A quantitative method for clustering size distributions of elements

Ann M. Dillner\textsuperscript{a,\*}, James J. Schauer\textsuperscript{b}, William F. Christensen\textsuperscript{c}, Glen R. Cass\textsuperscript{a,\*}

\textsuperscript{a}Department of Civil and Environment Engineering and Department of Chemical and Materials Engineering, Arizona State University, Tempe, AZ 85287-5306, USA
\textsuperscript{b}Environmental Chemistry and Technology Program, University of Wisconsin-Madison, Madison, WI 53706, USA
\textsuperscript{c}Department of Statistics, Brigham Young University, 219 TMCB, Provo, UT 84602-657, USA

Received 14 July 2004; received in revised form 18 November 2004; accepted 24 November 2004

Abstract

A quantitative method was developed to group similarly shaped size distributions of particle-phase elements in order to ascertain sources of the elements. This method was developed and applied using data from two sites in Houston, TX; one site surrounded by refineries, chemical plants and vehicular and commercial shipping traffic, and the other site, 25 miles inland surrounded by residences, light industrial facilities and vehicular traffic. Twenty-four hour size-segregated (0.056 < \( D_p \) (particle diameter) < 1.8 \( \mu \)m) particulate matter samples were collected during five days in August 2000. ICP-MS was used to quantify 32 elements with concentrations as low as a few picograms per cubic meter. Concentrations of particulate matter mass, sulfate and organic carbon at the two sites were often not significantly different from each other and had smooth unimodal size distributions indicating the regional nature of these species. Element concentrations varied widely across events and sites and often showed sharp peaks at particle diameters between 0.1 and 0.3 \( \mu \)m and in the ultrafine mode (\( D_p < 0.1 \) \( \mu \)m), which suggested that the sources of these elements were local, high-temperature processes. Elements were quantitatively grouped together in each event using Ward’s Method to cluster normalized size distributions of all elements. Cluster analysis provided groups of elements with similar size distributions that were attributed to sources such as automobile catalysts, fluid catalytic cracking unit catalysts, fuel oil burning, a coal-fired power plant, and high-temperature metal working. The clustered elements were generally attributed to different sources at the two sites during each sampling day indicating the diversity of local sources that impact heavy metals concentrations in the region.

\( \text{\copyright} \ 2004 \) Elsevier Ltd. All rights reserved.

Keywords: ICP-MS; Cluster analysis; Metals; Particulate matter; Source attribution

1. Introduction

Size distributions of elements in ambient particulate matter have been used to suggest sources of elements in urban areas such as Los Angeles (Singh et al., 2002) and in regions with large numbers of point sources such as an area with many coal-fired power plants (Dodd et al., 1991). Supercmicrometer particles, primarily generated...
by mechanical processes including soil, sea spray or industrial mechanical processes, are composed of largely crustal elements (Seinfeld and Pandis, 1998). High-temperature processes, like combustion and metal working, emit submicrometer heavy metals (Ondov and Wexler, 1998). Condensational growth of particles during the cooling stages of high-temperature processes produce narrow size distributions of elements which peak between 0.1 and 0.3 μm (Dodd et al., 1991).

In addition to source information, size distributions can provide information about the proximity of the sources to the sampling site. For example, levels of primary ultrafine particles \((D_p)\) particle diameter <0.10 μm have been shown to be highest close to their source (Reponen et al., 2003). Size distributions with peaks above 0.3 μm may be due to particle growth during transport and or due to changes in relative humidity, and therefore particle size, during the sampling (Ondov and Wexler, 1998). Broad size distributions dominated by sulfate which peak around 0.7 μm result from slow condensation of gas-phase species and cloud processing of particles (Ondov and Wexler, 1998) and are regional in nature.

Other approaches used to identify sources of heavy elements in ambient particulate matter are principle component analysis (PCA) techniques used on bulk aerosol samples (Kim et al., 2003; Song et al., 2001) and cluster analysis techniques used on single particle data (Osan et al., 2001; Phares et al., 2003). The PCA approach relies on a large data set (usually greater than 60–100 sample events) to obtain statistical significance and therefore typically cannot exploit size-resolved chemical composition information due to the extreme cost required to analyze such large numbers of samples. Large data sets from single particle analysis have been analyzed using cluster analysis to group particles with similar composition (Hoornaert et al., 2003; Osan et al., 2001). Phares et al. (2003) used this approach to analyze ultrafine particle composition in Houston, TX, during TEXAQS 2000.

Our goal in the present study was to elucidate sources of elements at two locations in Houston, TX, a highly industrialized city, using size-resolved elemental concentration data. ICP-MS was used to quantify the fine and size-segregated concentrations of 32 elements from sodium to uranium with concentrations ranging from picograms to micrograms per cubic meter. The current study exploits the idea, suggested by Ondov and Wexler (1998), that elements with similar size distributions often originate from the same source. To our knowledge, the present study is the first to use a quantitative method to cluster the size distributions of elements with similar shapes. Sources of the clustered elements are suggested based on the clustered elements and the shape of the size distribution. In addition to element concentrations, fine and size-segregated concentrations of inorganic ions and elemental and organic carbon (OC) were analyzed to give the overall composition of the aerosol at the two sites.

## 2. Experimental methods

### 2.1. Sampling

Size-resolved particulate matter samples were collected at two TEXAQS 2000 sites in Houston, TX, during August 2000. One sampling site, LaPorte, is located near the Houston ship channel, which has a high density of oil refineries, chemical plants and vehicular and ship traffic (Russell et al., 2004). The second site, Aldine, is located approximately 25 miles inland, north of downtown Houston and is characterized by vehicular traffic and light industry. Twenty-four hour, PM_{1.8} \((D_p <1.8 \text{ μm})\) were collected on 23, 25, 27, 29, and 31 August at both sites. Each PM_{1.8} sampler contained two Teflon filters (Gelman Teflo, 47 mm) and one quartz filter (Pallflex, 47 mm) and operated at a nominal flow rate of 24 liters per minute (lpm). Simultaneously, three sets of impactor samples were collected using Micro-Orifice, Uniform Deposit Impactors (MOUDI, MSP Corp.) at both sites during all events except 25 and 29 August at Aldine. At each site, two MOUDIs were loaded with Teflon filters (Gelman Teflo, 47 mm) and one MOUDI was loaded with aluminum substrates (MSP Corp., 47 mm). The MOUDIs were operated at 30 lpm with an AIHL-design cyclone separator upstream of the inlet to remove particles larger than 1.8 μm. Particles were collected in six aerodynamic particle size bins: 0.056–0.1, 0.1–0.18, 0.18–0.32, 0.32–0.56, 0.56–1.0 and 1.0–1.8 μm. Flowrates of all impactors and samplers were measured before and after each sampling event. Sampled filters and substrates were frozen until physical and chemical analysis was performed.

### 2.2. Physical and chemical analysis

Gravimetric mass was determined by weighing each teflon filter and aluminum substrate a minimum of three times before and after sampling using a Mettler Toledo Model M-55 microgram balance at a constant temperature \((20.5\pm1.6 \text{ °C})\) and relative humidity \((44.5\pm4.3\% \text{ RH})\). Sulfate, nitrate and chloride concentrations were obtained from teflon filters using ion chromatography (Mulik et al., 1976) and for ammonium by colorimetry (Bollette et al., 1961). Elemental and OC concentrations were obtained using the thermal optical methods as described by Birch and Cary (1996) for quartz filters and the method described by Kleeman et al. (1999) for foil substrates. Organic compound concentrations were calculated from the OC by multiplying by a factor of 1.4 to account for atomic species associated with the OC. Foils and quartz filters were baked prior to
sampling to remove organic contamination and were stored in annealed, foil-lined Petri dishes.

Concentrations of 32 light and heavy elements were obtained from teflon filters and substrates using inductively coupled plasma/mass spectrometry (ICPMS, VP PlasmaQuad Excell, TJA Solution). Prior to analysis, filters were digested in a microwave-assisted acid bath. Heavy elements were quantified using standard hot-plasma ICPMS (1350 watt RF): titanium (Ti), vanadium (V), chromium (Cr), manganese (Mn), nickel (Ni), copper (Cu), zinc (Zn), arsenic (As), rubidium (Rb), strontium (Sr), molybdenum (Mo), ruthenium (Ru), rhodium (Rh), palladium (Pd), cadmium (Cd), tin (Sn), antimony (Sb), cesium (Cs), barium (Ba), lanthanum (La), cerium (Ce), tungsten (W), platinum (Pt), thallium (Tl), lead (Pb) and uranium (U). In addition, five light elements, sodium (Na), magnesium (Mg), aluminum (Al), potassium (K), and calcium (Ca), and one heavy element, iron (Fe) were quantified in the cool plasma/shielded torch mode of the ICPMS. The digestion and ICPMS analysis method has been validated previously and the details are discussed by Schauer et al. (2004).

One set of blank filters and substrates at each site was used to blank correct the measured species mass. Uncertainties in calculated species concentrations were obtained using a propagation of errors technique that included uncertainties in the mass of the species on the sampled and blank filters.

To assess whether the species concentrations at the two sites ($X_A$ and $X_L$, concentration of species $X$ at Aldine, A, and LaPorte, L) were equal, a test statistic that was constructed based on the assumption that the measurement errors for the two sites were independent and normally distributed with standard deviations equal to the measured uncertainties ($\sigma_A$ and $\sigma_L$). Under the hypothesis of equal gravimetric mass concentrations, the statistic, $z = \left( X_A - X_L \right) \sqrt{\sigma_A^2 + \sigma_L^2}$ follows a normal distribution. $p$-Values greater than 0.05 indicate that the concentrations at the two sites are not significantly different from each other.

### 3. Method for clustering size distributions

A method that uses cluster analysis to group elements with similar size distributions was developed. The clustered elements were then used to identify sources of elemental particulate matter.

Any algorithm used to cluster $N$ objects in a hierarchical manner is based on an $N \times N$ matrix of inter-object distances which looks much like a table of mileages. For this work, the objects to be clustered were the element size distributions measured at one site during one event. The size distribution for element 1 was represented by the vector $x_1 = (x_{11}, \ x_{12}, \ldots, \ x_{1j})$ where each value in the vector was the concentration of element 1 for a given size range. The measurement for each size range on each element was used in the analysis, regardless of the closeness of the measurement to zero. The value $x_{1j}$ represents the $j$th size bin ($j = 1, \ldots, 6$) collected by the MOUDIs. (For example, size bin 1 contained particles with aerodynamic equivalent diameters between 0.056 and 0.1 $\mu$m.) In order to compare the size distributions for species with dramatically different mass concentrations, measured concentrations were converted to relative concentrations as follows: $p_i = \left( p_{i1}, \ p_{i2}, \ldots, \ p_{i6} \right) = \frac{\left( x_{i1} + x_{i2} + \cdots + x_{i6} \right) - \bar{p}_K}{\bar{p}_K}$.

Ward’s method (Ward, 1963), an agglomerative hierarchical clustering approach, was used to cluster the elements because it has the advantage of interpretability. In Ward’s method, each object (element size distribution) begins as its own cluster. During each step of the process, the two closest clusters are joined to form a new cluster which replaces the two old clusters. To calculate the distance between the normalized size distributions of two species, $p_1 = \left( p_{11}, \ p_{12}, \ldots, \ p_{16} \right)$ and $p_2 = \left( p_{21}, \ p_{22}, \ldots, \ p_{26} \right)$, the squared Euclidian distance, $d_{12} = \sum_{j=1}^{6} \left( p_{1j} - p_{2j} \right)^2$, was used. To calculate the distance between clusters with more than one element, Ward’s method uses the ANOVA sum of squares between the two cluster means summed over all the size bins with the distances between objects defined as half of the squared Euclidean distance. That is, the distance between clusters for Ward’s method is the ANOVA sum of squares between the two cluster means summed over all the size bins. Let $p_{Kl} = \left( p_{Kl1}, \ p_{Kl2}, \ldots, \ p_{Kl6} \right)$, for $l = 1, \ldots, N_K$, and $N_K$ be the number of elements in cluster $K$, and let $\bar{p}_K = \left( \bar{p}_{K1}, \ldots, \bar{p}_{K6} \right)$ be the mean vector for the $K$th cluster where $\bar{p}_K = \frac{1}{N_K} \sum_{l=1}^{N_K} p_{Kl}$. Then, the distance between cluster $K$ and cluster $L$ using Ward’s Method is

$$D_{K,L} = \frac{\sum_{j=1}^{6} \left( \bar{p}_{Kj} - \bar{p}_{Lj} \right)^2}{1/N_K + 1/N_L}.$$ 

Thus, in Ward’s method, we consider the within-cluster sum of squares after combining two clusters $K$ and $L$. After considering all possibilities for the pair $(K, L)$, we combine the pair of clusters that minimizes the within-cluster sum of squares. This process is continued until the $(N-1)$th step, when the final two clusters are joined and the clustering is complete. In contrast with some other clustering approaches, this approach requires only one pass through the algorithm (i.e., the approach is not iterative). Thus, once two elements are joined in a common cluster, they are never again separated from each other.

A metric for identifying a natural number of clusters is the squared multiple correlation, $R^2$.

$$R^2 = 1 - \frac{1}{T} \left( \sum_{K=1}^{G} \sum_{i=1}^{N_K} \sum_{j} \left( p_{Kij} - \bar{p}_{Kj} \right)^2 \right).$$
Note that \( T = \sum_{k=1}^{G} \sum_{i=1}^{N_k} \sum_{j=1}^{6} (P_{ki} - \bar{P}_j)^2 \) is the total sum of squares for the data, \( \bar{P}_j = 1/N \sum_{k=1}^{G} \sum_{i=1}^{N_k} P_{ki} \), \( N_k \) is the number of elements in the \( k \)th cluster, and \( N = \sum_k N_k \) is the total number of objects being clustered. The value \( R^2 \) is the proportion of variance accounted for by the clusters and \( G \) is the total number of clusters. In practice, ad hoc rules for choosing the number of clusters based on \( R^2 \) are chosen to yield physical meaning for the given data. For this data set, the clustering of elements was determined to be meaningful when \( R^2 \) was greater than 0.96. By choosing a high \( R^2 \) value, only elements with very similar size distributions were clustered together and some elements were left in clusters with only one member. The high \( R^2 \) value provides for a clear interpretation of the sources of the clustered elements and recognizes that some sources emit only one measured element and some elements have multiple sources.

4. Results and discussion

4.1. Bulk particulate matter composition

Speciated concentrations of PM\(_{1.8}\) at Aldine and LaPorte for the five sampling events in this campaign are shown in Fig. 1. The PM\(_{1.8}\) mass concentrations at the two sites are not significantly different from each other (\( p \)-values > 0.05) during any of the five sampling events. The measured concentrations and similarity of concentrations between sites is consistent with other studies in southeast Texas (Russell et al., 2004). At both sites, every event has a unimodal mass size distribution that peaks at 0.32–0.56 μm. Speciated size distributions for 27 and 31 August at Aldine and LaPorte are shown in Fig. 2 as examples. The similar concentrations and size distributions indicate the regional nature of the bulk fine particulate matter.

The PM\(_{1.8}\) mass measured during this sampling campaign is dominated by OC and sulfate ion as was found in PM\(_{2.5}\) collected in this region by others (Russell et al., 2004). OC is the largest contributor to fine mass and concentrations of OC are not significantly different (\( p \)-values > 0.05) at the two sites during any event. Size distributions of OC aerosol have a single peak at 0.32–0.56 μm (same as total mass), except for 25 August at LaPorte where the peak is shifted to a smaller particle size. The similarity in size distributions and concentrations between the two sites indicates the regional nature of the OC. Large sources of primary gaseous and particulate matter organic compounds, including refineries and ships, surround the LaPorte site. However, these large sources did not yield elevated OC concentrations compared to Aldine during this study, further indicating the regional nature of the fine OC particulate.

Sulfate ion and ammonium ion size distributions are also very similar across locations and events with smooth unimodal size distributions that peak around the same particle size as particle mass indicating that the...
sulfate and ammonium are secondary and regional in nature. Nitrate concentrations are small and found mostly at 1–1.8 μm. Chloride concentrations for most of the fine and size-segregated samples are below detection.

Elemental carbon concentrations are not significantly different (p-values > 0.05) between the two sites during any event, likely due to diesel traffic and other combustion sources located throughout the city.

4.2. Element composition: general characterization

As shown in Fig. 3a, the total concentrations of elements (the sum of 32 elements quantified by ICP-MS) at the two sites are not significantly different (p-values > 0.05) from each other during any event. This suggests a high degree of similarity in elemental composition at the two sites, as was seen for OC and sulfate. However, further analysis of individual element concentrations and size distributions (discussed below) indicate that the elemental composition of the aerosol at the two sites is quite different likely as a result of different local sources.

As shown in Fig. 3a, two types of events related to the total element concentrations were observed; element concentrations above 1 μg m⁻³ (“high events”) which occurred on 27 and 29 August at both sites and 31 August at the Aldine site and elemental concentrations below 0.25 μg m⁻³ (“low events”) which occurred on 23 and 25 August at both sites and 31 August at LaPorte. These high and low events of total elements do not correlate with high and low PM₁.₈ mass, OC or sulfate, indicating that the elemental composition has different sources than OC and sulfate and is not driven by the regional sources and transport observed for OC and sulfate.

Fine elemental concentrations are dominated by the light elements Na, Mg, Al, K, and Ca (Fig. 3a). The high elemental concentration events are characterized by high Ca concentrations. The Ca concentration during high events (2.3 ± 1.3 μg m⁻³) is on average 82 ± 5% of the total elemental concentration and is two orders of magnitude higher than the average concentrations during low events (0.015 ± 0.003 μg m⁻³). The low events are characterized by large relative contributions from Na and Al although the absolute concentrations of these elements are not significantly different (p-values > 0.05) between the low and high events. Na, Mg, Al, K, and Ca contribute 77 ± 9% the mass of elements during low events, but contribute 97 ± 2% of the elemental mass during high events.

The sum of the remaining 28 elements, labeled “heavy” in Fig. 3a and shown in detail in Fig. 3b, do not follow the high/low patterns of the light elements and vary less between events than the light elements. Fe is the most abundant heavy element at both sites and accounts for 61–85% of the total heavy element concentration. V, a tracer of fuel oil burning (Divita...
et al., 1996), is significantly higher ($p$-values $< 0.05$) at LaPorte than at Aldine for all events, likely as a result of fuel oil burning by ships and at refineries located near LaPorte.

### 4.3. Element composition: cluster analysis of size distributions and resulting source attribution

For each event with measured size distribution data, a quantitative clustering technique was used to group size distributions of elements with similar shapes. There are eight events with size distribution data including all events at LaPorte and the 23, 27 and 31 August events at Aldine. Table 1 lists the clusters obtained for each event. For illustrative purposes, size distributions for 18 of the 32 elements are shown for Aldine on 31 August (high event, Fig. 4) and for LaPorte on 31 August (low event, Fig. 5). The remaining 14 elements are shown in the supplementary material. Both figures show the same elements in the same order with a common concentration scale for each element except for Ca and K, whose concentrations vary by orders of magnitude between events. In both figures, clustered elements are indicated by bar shading. Visual inspection of the figures, for example the Sb–Pb cluster in Fig. 4, shows that the clustering technique successfully groups elements with similar size distributions. A figure in the supplementary material shows the mean and standard deviation of the histogram for each cluster for each event, illustrating the data-reducing capability of this method. In the following sections, selected clusters of elements are discussed and sources of the clusters of elements are suggested based on the elements in the cluster and the shape of the size distribution.

**Lanthanum (La) and cerium (Ce):** La and Ce are in the same cluster during five of the eight events with size distribution data. At LaPorte on 25 August, the La–Ce cluster has a unimodal size distribution that peaks at 0.18–0.32 μm. The La:Ce ratio ranges from 0.9:1 to 1.3:1.
for the measured size bins, which is within the range of values in fluid catalytic cracker (FCC) catalysts (0.45:1–5.3:1) used at petroleum refineries (Kitto et al., 1992), suggesting that the La and Ce are from refineries. La and Ce also cluster (sometimes with other elements such as Mn, Fe, and Zn) at LaPorte and Aldine on 23 and 27 August with La:Ce ratios (1.0–6.7) that are also consistent with FCC catalysts. However, for these four events, the size distribution is in the form of a coarse tail, i.e., a size distribution that has the largest concentration at 1–1.8 μm and has decreasing concentrations with decreasing particle size. The different size distributions are likely a result of differences in emissions from the refinery FCC units. Catalyst containment is different during normal operations and transient operations which yields different catalyst inventories and different FCC cyclone collection efficiencies (Gary and Handwerk, 1994).

Ce is also used in automobile catalysts (Kitto et al., 1992). At LaPorte Ce, Mo and U cluster with a peak at 0.18–0.32 μm on 29 August and Ce (at very low concentrations), Mo and Ti cluster with a similar peak on 31 August (Fig. 5). Single particles containing Ce and Mo have been observed in automobile exhaust (Silva and Prather, 1997), indicating that these clusters might result from particulate emissions from automobile catalysts.

**Vanadium (V):** V size distributions are typically unimodal with peaks below 0.32 μm which supports...
Fig. 4. Element size distributions (ng m$^{-3}$) measured at Aldine, TX, on 31 August 2000. Distributions with the same shading are in the same cluster. Element distributions without shading do not cluster with other elements in the figure.
Fig. 5. Element size distributions (ng m⁻³) measured at LaPorte, TX, on 31 August 2000. Distributions with the same shading are in the same cluster. Element distributions without shading do not cluster with other elements in the figure.
the use of V as a tracer for fuel oil combustion. However, the shape of the V size distribution and the elements clustered with V vary between sites and events. On 25 August at LaPorte, high concentrations of V and Ni cluster with As, Pd, and Cd. The cluster size distribution peaks over two size bins (0.1–0.32 μm) and has on average 86 ± 7% of each element’s total mass in the peak. Nickel is also associated with fuel oil combustion as are minor amounts of As and Cd (Claes et al., 1998) and although Pd is not reported by Claes et al. (1998), it is likely that this cluster of elements is from fuel oil combustion. V forms a cluster with Sn at LaPorte on 27 August and with Cu at Aldine on 27 August and with Cr on 31 August at Aldine (Fig. 4). It is unclear what sources these clusters represent.

Rhodium (Rh), palladium (Pd), and platinum (Pt): Pd, Pt and Rh are active elements in automobile catalysts (Rauch et al., 2001). At LaPorte on 29 August, Pd and Pt form a cluster with W and Ti that has a coarse-tail distribution. During this event, a coarse-tail size distribution was also observed for an As-Rh cluster. These two clusters (Pd–Pt—W–Ti and As–Rh) would themselves cluster together at an \( R^2 = 0.95 \), (slightly lower than the \( R^2 = 0.96 \) used throughout this work). Rauch et al. (2001) measured Pt-Rh ratios between 2 and 6 in ambient particulates collected near roadways in Gotteborg, Sweden. During the 29 August event at LaPorte, the Pt-Rh ratios in the measured size bins ranged from 1.8 to 6.6, consistent with automobile catalysts. Rh, Pd, and Pt form a cluster with Ca, Rb and U at Aldine on 27 August with Pt-Rh ratios that range from 2.0 to 3.0 (except for one stage where both elements are within one standard deviation of zero), also consistent with automobile catalysts. A coarse tail cluster of Pd and Pt with Sn, W and Ti observed at LaPorte on 31 August (Fig. 5) is smaller in magnitude to that observed at LaPorte on 29 August. Although Rh does not cluster with Pd and Pt due to a large ultrafine (0.056–0.1 μm) peak of Rh, similar size distributions for Rh, Pd and Pt above 0.1 μm suggest that the source is also automobile catalyst. The coarse-tail size distributions observed for all of these clusters may indicate that the PM source is resuspended road dust contaminated with the catalytic elements (Schafer et al., 1999) rather than direct automobile catalyst emissions, which should be primarily in the submicrometer size range (Rauch et al., 2001). Interestingly, at LaPorte on 29 and 31 August, two of the events with Rh–Pd–Pt clusters attributed to automobile catalyst, a Mo–Ce cluster, also attributed to automobile catalysts, is observed.

Sodium (Na), aluminum (Al) and potassium (K): At LaPorte on 25 August, a low event, Na and Al, which are dominant species during low events, form a cluster with K that has a large, sharp peak at 0.32–0.56 μm, suggesting a high-temperature source. At LaPorte and Aldine on 27 August, both high events, Na, Al and K cluster with bimodal distributions. The relative proportions of Na, K and Al (Na:K:Al) in these events averages 4:1:1 suggesting that these elements may be from soil dust (1:1:3, Taylor and McLennan, 1985) but are not likely to be solely from seasalt (27:1:8E-8, Wilson, 1975). Due to the bimodal nature of the distribution, these elements may be from a mixture of sources. Na and Al cluster together during another low event, LaPorte on 31 August (Fig. 5), with a bimodal distribution and Na:Al ratios ranging from 1.0:1 to 1.4:1 which is similar to that in soil dust (0.4:1; Taylor and McLennan, 1985) but the elements may also be from a mixture of sources.

Potassium (K) and calcium (Ca): K and Ca concentrations, which dominate the elemental mass during high events, are one to two orders of magnitude higher than during low events and form a cluster with a large peak at 0.32–0.56 μm and 0.056–0.10 μm on 29 August at LaPorte (high event). Phares et al. (2003) observed a calcium oxide peak at 0.07 μm on the same day on the other side of the Houston Ship Channel from LaPorte using a single-particle mass spectrometer. In the same event, Na, Mg, Mn, Fe and Sb form one cluster and Al, Ti, Rb and Ba form another cluster both with sharp peaks between 0.18–0.32 μm. Slightly loosening the cluster criteria to \( R^2 = 0.94 \) produces a cluster containing all nine elements. Ca and K, as well as Mg, Ti, and Ba concentrations are all significantly higher (p-values<0.05) during this event than during any other. One possible source of these very large concentrations and narrow peaks is a coal fired power plant. All of the nine elements except for Na and Rb are reported to be from coal (Pacyna, 1998). Na and Rb are probably in coal fly ash but were not reported in Pacyna (1998) because of the analytical method used. Ca and K may have been emitted from the scrubber on the coal fired power plant as suggested by Phares et al. (2003).

Ca, K and Al form a cluster at Aldine on 31 August (Fig. 4). Ca and K are one to two orders of magnitude higher during this event than during low events; however, the Al concentration is fairly similar to concentrations throughout the study indicating that like LaPorte on 29 August, the cluster is dominated by Ca and K. Unlike LaPorte on 29 August, the size distribution is fairly uniform and a cluster containing a large group of elements does not occur. The ratios of Ca:K:Al are approximately 100:10:1 during this event but in soil the ratios are 0:4:0.41 (Taylor and McLennan, 1985) so the source is not likely to be soil.

For the two other events with high elemental concentrations and measured size distributions (27 August at LaPorte and Aldine), calcium does not cluster with potassium or aluminum. At LaPorte, Ca clusters with Mg and Fe and has a coarse tail size distribution. During this event the ratios of elements in the peak particle size (1.0–1.8 μm) are Fe:Ca:Mg (3.0:2.7:1) are
similar to those of soil (2.6:2.3:1; Taylor and McLennan, 1985) suggesting a soil dust source. As discussed above, at Aldine, Ca clusters with Rb, Rh, Pd, Pt and U and the cluster is likely due to automobile catalysts. Although the high elemental concentration events are dominated by calcium, the size distribution of calcium indicates that it is from different sources at different sites and days.

Iron (Fe), antimony (Sb) and barium (Ba): At LaPorte on 25 August, Fe, Sb and Ba, elements commonly found in brake wear dust (Garg et al., 2000), form a cluster with a coarse tail distribution. The ratios of Fe:Ba:Sb in the 1–1.8 μm size bin, 13:1.2:1, are similar to the 6:2:1 ratios measured in limited testing in Garg et al. (2000). Sb is enriched relative to Fe and Ba compared to soil dust (175,000:2750:1, Taylor and McLennan, 1985).

Copper (Cu), antimony (Sb) and lead (Pb): At LaPorte on 31 August (Fig. 5), Cu, Sb, and Pb form a cluster in which 86–91% of the mass of the three elements are in the peak at 0.18–0.32 μm. The small particle size of this peak suggests a local high temperature source. During this event, Fe and Zn also have a single concentration spike with 65–71% of the mass also at 0.18–0.32 μm. Although Cu, Sb, and Pb and Fe and Zn do not cluster together, it is possible that the mass at 0.18–0.32 μm is from the same high-temperature source.

Cu, Sb, and Pb also form a cluster at Aldine on 31 August (Fig. 4) and have a single concentration spike similar to LaPorte (Fig. 5). However, compared to the LaPorte values, the concentrations at Aldine are 20–40% lower, the percentage of the total mass in the peak (~80%) is lower, and the peak is shifted one bin larger at Aldine. These elements might be due to the same source with the shift in size distribution due to difference in transport time which could produce differences in particle size. Meteorological data measured at LaPorte and Aldine indicate that on 31 August the wind was predominately out of the west–southwest at both sites (59% of the sampling period at LaPorte and 66% of the sampling time at Aldine) making it unlikely that these particles traveled northwest from LaPorte to Aldine. Since the peak of the size distribution has more mass and is in a smaller size bin at LaPorte, it is likely that the source is closer to LaPorte than to Aldine. Fe and Zn, which cluster together and can be associated with Cu, Sb and Pb at LaPorte, are not associated with the cluster at Aldine. Furthermore, the Fe concentration is higher at Aldine than at LaPorte, indicating an additional source of Fe at Aldine.

Zinc (Zn): At LaPorte Cu and Zn form a cluster with Rb and Pt on 27 August while Cu and Zn form a cluster with Sr on 29 August. Cu and Zn are indicative of a high-temperature non-ferrous metal smelter (Claes et al., 1998) but the Cu–Zn particles collected at LaPorte during these 2 days are likely from two different sources because they cluster with different elements. Mg, Mn, Zn and Cd form a cluster with a peak at 0.32–0.56 μm at Aldine on 31 August (Fig. 4), suggesting a third source of zinc.

Ultrafine clusters: At LaPorte on 31 August (Fig. 5), a number of elements have high concentrations in the ultrafine size range. Ni, Ru and Cs form a cluster with high ultrafine concentrations with 52–67% of each element’s mass below 0.1 μm. Cr, As, Rb, Sr, Rh and Ba also have high but more modest ultrafine concentrations with 30–45% of the individual element’s mass in the ultrafine size bin. Ca does not cluster with any other elements but 65% of the Ca is in the ultrafine size range. Cd and U cluster together with three distinct peaks, one of them in the ultrafine range. This multitude of clusters and the high ultrafine mass contribution, suggest that there are several sources that produce heavy element ultrafine aerosol in the LaPorte area. High ultrafine concentrations are not seen to this extent on any other event.

5. Conclusions

A quantitative clustering technique was developed to group elements whose size distributions have similar shapes. The method was applied to size-segregated samples from two sites in Houston, TX. ICP-MS was used to obtain elemental concentrations because it provides data for a large number of elements over a wide range of concentrations. Elements known to be emitted from high-temperature sources such as combustion and metal working often had size distributions with sharp peaks below 0.3 μm. Soil related elements as well as some heavy elements occasionally had distributions that appeared to be the tail of a supermicrometer peak.

The clustering method quantitatively grouped elements together, which based on visual inspection, were very similar in shape. Some of the clusters contained elements that have been shown to be emitted from known sources types in Houston. For example, on 25 August at LaPorte: (1) a La–Ce cluster was attributed to an oil refinery FCC unit and (2) a V–Ni–As–Pd–Cd cluster was attributed to fuel oil burning by ships and refineries. In addition, some clusters of elements, for instance the V–Cr cluster observed on 31 August at Aldine (Fig. 4), are from unknown sources, indicating the need for additional source characterization. The clustering technique and source attribution indicates that element concentrations in Houston are due to local sources that are highly variable between sites and days. In contrast, concentrations and size distributions of particulate mass, inorganic ions and OC were very similar between sites and events, indicating the regional nature of these pollutants. This new mathematical clustering technique is a useful tool to
group elements based on their size distribution to enable source attribution.

Acknowledgements

Although the research described in this article has been funded in part by the United States Environmental Protection Agency through cooperative agreement R827354, it has not been subjected to the Agency’s required peer and policy review and therefore does not necessarily reflect the views of the agency and no official endorsement should be inferred. Thanks to Lynn Salmon for inorganic and elemental and organic carbon analysis and Dr. Martin Shafer and Dr. Michael Arndt for the analysis of the metals by ICPMS.

Appendix A. Supplementary material

The online version of this article contains additional supplementary data. Please visit doi:10.1016/j.jssc.2004.08.007.

References


and receptor sites in the Los Angeles basin. Atmospheric Environment 36, 1675–1689.


