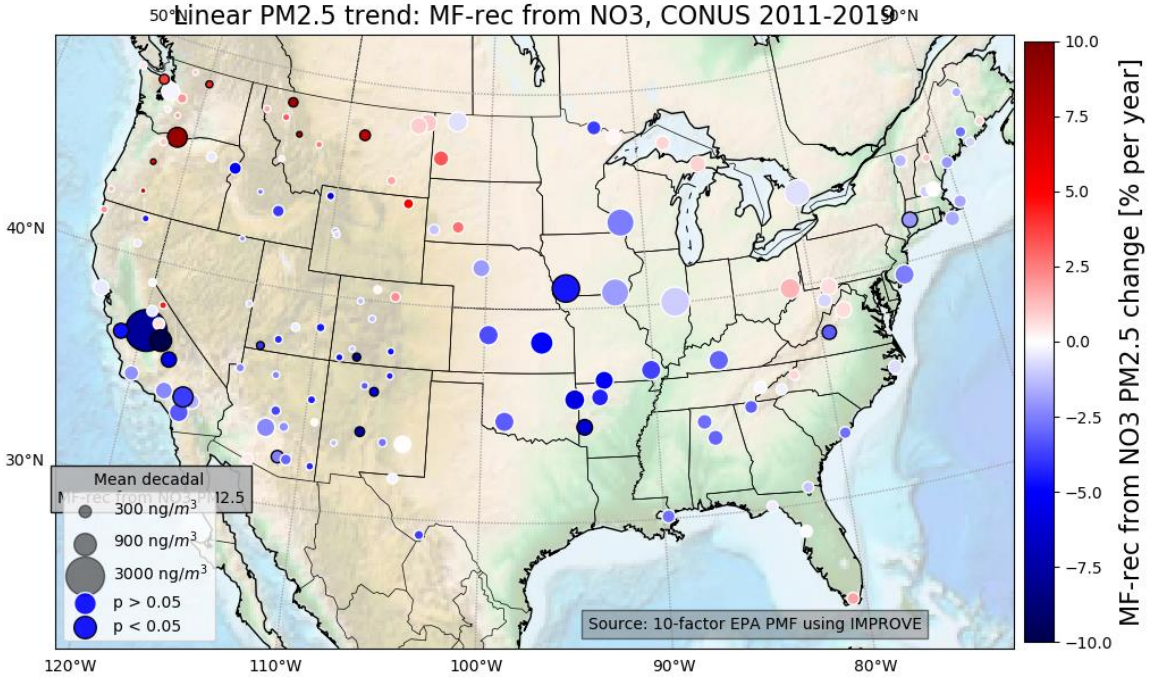

Fig A.7. Trends in the secondary nitrates factor: $PM_{2.5}$ from this factor shows an increase over parts of the Pacific Northwest and a decrease over the San Joaquin Valley and some of the Plains. The color of the circle shows linearly-modeled mass change per year from this factor, circle size indicates decadal-average mass per site from coal combustion, and the outline of the circle (white or black) indicates significance at the 95th percentile.

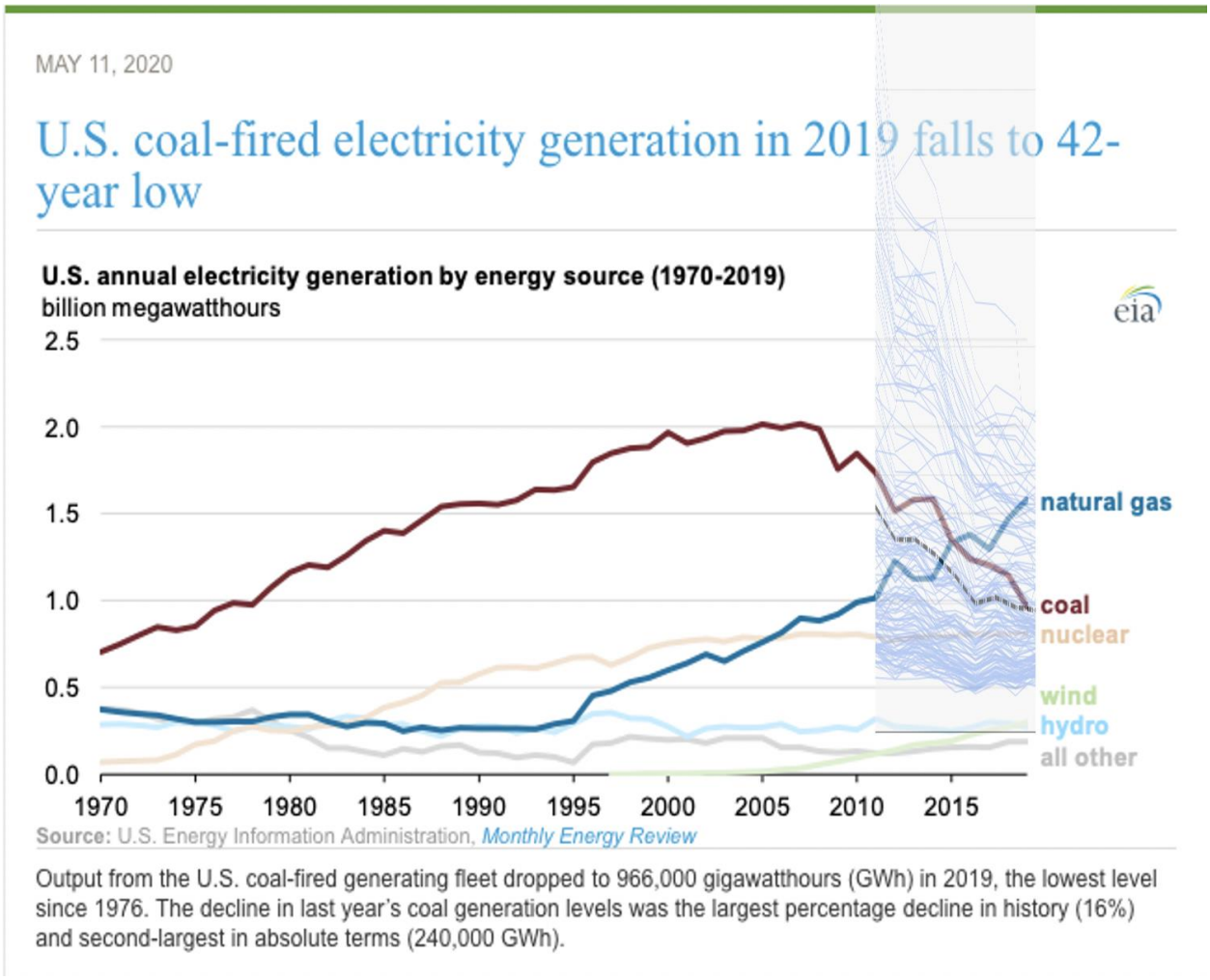

Figure A.8. U.S. coal-fired electricity generation (source: U.S. Energy Information Administration, https://www.eia.gov/todayinenergy/detail.php?id=43675) overlaid with $PM_{2.5}$ from the coal combustion factor (arbitrary scale) over all sites. The thin blue lines show the interannual variability for each site, and the black line is the mean over all sites.