

Source Identification of PM_{2.5} in Steubenville, Ohio Using a Hybrid Method for Highly Time-resolved Data

Ram Vedantham¹, Matthew S. Landis^{1*}, David Olson¹, Patrick Pancras²

¹United States Environmental Protection Agency, Office of Research and Development, RTP, NC 27709

² Alion Science and Technology, P.O. Box 12313, RTP, NC 27709

*Corresponding Author: Tel.: (919) 541-4841, landis.matthew@epa.gov

Abstract

A new source-type identification method, Reduction and Species Clustering Using Episodes (ReSCUE), was developed to exploit the temporal synchronicity typically observed between ambient species in high time resolution fine particulate matter (PM_{2.5}) data to form clusters that vary together. High time-resolution (30 min) PM_{2.5} sampling was conducted for a month during the summer of 2006 in Steubenville, OH, an EPA designated non-attainment area for the U.S. National Ambient Air Quality Standards (NAAQS). When the data were evaluated, the species clusters from ReSCUE matched extremely well with the source types identified by EPA Unmix demonstrating that ReSCUE is a valuable tool in identifying source types. Results from EPA Unmix show that contributions to PM_{2.5} are mostly from iron/steel manufacturing (36% ± 9%), crustal matter (33% ± 11%), coal combustion (11% ± 19%) and residual oil burning (15% ± 12%). More importantly, ReSCUE was useful in (i) providing objective data driven guidance for the number of source factors and key fitting species for EPA Unmix, and (ii) detecting tenuous associations between some species and source types in the results derived by EPA Unmix.

Introduction

The availability of continuous or semi-continuous instrumentation has been a significant technological advance in the measurement of air pollutants (1,2). Highly time-resolved measurements have been reported for a wide range of PM_{2.5} components including mass (3, 4, 5, 6), organic and elemental carbon (7, 8), and inorganic compounds (9,10, 11, 12, 13, 14). Data from these instruments offer significant advantages over traditional 24-h integrated filter-based measurements. First, highly time-resolved data can be used to both spatially and temporally isolate impacts from specific point sources, especially when coupled with meteorological data of similar time scales. Secondly, sources can be identified and quantified based on known activity patterns (e.g., vehicular traffic), and recent health studies have reported acute outcomes at time scales finer than traditional 24-h measurement durations (15,16).

Despite advances in instrumentation resulting in time-resolved data sets with low associated uncertainties, techniques have not been specifically developed to process these content-rich data. Traditional multivariate source-apportionment models use techniques such as the gradient method (EPA PMF) or the Singular Value Decomposition method (EPA Unmix) to develop species groupings or fingerprints. While such models do utilize underlying data characteristics, they do not take advantage of some other features typically

present in highly time-resolved data such as serial correlation. Thus, these traditional methods not only fail to take advantage of existing data features but the user implicitly assumes features in their data set (e.g., profile invariance) that may not be true all the time. Additional challenges include selection of appropriate input species that adequately characterize the contributors to the air shed, identifying the number of factors (source types) to model, and interpretation of the resulting source types.

The most common technique to using traditional receptor models is an iterative approach whereby modelers utilize various analysis approaches to find the best combination of species and the maximal number of expected or interpretable source types. ReSCUE can reduce this iterative process while providing an independent evaluation of the input data set, and is also designed to mine the rich content of a highly time-resolved data set. ReSCUE analyzes the data for temporal patterns and creates clusters of species that represent source types or a combination of sources with similar composition, temporal relationships, and relative location compared to the sampling site. ReSCUE can also be used to evaluate source apportionment assumptions such as profile invariance, made by the traditional receptor models (17,18,19). The number of distinct ReSCUE clusters can be viewed as the expected number of sources and the species can be viewed as input data for traditional receptor models. Although EPA Unmix (hereinafter, Unmix) can provide guidance on possible number of recoverable source profile signals using the NUMFACT algorithm (20), it is highly sensitive to the combination of observations and species. In Unmix, a small set of observations and/or the selected species can disproportionately affect the result. In general, these issues are magnified with highly time-resolved data due to transient high concentration source impacts and other periods of no observable source impact. ReSCUE is an independent approach that also helps receptor modelers to (i) select appropriate species from a data set, (ii) choose the number of source types, (iii) independently investigate source type groupings without assuming multivariate analysis requirements, and (iv) receive independent confirmation of source profiles.

This manuscript (i) introduces the ReSCUE algorithm and presents its application to a high time resolution $PM_{2.5}$ data set collected in the Ohio River Valley, USA, (ii) compares the ReSCUE results with those from a traditional receptor model (Unmix), (iii) introduces *sector* apportionment model to evaluate model results by identification of likely location of known source types, and (iv) discusses high time resolution data issues related to the scope and applicability of inverse models like ReSCUE and Unmix.

Site Description

In the summer of 2006, 30-minute measurements were made at the Steubenville, OH monitoring site located on the Franciscan University Campus (40° 22' 45.69 N 80° 37' 10.29 W) overlooking the Ohio River. The industrially impacted site is located in an area with a high density of large atmospheric emission sources such as coal-fired power plants, iron/steel production, and other general industrial activity (Figure 1). Previous long-term source-apportionment studies of wet deposition (21,22) and $PM_{2.5}$ (20,23,24) in Steubenville found significant impacts from coal combustion, oil combustion, iron/steel production, and windblown dust.

Sample Collection and Analysis

An EPA modified Semi-Continuous Elements in Aerosol Sampler-III (SEAS-III; OEI, Clarksville, MD) was set up and operated at the Steubenville, OH site between July 12 and August 17, 2006 during an EPA measurement intensive field experiment. The SEAS-III collects 30 min integrated $PM_{2.5}$ samples as a liquid slurry (suspension of $PM_{2.5}$ in water) for off-line chemical analysis. Operation of the SEAS-III instrument, subsequent analysis of samples, and data QA/QC is discussed by Pancras and Landis (25). Briefly, the SEAS-III was configured with a $PM_{2.5}$ cyclone inlet and sampled at a flow rate of 92 l min^{-1} . Direct steam injection is used to nucleate aerosols that are subsequently collected by direct impingement. Samples were extracted with concentrated ultrapure nitric acid (to 0.2% v/v), sonicated for 2 h, and left to leach for 30 days prior to analysis. The samples were analyzed for a comprehensive suite of trace elements using a ThermoFinnigan Element2 (Bremen, Germany) high-resolution magnetic sector field inductively coupled plasma mass spectrometer (HR-ICPMS) housed in a class 100 clean laboratory at the EPA facility in Research Triangle Park, NC. A total of 1453 valid 30 min samples were used in this analysis. The final species matrix included 24 trace elements, 2 criteria gases, and $PM_{2.5}$ mass (Table 1). Meteorological data including wind direction and wind speed (10 m above ground level) were measured using RM Young (Traverse City, MI) instrumentation. Ambient concentrations of NO_x and SO_2 , were continuously measured at 1 min intervals using ThermoEnvironmental (Franklin, MA) models 42CTL and 43CTL, respectively. $PM_{2.5}$ mass was continuously measured using a Rupprecht & Patashnick Model 1400b Tapered Element Oscillating Microbalance (TEOM) at a 1 min resolution. A summary of the observed criteria gas and $PM_{2.5}$ species concentrations, and the SRM recoveries are presented in Supporting Information (SI) Table 1 and SI Table 2, respectively.

ReSCUE Algorithm

ReSCUE analyzes data to identify time series synchronicity and uses it to create intra-species clusters. The annotated ReSCUE algorithm code is provided in the Supporting Information and consists of the following steps:

- Step 1. Episodes are identified for each species separately. An episode is identified as a set of temporally ordered values that starts and stays consistently above the base value followed by an eventual decrease to the base value. Due to the consistency requirement, excursions above the base value alone may not be enough to be categorized as an episode (Figure 2). For this study, the 5th percentile value was assigned to be the base value. Also, only those data identified as being part of an episode will be retained for further analysis. For instance, only 46% of Fe was flagged as episodic in nature (Figure 2).
- Step 2. Pair-wise Pearson correlations between species during identified episodes, Episodic Species Correlations ($ESC_{(X,Y)}$), are calculated using Equation 1. This measures the synchronicity between any two species during episodes and does not include any contributions from non-episodic intervals. If X_i and Y_i were species values associated with two species X and Y , with their associated respective means \bar{X} and \bar{Y} , then $ESC_{(X,Y)}$ is defined as:

$$ESC_{(X,Y)} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{[\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2]^{1/2}} \quad (1)$$

Step 3. The fraction of observations over which any two species share their episodes, Episodic Intersection Fraction (EIF), is then calculated using Equation 2. For each species, the rows associated with the episodes are assigned a value of 1 and the non-episodic rows are assigned a zero value. This is called the indicator function associated with the criteria. A dot product of the indicator functions associated with two species divided by the total number of observations provides the fraction of episodic overlap.

$$EIF_{(X,Y)} = \frac{(I_X \cdot I_Y)}{\# \text{ of observations}} \quad (2)$$

where $I_{()}$ denotes the indicator function associated the chosen species and parenthesis in the numerator indicates the dot product.

Step 4. Episodic Species Correlations are weighted by their Episodic Intersection Fraction using Equation 3, and the resulting value is a Weighted Episodic Species Correlation ($WESC_{(X,Y)}$) Value with $EIF_{(X,Y)}$ as weights.

$$WESC_{(X,Y)} = EIF_{(X,Y)} \cdot ESC_{(X,Y)} \quad (3)$$

Step 5. For each species (X), its associated species are categorized using their WESC values as follows: “Very Strong” (values > 0.85), “Strong” (0.75-0.85), and “Moderate” (0.60-0.75). These descriptions are meant to capture the strength of association between species.

Step 6. Finally, for each species, the associated species cluster is formed using two steps:

- I. First, an initial cluster is formed by combining the “Very Strong”, “Strong”, and the “Moderate” species associated with the species using the criteria stated in Step 5.
- II. Next, the species in the “Very Strong” category (from the previous step) are re-considered in turn. For each of those species, the species that are in its “Strong” and “Moderate” categories are also combined with the initial cluster of the chosen species. For these secondary associations, the cutoffs are more stringent: “Strong” (>0.80) and “Moderate” (> 0.70).

A comparison of Steubenville $PM_{2.5}$ ReSCUE clusters (episodic data only) to the clusters formed using simple correlations (all data) is presented in SI Table 3, and shows that only ReSCUE clusters closely resemble emission profiles of known source types (13) in the region.

Results obtained using ReSCUE

A total of 27 species (Table 1) were used in the ReSCUE analysis. As previously described, the ReSCUE algorithm creates clusters of species that are episodically correlated. Species in the “Very Strong” and “Strong” categories of a cluster are boldfaced in Table 1. Source type associations were identified by looking for the presence and absence of known tracer species. For each species, its associated species are listed in the order of the strength of their association. For example, Cr is strongly associated with the base species Ni while Fe, Ba, K, and Mo are progressively less associated. Trace metals such as Ce, La are categorized as multi-source components due to their (“Moderate”) associations with many other source-defining species. Conversely, Cd is not strongly associated with any species and is categorized as a sole species source type. In some cases, two species may be associated in a species cluster but also be exclusively associated with other clusters. For instance, Fe and Ge are temporally correlated enough to be clustered. But, there were also enough periods when there were Fe or Ge episodes, but not both (SI Figure 1). This implies that Fe and Ge have both common *and* distinct sources. In this way ReSCUE clusters can be identified as either distinctive emission source types such as local/regional coal combustion (presence of S, Se, SO₂), long range transported coal-combustion (presence of S and Se and the absence of SO₂), or as part of a composite coal combustion emission source.

Results obtained from Unmix and comparison with ReSCUE Results

Unmix is one of the most widely used receptor modeling methods. The current version of Unmix (v 6.0) uses measured data only and does not require any *a priori* knowledge of sources or their emission characteristics. The algorithm assumes that, for every identifiable source type, there are at least a few observations that contribute insignificantly to that source and uses them to construct the edges that define the region of feasible solutions (26). Periods of absence (or near absence) of at least one source are more likely in a highly time-resolved data set. Therefore, Unmix has been found to be highly suitable for identifying and apportioning the PM_{2.5} mass (12).

The same high time resolution PM_{2.5} species and criteria gases that were screened and used in the ReSCUE analysis were also utilized by Unmix. Seven profiles resulting from using Unmix were identified to be those that are typically found in the Steubenville air shed (Table 2). The last column in Table 1 shows the closest associated Unmix source with each ReSCUE cluster. This association was created by comparing the strong species from ReSCUE that are usually used by receptor modelers to identify source types. These clusters fall in two broad categories: closest matching Unmix source and multi-source component.

Profile1 has the highest loading of Zn and smaller portions of Rb, Mg, Mn, Fe and K. This matches quite well with ReSCUE’s Zn cluster: **Rb, Mn, K, Fe, Mg**, Ge, NO_x, La, Ce. Roughly 80% of the apportioned Zn is associated with this profile and is most likely associated with iron/steel manufacturing. The basic oxygen furnace (BOF) and the Electric Arc Furnace (EAF) usage in steel

production is a significant source of Zn. The BOF flue dust is known to contain high levels (1.5%-4.0%) of Zn and the EAF contains even higher amounts (15% - 25% of the resulting flue dust) of Zn (27, 28). There are multiple steel manufacturing facilities (Figure 1) within 30 km of Steubenville producing Zn-coated (galvanized) steel. Zinc can also be emitted from iron/steel production facilities charging recycled scrap steel into the BOF (29) and steel galvanizing operations. Manganese is also used in ferroalloy production and steel coating operations which accounts for the strong association of Zn with Mn. Connell et al. (23) found a similar Zn-Mn dominated factor when apportioning PM_{2.5} in Steubenville using 24-h integrated samples and attributed the source to local metal related industries. The US EPA's 2001 Toxic Release Inventory (TRI) reported that over 90% of Zn emissions and over 80% of Mn emissions are attributable to steel production near Steubenville.

Profile 2 is dominated by Ni and Cr with small contributions from Al and Fe. ReSCUE did not associate Al with this cluster. The corresponding ReSCUE cluster is the Cr cluster: **Ni**, Fe, Mo, V, Ba, La. Thus, it appears that Ni/Cr is almost a complete source cluster by itself due to the unique nature of Cr and Ni time series. There are two matching episodes (both Cr and Ni exceeding 15 µg m⁻³) at the beginning of the intensive followed by no discernible activity simultaneous or otherwise. This profile may be attributable to steel processing industries and their associated use of Ni and Cr in ferroalloys production and plating in and around Steubenville. In addition, austenitic stainless steels (over 70% of total stainless steel production) have a typical composition of 18% Cr and 10% Ni (30).

Over 75% of Cd is present in Profile 3 along with smaller loadings of Pb and K. ReSCUE, essentially identifies Cd as a single species cluster. The presence of Pb in this source type is not confirmed by ReSCUE. It is likely that Unmix edges (resulting in the Pb loading) were disproportionately influenced by a small set of co-varying values (see SI Figure 2). Thus, it may be appropriate to interpret this profile as a Cd-only source type. The major anthropogenic sources of atmospheric Cd emission (31) are both ferrous production (e.g., iron/steel manufacturing) and non-ferrous metal operations (e.g., Zn smelting). Other sources include coal combustion and residual heavy fuel oil combustion. The time series shows Steubenville being affected with Cd peaks from both the northeast and the south. SI Figure 2 highlights the wind events from the northeast (0° – 90°) and the south (120° – 240°) as being significant contributors of Cd. Hence, it is most likely that the iron/steel production facilities and the power plants from the south may be the primary sources of the Cd found in Steubenville. A thin coating (<25 µm) of Cd is usually applied to iron and steel to withstand atmospheric corrosion (32).

The strong presence of Se and SO₂ in Profile 4 suggests a local/regional coal combustion source (33, 34). ReSCUE's Se cluster (**SO₂**, S, PM_{2.5}) appears to be the closest to this profile. Again, Pb is associated with this source by Unmix (and not by ReSCUE). The ubiquitous nature of Pb emitted by a variety of combustion sources and manufacturing processes in and around Steubenville is most

likely the reason for the smeared presence of Pb in a majority of the profiles generated by Unmix for this data set. The time series data shows that winds from the south accounted for almost all of Se and SO₂ measured at the site. One of the largest coal-fired power plants in the US (Cardinal Power) is located in Brilliant, OH less than 16 km south of the receptor site. While PM_{2.5} mass and S track each other well, SO₂ peaks appear have a corresponding S peak only when the wind comes from the south. Since S is emitted from fossil fuel combustion primarily as SO₂ which is subsequently oxidized to particulate SO₄⁻ in the atmosphere, the presence of simultaneous SO₂ and Se peaks indicates a local/regional coal combustion source. The observation of simultaneous S and Se peaks without SO₂ would be indicative of aged or transported coal combustion source(s). In fact, there are quite a few S peaks for which the corresponding SO₂ values are not significant (SI Figure 3). This weak association between SO₂ and S could be the reason why Unmix did not associate S with this profile.

Profile 5 is the source with the least amount signal amplitude, likely representing mixed industrial emissions from the local air shed yielding an often present moderate source signal. This profile is a diverse mix of metals associated with iron/steel production. The ReSCUE cluster that appears closest to this profile is the Fe cluster (**Ge, Rb, K, Mn, Mg, Zn**, La, Cu, Ce, Cr, Ni, NO_x, Ba, S, Sr, Pb, As). Unmix attributes over 70% of Ge to this profile along with elevated loadings of Fe, Cu, Mn, Rb, Pb, and Mg that matches ReSCUE results. This is remarkable considering the different numerical approaches that the Unmix (whole data set approach) and ReSCUE (pair-wise episodic weighted correlation approach) models utilize, which may not highlight the same underlying data features. Along with NO_x, this profile aptly characterizes the expected emissions from integrated iron/steel production and slag/sintering operations (35, 13). Profile 6 is crustal in nature with substantial loadings of Al, Sr, Mg, S and PM_{2.5} mass. The closest matching ReSCUE cluster is the Sr cluster: **Mg, Mn, Al, Ce, Ba, Fe, Ge, Rb, Zn, K, NO_x**. The other cluster that resembles this source type from Unmix is the Al cluster: Sr, Mg. But all associations in this cluster are weak. However, even in Sr cluster, ReSCUE seems to have been influenced by species such as Fe, Ge, and Rb that appear to be ubiquitous in Steubenville air shed.

The strong presence of V, Ti, Mo, As, Cr and NO_x in Profile 7 is indicative of iron/steel production or heavy residual oil burning. The closest ReSCUE cluster is the V cluster: Ti, As, Cr, Mo, NO_x. Note that the associations are all moderate. Even in Unmix, this source type is calculated as the last source due to low signal strength. About 90% of worldwide V production is used in iron/steel production due to its high tensile strength, hardness, and fatigue resistance (36, 37). Thus, Profile 7 may be attributed to multiple iron/steel production facilities or residual heavy fuel oil combustion near Steubenville (38).

Identification of Sources Using Sector Apportionment Methods

One way to evaluate receptor modeling results is to investigate the link between estimated source contribution estimates and the physical location of known sources in the area. But, complications in associating physical sources to identified source types include

co-linearity of sources, diffused sources, topographic impediments (e.g., mountains), and changes to emission profiles due to photochemical reactions. Nevertheless, for many source types identified using highly time-resolved data, it is possible to identify the origin of the identified source type using additional information such as meteorology (14).

The Sustained Wind Incidence Method (SWIM) is a spatial apportionment model that highlights regions of highest contribution using measured concentrations (or predicted source contributions) and meteorological data (39). The method uses a kernel smoothing approach to highlight directional domains of significant contributions. Then, by utilizing resources like the EPA Toxics Release Inventory (TRI), specific sources may be identified. The SWIM model is a directional apportionment tool and is not capable of determining the distance. If the directional domain suggested by SWIM is devoid of a known local (< 30 km) source(s), then it is likely that the source type is primarily non-local. However, it is unlikely that both local and non-local sources will be identified by receptor models together from a highly time-resolved data due to distinct variance structures (important to factor analytic models) of local and non-local source types. Nevertheless, source types that contain both local and non-local contributions may still be identified. For instance, measured Se may be a combination of near field and regional power plant emissions. In such cases, it would be a mistake to apportion all the Se contributions only to the local source. Additional analysis (e.g., Wavelet) must be done to be able to find the appropriate share of the local source.

Using the SWIM model on the Steubenville receptor modeling results, all source types identified by ReSCUE/Unmix were confirmed (Figure 3a-d). For instance, Figure 3b shows that almost all the high values of the source type 1 (Zn-source) are from the south whereas the source type 4 in Figure 3c have contributors in both in the southwestern and northeastern sector. The lack of any contribution from the northwestern sector is due to lack of wind events from that sector (Figure 3a). The Ohio River, just east of the receptor site, runs from north-northeast to south-southwest direction. The river valley channels the wind to be predominantly from either from the southwest or the northeast.

Implications from using ReSCUE on highly time resolved data

The fact that nearly identical results for major source types were derived using the two unrelated numerical approaches of ReSCUE and Unmix (once the species selection and number of factors were optimized using ReSCUE) strengthens the results from both models. The results from any source apportionment analysis gain credibility through weight of evidence provided by the application of multiple receptor modeling tools. When Unmix produces a model solution, it is nearly impossible to know if those solutions were influenced by a representative subset of the data. Bootstrapping resampling tools are available to evaluate the robustness of the factors, but having an independent, data value based model such as ReSCUE is a valuable interpretive tool. The ReSCUE method is especially relevant to highly time-resolved measurements, where comparatively short signal or plume impact periods can be

numerically obscured by relatively more numerous and noisy baseline periods by whole data set models like Unmix. The disadvantage of ReSCUE is that the results are qualitative and it does not provide quantitative source contributions.

Limitations of traditional source-apportionment methods, including an assumption of source invariance and source interpretation, can be evaluated by ReSCUE by exploiting the autocorrelation structure typically present in highly time-resolved environmental data. One example from this study was the obfuscating effect of Pb in the Steubenville air shed, an effect revealed only by examining the ReSCUE clusters. Designations of weighted species correlations (e.g., “Strong,” “Moderate”) can further refine source interpretations, and clusters with multiple species associations likely indicate ubiquitous or non-unique sources. In addition, the very nature of time-resolved data calls into question the traditional definition of a source. The sporadic nature of some source types identified in this study (e.g., the Ni-Cr source) is likely the result of temporal changes to industrial emissions (e.g., batch processes, raw materials, fuels). This is especially relevant in areas with a variety of industrial sources (e.g., Steubenville), where multiple profiles could be attributable to the same physical source.

As previously discussed, ReSCUE helped elucidate the Unmix “smeared” apportionment of Pb. Another instance of differences between Unmix and ReSCUE is the trace metal Sm. Unmix associates Sm with iron/steel production and crustal source types whereas ReSCUE does not associate Sm with any major species. There are two potential explanations for this observation. First, as a mathematical model, Unmix tries to associate every input species with some source type whether or not the alignment is natural. If the signal strength were strong, Sm would have been identified as its own source type. However, Sm was smeared across most Unmix source types. From the air pollution emission source perspective, the most likely explanation is that Sm may be a part of the raw material processed to obtain iron (e.g., taconite), resulting in a weak association with the crustal source as part of wind-blown dust. Irrespective of the reasons, Sm should not be interpreted with confidence. Thus, use of both ReSCUE and Unmix can prevent over-interpretation of trace metals with low signal strengths.

An additional advantage of ReSCUE is its ability to test scenarios that may be unrealizable using another traditional multivariate receptor model. For instance, it is desirable to help associate P to appropriate sources but Unmix was unable to find a feasible solution when P was added to the species list. By using ReSCUE however, it is possible to test this scenario. ReSCUE associated P with the crustal source along with Al, Sr, and Mg. This grouping can be strongly identified with windblown dust with P most likely associated with farm applications as fertilizer. Thus, ReSCUE can be used in situations where the inclusion of species in Unmix is precluded. Perhaps one of the more compelling reasons to use ReSCUE is its ability to identify more factors or source types than Unmix by using all of the available episode data to resolve the source solution. Clearly in a complex air shed like Steubenville there are more than the seven sources types that were identified by Unmix contributing to the measured PM_{2.5} mass. ReSCUE can resolve additional sources

such as local/regional coal combustion versus long range transported coal combustion emissions that may further investigated by methods such as SWIM and are of importance to the community. The ReSCUE algorithm has been found to be a useful data screening and receptor modeling tool, and as a result EPA is planning to incorporate it into the next EPA Unmix model release.

Disclaimer

The United States Environmental Protection Agency (EPA) through its Office of Research and Development funded, managed, and participated in the research described here under contract EP-D-05-065 with Alion Science and Technology. The views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of EPA. It has been subjected to Agency's administrative review and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Acknowledgement

The authors wish to thank the anonymous peer reviewers whose insightful comments and suggestions helped improve this manuscript and Vasu Kilaru (EPA) for creating the overlay of the receptor site and surrounding sources.

Supporting Information Available: This information is available free of charge via the Internet at <http://pubs.acs.org/>.

Mass Rank	Species	Associated Species	Source Type Association	Associated Unmix Source
1	PM _{2.5}	S , Se	Coal Combustion	4
2	S	PM_{2.5} , Se, SO ₂ , Fe	Coal Combustion	4
3	K	Rb, Mg, Fe, Zn, Mn, La , Ce, Ge, Cu, Cd, NO _x , As, Ni, Ba,	Multi-source Component	
4	Zn	Rb, Mn, K, Fe, Mg , Ge, NO _x , La, Ce,	Zinc	1
5	Al	Sr, Mg	Crustal	6
6	Fe	Ge, Rb, K, Mn, Mg, Zn , La, Cu, Ce, Cr, Ni, NO _x , Ba, S, Sr, Pb, As	Iron/Steel Industries	5
7	Mg	Mn, Sr, Rb, K, Fe, Zn , Ge, Ce, NO _x , Ba, La, Al	Multi-source Component	
8	NO _x	La, Ce, Rb, Mn, Ge, Mg , As, K, Zn, Fe, Ti, Ba, V	Multi-source Component	
9	Pb	Cu,SO ₂ ,Fe	Secondary emission	
10	SO ₂	Se , S, Pb	Coal Combustion	4
11	Mn	Rb, Mg, Ge, Zn, NO_x,K, Fe , Ce, La, Sr	Multi-source Component	
12	Cu	Fe, K, Ge, Pb, Ce, La, Rb	Iron/Steel Industries	5
13	Ba	Fe, NO _x , Mg, Cr, Ni, Sr, K, Cd, Rb	Iron/Steel Industries	5
14	Se	SO₂,S,PM Fine	Coal Combustion	4
15	As	La, NO _x , Rb, Ge, Ce, V, K, Cd, Fe	Multi-source Component	5
16	V	Ti, As, Cr, Mo, NO _x	Oil Combustion	7
17	Mo	Cr, V, Ti	Oil Combustion	7
18	Ti	V, NO _x , Mo	Oil Combustion	7
19	Sr	Mg , Mn, Al, Ce, Ba, Fe, Ge, Rb, Zn, K, NO _x	Crustal	6
20	Cd	K, As, Ba	Cadmium	3
21	Ni	Cr , Fe, K, Ba, Mo	Nickel/Chromium	2
22	Cr	Ni , Fe, Mo, V, Ba, La	Nickel/Chromium	2
23	Ge	Rb, Mn, Fe, NO_x, Ce, La , Mg, Zn, K, Cu, As, Sr	Multi-source Component	
24	Rb	Zn, Mn, K, Ge, Mg, La, NO_x, Fe, Ce , As, Cu, Ba, Cd	Multi-source Component	
25	Ce	La, NO_x, Rb, Ge , Fe, K, Mg, Mn, As, Cu, Zn, Sr, Sm	Multi-source Component	
26	La	Ce, NO_x, Rb, Ge , K, Fe, As, Mn, Cu, Zn, Mg, Cr	Multi-source Component	
27	Sm	Ce	Trace component	

Table 1: Summary of ReSCUE results. NOTE: Strong associations (WESC values >0.75) are boldfaced.

Source Types	Source 1 Zn	Source 2 Ni/Cr	Source 3 Cd	Source 4 Coal Combustion	Source 5 Iron/Steel	Source 6 Crustal	Source 7 Ti/V/Mo
Mg	2.25 ± 0.29	1.3 ± 0.294	1.2 ± 0.334	0.0295 ± 0.26	5.32 ± 0.473	8.66 ± 0.687	3.25 ± 0.348
Al		3.97 ± 1.01		2.73 ± 0.756		30.2 ± 1.88	
S		123 ± 44.7		532 ± 95.3	1160 ± 110	1000 ± 110	171 ± 53.2
K	7.54 ± 0.623	3.5 ± 0.645	11.4 ± 1.85		11.2 ± 1.65	12 ± 1.58	3.82 ± 1.7
Ti	0.0261 ± 0.00958		0.0391 ± 0.0132	0.0535 ± 0.0124	0.197 ± 0.0232		0.651 ± 0.0355
V			0.319 ± 0.0318	0.119 ± 0.0265	0.24 ± 0.0491		1.06 ± 0.0551
Cr	0.0286 ± 0.00967	0.232 ± 0.0142		0.0137 ± 0.00874	0.11 ± 0.0341		0.317 ± 0.0345
Mn	0.705 ± 0.057	0.108 ± 0.0419		0.0173 ± 0.0559	1.41 ± 0.0779	0.893 ± 0.094	0.721 ± 0.075
Fe	3.8 ± 0.348	2.72 ± 0.321		1.28 ± 0.358	10.3 ± 0.657	6.5 ± 0.675	2.18 ± 0.486
Ni		0.442 ± 0.0179	0.0755 ± 0.0502			0.114 ± 0.0356	
Cu	0.0292 ± 0.0366	0.16 ± 0.0431	0.573 ± 0.0645	0.243 ± 0.0501	1.46 ± 0.109	0.497 ± 0.0765	
Zn	36.8 ± 1.5		3.51 ± 0.632		4.04 ± 0.91		
Ge	0.0414 ± 0.0052		0.00709 ± 0.00871		0.205 ± 0.0102		0.0289 ± 0.0097
As	0.0834 ± 0.0155		0.281 ± 0.0218		0.431 ± 0.0345		0.643 ± 0.0363
Se	0.097 ± 0.0449			1.39 ± 0.261	0.536 ± 0.158	0.252 ± 0.168	0.157 ± 0.0643
Rb	0.0439 ± 0.00257	0.00242 ± 0.00176	0.0143 ± 0.00286		0.0649 ± 0.00395	0.0204 ± 0.00376	0.0215 ± 0.00324
Sr	0.00224 ± 0.00698	0.0312 ± 0.00729	0.0304 ± 0.00675	0.0203 ± 0.00636	0.118 ± 0.0125	0.221 ± 0.0177	0.133 ± 0.0102
Mo	0.0174 ± 0.0235		0.107 ± 0.0495	0.0843 ± 0.0259	0.16 ± 0.0554		1.44 ± 0.0831
Cd	0.0415 ± 0.0111	0.0108 ± 0.0145	0.495 ± 0.0367				0.0797 ± 0.0303
Ba	0.0475 ± 0.0257	0.162 ± 0.0262	0.27 ± 0.0503	0.168 ± 0.0321	0.641 ± 0.0682	0.566 ± 0.0756	0.642 ± 0.0535
La	0.00594 ± 0.00108	0.00133 ± 0.000898	0.00803 ± 0.000973	0.00705 ± 0.0013	0.0303 ± 0.0023	0.0163 ± 0.0023	0.0378 ± 0.00219
Ce	0.00526 ± 0.00132	0.00367 ± 0.00121	0.00792 ± 0.00122	0.0086 ± 0.00153	0.0351 ± 0.00285	0.0298 ± 0.00326	0.0413 ± 0.00242
Sm	0.00123 ± 0.000244	0.000179 ± 0.000179	0.000244 ± 0.000299	0.00137 ± 0.000348	0.00332 ± 0.000523	0.00305 ± 0.000478	0.0101 ± 0.00053
Pb	0.244 ± 0.0921		1.56 ± 0.3	1.51 ± 0.267	1.71 ± 0.352	0.813 ± 0.308	
SO ₂	0.0604 ± 0.223			4.12 ± 0.961	0.267 ± 0.699		0.433 ± 0.194
NO _x	1.51 ± 0.187		0.445 ± 0.308	0.43 ± 0.179	5.17 ± 0.361		7.16 ± 0.476
PM _{2.5} Mass		1.18 ± 0.36		3 ± 0.57	9.49 ± 0.836	8.56 ± 0.923	4.03 ± 0.501
Strong species	Zn, Rb	Ni, Cr	Cd, Pb	SO ₂ , Se, Pb	Fe, PM Fine, Cu, S, NO _x , Mn, Ge, Rb, As, Pb	Al, Mg, Sr, S, PM Fine	V, Ti, Cr, Mo, As, NO _x

Table 2: Seven-source solution from Unmix ($\mu\text{g m}^{-3}$). Bold numbers signify source defining species, blank cells indicated non-significant contributions, and uncertainties were calculated using the Unmix bootstrap function.

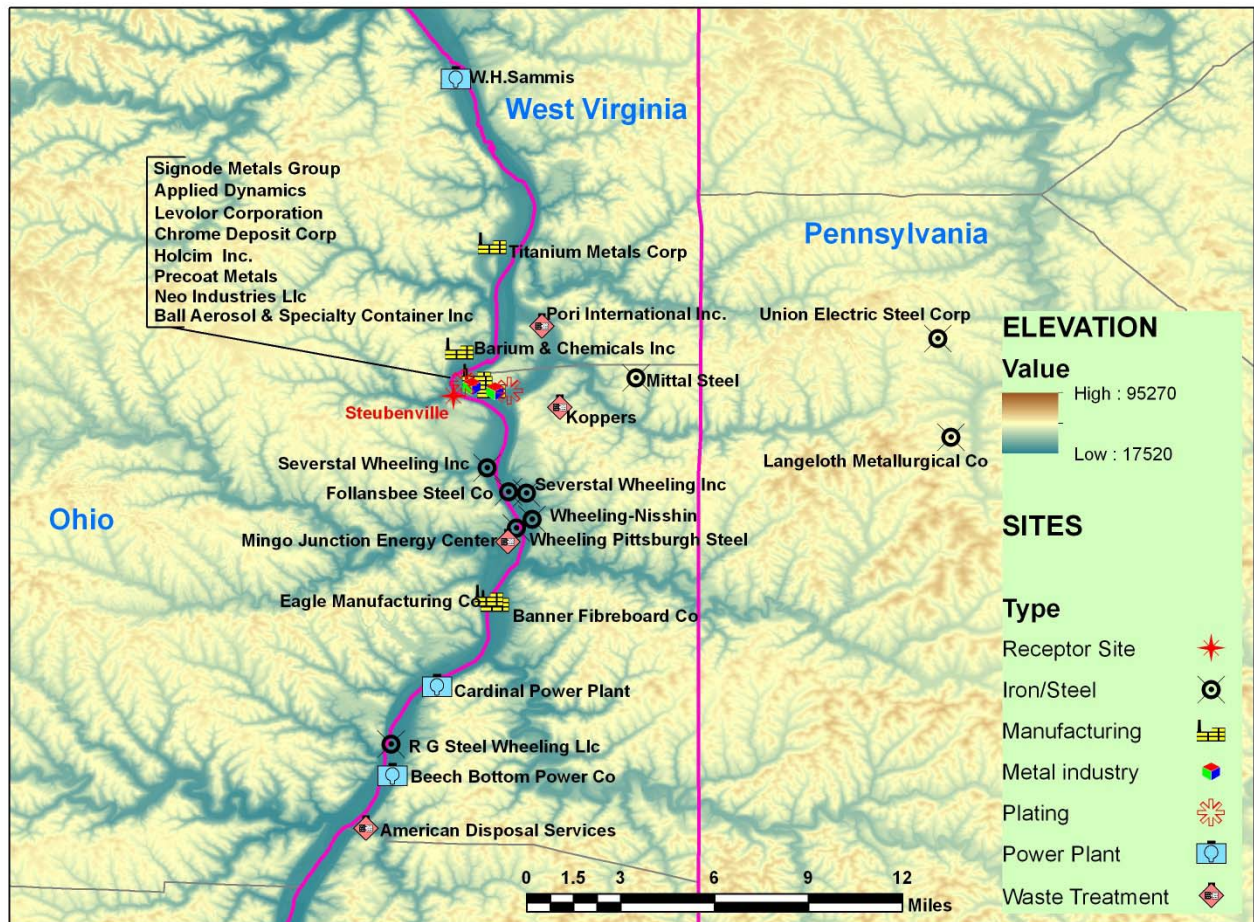


Figure 1: Steubenville site (marked in red) in the Ohio River Valley and surrounding industrial sources.

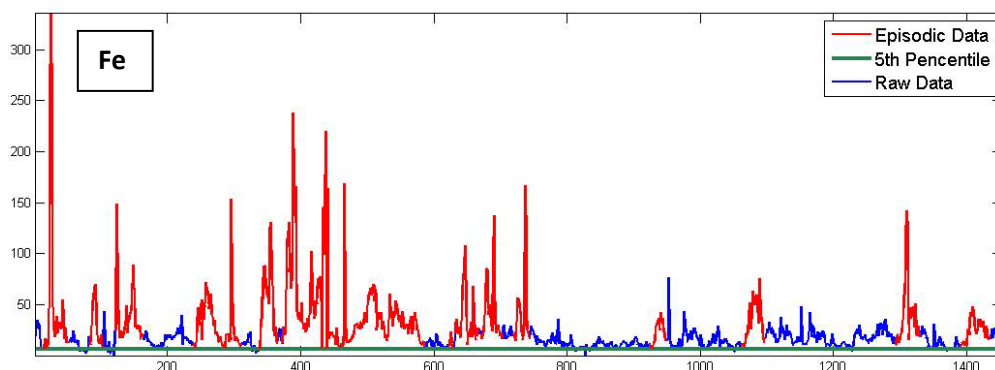


Figure 2: Time series plot of all 30 m SEAS-III PM_{2.5} Fe results (ppb), the red overlay indicates the ReSCUE identified "episodes".

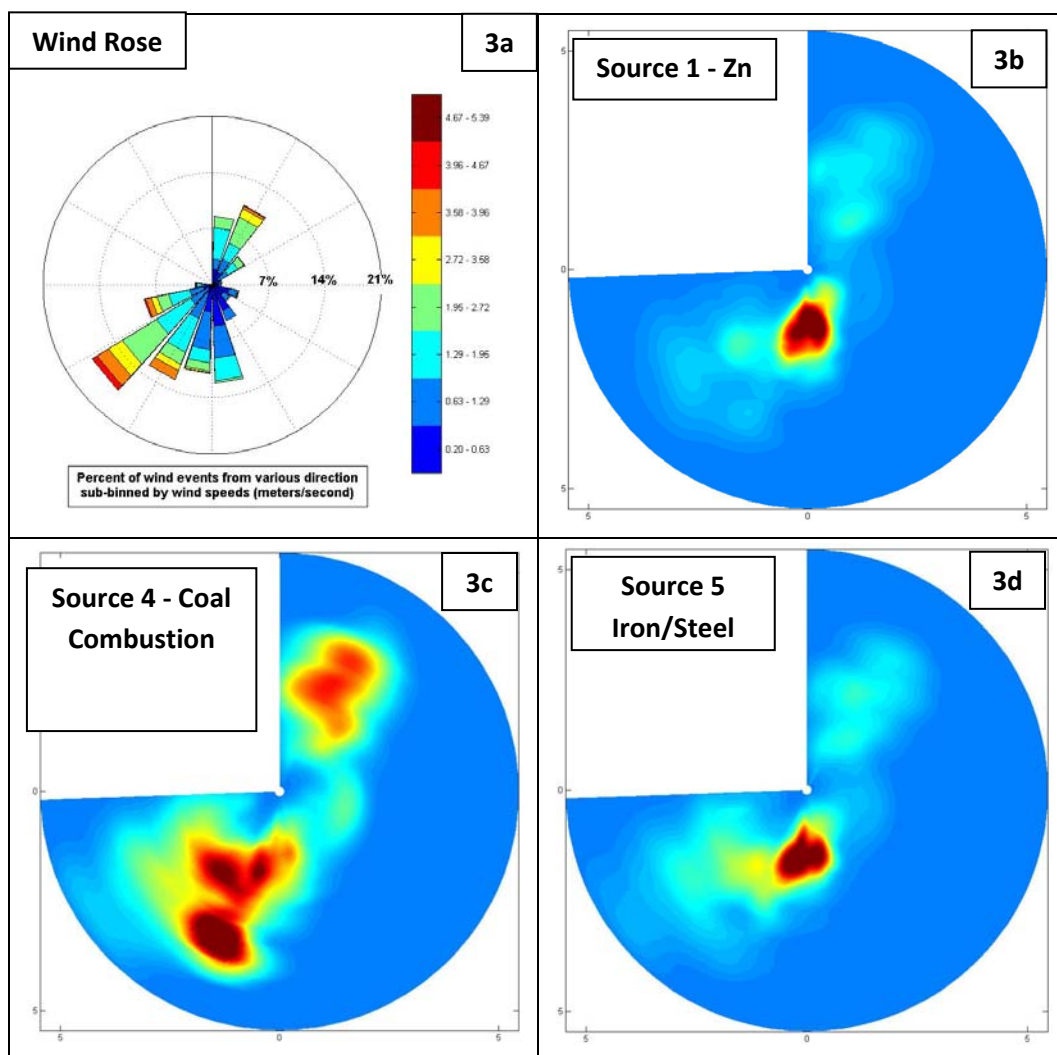


Figure 3: SWIM model spatial probability results for Unmix source contribution factors. Concentration units are $\mu\text{g m}^{-3}$ and wind speed is m s^{-1} . Source 1 values range from $0 \mu\text{g m}^{-3}$ (blue) to $0.0018 \mu\text{g m}^{-3}$ (red). Source 4 values range from $0 \mu\text{g m}^{-3}$ (blue) to $0.0008 \mu\text{g m}^{-3}$ (red). Source 5 values range $0 \mu\text{g m}^{-3}$ (blue) to $0.015 \mu\text{g m}^{-3}$ (red).

- (1) Wexler, A.S., M.V. Johnston, (2008). What have we learned from highly time resolved measurements during the EPA supersite program and related studies? *J. AWMA* **58**, 303-319, 2008.
- (2) Dreyfus, M.A., Adou, K., Zucker, S.M., Johnson, M.V. (2009). Organic aerosol source apportionment from highly time resolved molecular composition measurements. *Atmos. Environ.* **43**, 2901-2910.
- (3) Chuersuwan N., Turpin B.J., and Pietarinen C. (2000) Evaluation of time-resolved PM_{2.5} data in urban/suburban areas of New Jersey. *J. AWMA* **50**, 1780-1789.
- (4) Weber, R., Orsini, D., Bergin, M., Kiang, C.S., Chang, M., John, J.S., Carrico, C.M., Lee, Y.N., Dasgupta, P., Slanina, J., Turpin, B., Edgerton, E., Hering, S., Allen, G., Solomon, P., Chameides, W., 2003a. Short-term temporal variation in PM_{2.5} mass and chemical composition during the Atlanta supersite experiment. *J. AWMA* **53**, 84-91.
- (5) Park et al., 2006 Park, S. S.; Pancras, J. P.; Ondov, J. M.; Poor, N.; Robinson, A. Application of the pseudo-deterministic receptor model to resolve power plant influences on air quality in Pittsburgh. *Aerosol Sci. Technol.* 2006, **40**, 883-897.
- (6) Solomon, P.A.; Sioutas, C. 2008. Continuous and Semicontinuous Monitoring Techniques for Particulate Matter Mass and Chemical Components: A Synthesis of Findings from EPA's Particulate Matter Supersites Program and Related Studies; *J. AWMA* **58**(2), 164 - 195.
- (7) M. S. Bae, J. J. Schauer, J. T. DeMinter, and J. R. Turner, 2004. Hourly and Daily Patterns of Particle-Phase Organic and Elemental Carbon Concentrations in the Urban Atmosphere. *J. AWMA* , **54**, 823-833.
- (8) S. S. Park, M. S. Bae, J. J. Schauer, S. Y. Ryu, Y. J. Kim, S. Y. Cho, and S. J. Kim, 2005. Evaluation of the TMO and TOT Methods for OC and EC Measurements and Their Characteristics in PM_{2.5} at an Urban Site of Korea during ACE-Asia. *Atmospheric Environment*. **39**, 5101-5112
- (9) Harrison, R. M., Jones, A. M., and Lawrence, R. G.: Major component composition of PM₁₀ and PM_{2.5} from roadside and urban background sites, *Atmos. Environ.*, **38**, 4531-4538, 2004.
- (10) Wittig, A.E., Anderson, N., Khlystov, A.Y., Pandis, S.N., Davidson, C., Robinson, A.L. (2004). Pittsburgh air quality study overview. *Atmos. Environ.* **38**, 3107-3125.
- (11) Weitkamp, E.A., Lipsky, E.M., Pancras, J.P., Ondov, J.M., Polidori, A., Turpin, B.J., Robinson, A.L. (2005). Fine particle emission profile for a large coke production facility based on highly time-resolved fence line measurements. *Atmos. Environ.* **39**, 6719-6733.
- (12) Pancras, J.P.; Ondov, J.M.; Poor, N.; Landis, M.S.; Stevens, R.K. (2006). Identification of sources and estimation of emission profiles from highly time-resolved pollutant measurements in Tampa, FL. *Atmos Environ.* **40**, S467-S481.
- (13) Pancras, J.P.; Vedantham, R.; Landis, M.S.; Norris, G.A.; Ondov, J.M. (2011). Application of EPA UNMIX and Non-parametric Wind Regression on High Time Resolution Trace Elements and Speciated Mercury in Tampa, Florida *Aerosol. Environ. Sci. Technol.* **45**, 3511-3518.
- (14) Pancras, J.P.; Landis, M.S.; Norris, G.A., Dvonch, J.T. (2012). Source apportionment of ambient fine particulate matter in Detroit, Michigan, using hourly resolved PM chemical composition data. *Science of the Total Environment* (In Press).
- (15) Brook, R.D., Brook, J.R., Urch, B., Vincent, R., Rajagopalan, S., Silverman, F. (2002). Inhalation of fine particulate air pollution and ozone causes acute arterial vasoconstriction in healthy adults. *Circulation*. **105**, 1534-1536.
- (16) Urch, B., Silverman, F., Corey, P., Brook, J.R., Lukic, K.Z., Rajagopalan, S., Brook, R.D. (2005). Acute blood pressure responses in healthy adults during controlled air pollution exposures. *Environ Health Perspect.* **113**, 1052-1055.
- (17) Hopke, P.K. (1991). Receptor modeling for air quality management. Elsevier, NY.
- (18) Henry, R. (2003). Multivariate receptor modeling by N-dimensional edge detection. *Chemom. Intell. Lab. Sys.* **65**, 179-189.
- (19) Paatero, P., Tapper, U. (1994) Positive matrix factorization: a non negative factor model with optimal utilization of error-estimates of data values. *Environmetrics*, **5**, 111-126.
- (20) Henry, R.C., Park, E.S., Spiegelman, C.H. (1999). Comparing a new algorithm with the classic methods for estimating the number of factors. *Chemometrics and Intelligent Laboratory Systems*, **48**, 91-97.
- (21) Keeler, G.J.; Landis, M.S.; Norris, G.A.; Christianson, E.M.; Dvonch, J.T. (2006). Sources of mercury wet deposition in eastern Ohio, USA. *Environ. Sci. Technol.* **40**, 5874-5881.
- (22) White, E.M.; Keeler, G.J.; Landis, M.S. (2009) Spatial Variability of Mercury Wet Deposition in Eastern Ohio: Summertime Meteorological Case Study Analysis of Local Source Influences. *Environ. Sci. Technol.* **43**, 4946-4953.
- (23) Connell, D.P., Winter, S.E., Conrad, V.B., Kim, M., Crist, K.C. (2006) The Steubenville comprehensive air monitoring program (SCAMP): concentrations and solubilities of PM(2.5) trace elements and their implications for source apportionment and health research. *J Air Waste Manag Assoc.*, **56**, 1750-1766
- (24) Laden, F., Neas, L.M., Dockery, D.W., Swartz, J. (2000) Association of fine particulate matter from different sources with daily mortality in six U.S. cities. *Environ Health Perspect.*, **108**, 941-947.
- (25) Pancras, J.P.; Landis, M.S. (2011). Performance Evaluation of modified Semi-continuous Elements in Aerosol Sampler-III. *Atmos Environ.* **45**, 6751-6759.
- (26) Henry, R.C. (1994) Multivariate receptor models-current practice and future trends. *Chemometrics and Intelligent Laboratory Systems*. **1**, 43-48.
- (27) Nyirenda, R.L. (1991). The processing of steelmaking flue-dust: A review. *Minerals Engineering*. **4**, 1003-1025.

-
- (28) Antrekowitsch, J, Antrekowitsch, H. (2001). Hydrometallurgically recovering zinc from electric arc furnace dust. JOM. 53, 26-28.
- (29) Kovalcik, K., Landis, M.S., Norris, G.N., Turner, J., Duvall, R. (2014) Near Field PM_{2.5} Impacts from a Primary Integrated Steel Mill and a Secondary Electric Arc Furnace Facility in Granite City, Illinois, USA: Source Profiles, Solubility, and Aerosol Morphology, *in preparation*
- (30) Di Schino, A., Kenny, J.M., Mecozzi, M.G., Barteri, M. (2000). Development of high nitrogen, low nickel, 18% Cr austenitic stainless steels. Journal of Materials Science 35.19: 4803-4808.
- (31) Pacyna, Lead, Mercury, Cadmium and Arsenic in the Environment, Edited by T. C. Hutchinson and K. M. Meema @ 1987 SCOPE. Published by John Wiley & Sons Ltd, CHAPTER 7, Atmospheric Emissions of Arsenic, Cadmium, Lead and Mercury from High Temperature Processes in Power Generation and Industry
- (32) ASM Handbook Volume 5, Surface Engineering ASM International, pp. 214-215, Cadmium Plating, Milton F. Stevenson
- (33) Ondov, J.M., Wexler, A.X. (1998), Where do particulate toxins reside? An improved paradigm for the structure and dynamics of the urban mid-Atlantic aerosol. Environ Sci Technol, 32: 2547-2555.
- (34) Keeler, G.J., Landis, M.S., Norris, G.A., Christianson, E.M., Dvonch, J.T. (2006). Sources of Mercury Wet Deposition in Eastern Ohio, USA Environ. Sci. Technol. 40, 5874-5881.
- (35) Jiun-Horng, T., Kuo-Hsiung, L., Chih-Yu, C., Jian-Yuan, D., Ching-Guan, C., Hung-Lung, C. (2007). Chemical constituents in particulate emissions from and integrated iron and steel facility. Journal of Hazardous Materials, 147, 111-119.
- (36) Moskalyk, R.R., Alfantazi, A.M. (2003). Processing of Vanadium: a review. Mineral Engineering. 9, 793-805.
- (37) Monakhov, I.N., Khromov, S.V., Chernousov, P.I., Yusfin, Y.S., (2004). The flow of vanadium-bearing materials. Metallurgist. 48, 381-385.
- (38) Grahame, T. and Hidy, G, (2004), Using Factor Analysis to Attribute Health Impacts to Particulate Pollution Sources, Inhalation Toxicology, 16(suppl. 1):143-152.
- (39) Vedantham et. al. Atmospheric Pollution Research 3 (2012) 105-111, Atmospheric Pollution Research, Combining continuous near-road monitoring and inverse modeling to isolate the effect of highway expansion on a school in Las Vegas.