AIR POLLUTION SOURCE IDENTIFICATION

by J. Stuart Fordyce
Lewis Research Center
Cleveland, Ohio 44135

TECHNICAL PAPER presented at
Sources and Emissions Workshop of the Second
Interagency Committee on Marine Science and
Engineering Conference on the Great Lakes
Argonne, Illinois, March 25-27, 1975
I. SUMMARY

The regulatory agency faced with developing abatement strategies and enforcing emission standards has inadequate tools available to identify sources that may not be in compliance and to associate source emissions of specific types with ambient levels in a cause and effect relationship. This inadequacy is particularly important as federal standard setting shifts from gross pollutant loadings to control of specific pollutants in a multimedia (air, water, land) context.

The techniques available for source identification are reviewed. Remote sensing, injected tracers and pollutants themselves as tracers are described. The use of the large number of trace elements in the ambient airborne particulate matter as a practical means of identifying sources is discussed in detail. The availability of sensitive, inexpensive, non-destructive, multielement analytical methods such as instrumental neutron activation and charged particle x-ray fluorescence permit the determination of fifty or more trace constituents. The application of pairwise correlation, the more advanced pattern recognition-cluster analysis approach with and without training sets, enrichment factors and pollutant concentration rose displays for each element to a large data set obtained in Cleveland, Ohio are described. It is shown that elemental constituents can be related to specific source types: earth crustal, automotive, metallurgical and more specific industries. A field-ready source identification system based on time and wind direction resolved sampling is described which is to
be evaluated for use by local agencies.

II. INTRODUCTION

The Federal Clean Air Act of 1970 and its subsequent implementation by the states under approval by the U. S. EPA requires the ambient monitoring of pollutants for which ambient air quality standards have been adopted. At the present time, this includes sulfur dioxide, nitrogen dioxide, total suspended particulate, carbon monoxide, total non-methane hydrocarbons, and total oxidants (ozone). Under the Hazardous Air Pollutant category, asbestos, mercury and beryllium are controlled by emission standards and the ambient monitoring for them is required. Other pollutants are already recognized and will come under regulation and will have to be monitored. These are listed in Table I directly quoted from the paper by Forziati (ref. 1), which also discuss monitoring techniques. The list in Table I is extensive and presents a reasonable estimate of air pollutants that are important to the national environment. Specifically in the Great Lakes area, EPA Region V personnel have indicated that beyond the 6 major pollutants and the 3 HAP, the need for trace element monitoring is paramount. They would place lead, arsenic, cadmium, antimony, chromium, iron, nickel, selenium, vanadium and zinc as being particularly important.

The ambient monitoring is required in order to define the appropriate emissions limitations that various sources must adhere to in order to meet the air quality standards under various meteorological conditions. It is clear, however, that enforcement agencies who have to plan and implement abatement strategies and enforce the emission standards established are inadequately staffed and equipped to identify unambiguously and on a quantitative basis the sources that may not be in compliance. A related issue is the difficulty
Table I. Important Air Pollutants

<table>
<thead>
<tr>
<th>Pollutant</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sulfur Dioxide</td>
<td>$SO_x$</td>
</tr>
<tr>
<td>Sulfur Trioxide</td>
<td></td>
</tr>
<tr>
<td>Sulfur Acid Mist</td>
<td></td>
</tr>
<tr>
<td>Nitric Oxide</td>
<td>$NO_x$</td>
</tr>
<tr>
<td>Nitrogen Dioxide</td>
<td></td>
</tr>
<tr>
<td>Nitric Acid Mist</td>
<td></td>
</tr>
<tr>
<td>Oxidants ($O_3$)</td>
<td></td>
</tr>
<tr>
<td>Total Particulate</td>
<td>Particulate</td>
</tr>
<tr>
<td>Visible emissions</td>
<td></td>
</tr>
<tr>
<td>Particle size distribution</td>
<td></td>
</tr>
<tr>
<td>Number of particles</td>
<td></td>
</tr>
<tr>
<td>Particle composition</td>
<td></td>
</tr>
<tr>
<td>Particulate sulfate</td>
<td></td>
</tr>
<tr>
<td>Particulate nitrate</td>
<td></td>
</tr>
<tr>
<td>Asbestos</td>
<td>HAP</td>
</tr>
<tr>
<td>Mercury</td>
<td></td>
</tr>
<tr>
<td>Beryllium</td>
<td></td>
</tr>
<tr>
<td>CO</td>
<td></td>
</tr>
<tr>
<td>Total non-$CH_4$ Hydrocarbon - organics</td>
<td>Organic</td>
</tr>
<tr>
<td>Specific hydrocarbons</td>
<td></td>
</tr>
<tr>
<td>Polychlorinated biphenyls</td>
<td></td>
</tr>
<tr>
<td>Polynuclear organic matter</td>
<td></td>
</tr>
<tr>
<td>Reactive hydrocarbons (class)</td>
<td></td>
</tr>
<tr>
<td>Hydrogen sulfide</td>
<td>Odors</td>
</tr>
<tr>
<td>Mercaptans</td>
<td></td>
</tr>
<tr>
<td>Ammonia and amines</td>
<td></td>
</tr>
<tr>
<td>Organic acids</td>
<td></td>
</tr>
<tr>
<td>Aldehydes</td>
<td></td>
</tr>
<tr>
<td>Odor (human perception)</td>
<td></td>
</tr>
<tr>
<td>Hydrogen Chloride</td>
<td>Halogens</td>
</tr>
<tr>
<td>Chlorine gas</td>
<td></td>
</tr>
<tr>
<td>Hydrogen fluoride</td>
<td></td>
</tr>
<tr>
<td>Copper</td>
<td>Elements and other</td>
</tr>
<tr>
<td>Zinc</td>
<td></td>
</tr>
<tr>
<td>Boron</td>
<td></td>
</tr>
<tr>
<td>Tin</td>
<td></td>
</tr>
<tr>
<td>Lithium</td>
<td></td>
</tr>
<tr>
<td>Chromium</td>
<td></td>
</tr>
<tr>
<td>Vanadium</td>
<td></td>
</tr>
<tr>
<td>Manganese</td>
<td></td>
</tr>
<tr>
<td>Selenium</td>
<td></td>
</tr>
<tr>
<td>Arsenic</td>
<td></td>
</tr>
<tr>
<td>Phosphoric acid mist</td>
<td></td>
</tr>
<tr>
<td>Cadmium</td>
<td></td>
</tr>
<tr>
<td>Lead</td>
<td></td>
</tr>
</tbody>
</table>
of associating source emissions of specific types with ambient levels in a cause and effect relationship. This is exemplified by the recent findings regarding ozone levels in the eastern part of the country where ozone standards are exceeded in rural as well as urban locations yet abatement strategies concentrate on urban traffic restrictions (ref. 2). The development and application of techniques for identifying sources of air pollution and advancing the understanding of the atmospheric chemistry and transport of pollutants are important areas of study. This is particularly true as federal standard setting shifts from gross pollutant loadings to control of specific pollutants in a multimedia (air, water, land) context (ref. 3).

This report reviews efforts to date on the source identification problem with a focus on the use of trace elemental constituents in the suspended particulate matter. A practical source identification system under development at NASA Lewis Research Center which is based on time and wind direction resolved sampling of ambient particulates and comparison of observed elemental constituents in the samples with source signatures will be discussed.

Techniques for air pollution source identification can utilize both the physical and chemical properties of particulates and gases emitted from sources or materials added intentionally for the purpose. The techniques can rely upon (1) spectral absorption, emission or scattering properties observed from a distant position (i.e., remote sensing), or (2) direct in-situ sampling at various locations. The techniques are variously applicable to global, regional or local distance scales dependent upon the optical resolution and sensitivity available.

The scope of this paper does not permit a comprehensive discussion of all the applicable techniques but some key examples will be covered.
III. TECHNIQUES

(1) Remote Sensing

Remote sensing techniques in which NASA has played a pioneering role from both satellite and aircraft platforms have been used to demonstrate the method's potential for source identification. One striking example is described by Lyons (refs. 4 and 5) from LANDSAT 1 (formerly called ERTS-1, Earth Resources Technology Satellite) imagery of the southern half of Lake Michigan taken with band 6 (0.7-0.8μm) of the Multispectral Scanner. The contrast of the plume from seven sources near Gary, Indiana against the water was sufficient to delineate them and also downwind nucleation of clouds from particulates in the plumes is well defined. Similar imagery has delineated the "Lake Breeze" circulation that influences shoreline air quality. LANDSAT-1 imagery over land areas has also revealed plumes (ref. 6). The major limitations are the need for relatively cloud-free observation and sufficient contrast to the background but it has potential for rapid large area coverage. A wide range of remote sensing instruments, including passive correlation spectrometers, interferometers, radiometers, and active laser systems are under development (ref. 7) and their potential applicability to global and regional energy-related pollution problems has been reviewed (refs. 8 and 9). Of special interest is the planned launch of the Nimbus G satellite dedicated to atmospheric pollution monitoring and studies of the use of remote sensors aboard aircraft (ref. 10).

(2) In-Situ

In-situ techniques are the most highly developed of the two general approaches to pollution monitoring. The techniques for sampling and analysis of atmospheric pollutants have been comprehensively reviewed (ref. 11) and the
three volume monograph edited by Stern (ref. 12) covers all aspects of the problem. For source identification, tracers can be injected into the source or the pollutant gases or particulates themselves can be utilized.

(2a) **Injected Tracers**

Injected tracers have included fluorescent particles or inert gases like sulfur hexafluoride which are easily detected at very low concentrations (ref. 13). These techniques have been extensively used for the study of plume dispersion characteristics and sulfur hexafluoride appears feasible for distances up to 100 km from the source using fixed, mobile or airborne platforms (refs. 14 and 15). The uses of these methods are limited because source injection is required.

(2b) **Pollutants as Tracers**

The use of the pollutants themselves for source identification is applicable providing the pollutants are specific enough for the source in question.

(i) **Gases**

Carbon monoxide has been studied by Stevens and coworkers (ref. 16) to estimate the respective magnitude of man-made and natural sources of the compound. This was done by comparison of the average isotopic composition of atmospheric CO with that of a CO species whose production rate is known and thus derive the production rate of other species. $^{18}O/^{16}O$ and $^{13}C/^{12}C$ ratios were measured and $^{18}O$ enrichment and $^{13}C$ depletion determined with respect to accepted oxygen and carbon isotopic references. In this way the total natural sources of CO in the northern hemisphere were estimated to be ten times the man-made ones.
On a regional and local scale sulfur dioxide has been utilized extensively to identify fossil fuel burning sources. As one example, the impact of generalized sources in New York City and local power plants on suburban Long Island has been delineated over distances >100 km (ref. 17). Aircraft carrying appropriate in-situ instrumentation have been used to follow the composition and hence the chemistry of plumes to 80 km downwind (refs. 18, 19, and 20). Variations of sulfur isotopes have also been utilized (refs. 21 and 22).

Mercury vapor as a tracer has been used extensively as a tool for geological exploration and has been observed in plumes from coal-burning power plants, municipal incinerators, and chloralkali plants among others, utilizing helicopter platforms (ref. 23).

(ii) **Particulates**

Airborne particulate matter is ubiquitous in the environment. The chemical composition and physical character; i.e., particle shapes and size distributions relate closely to its origins. A monograph on the subject has been edited by Hidy (ref. 24). It has been recognized for some time that characterization of particulate matter in the ambient air can be utilized for air pollution source identification. Highly developed are the techniques of chemical microscopy articulated by McCrone (refs. 25 and 26). The application of sensitive trace element analytical techniques, particularly non-destructive multielement methods such as neutron activation and X-ray fluorescence (refs. 27, 28, and 29), have permitted the study of as many as 50 or more trace constituents (refs. 30 and 31).
Generally, hi-volume filtration or impaction techniques have been used for sampling. This requires the use of high purity substrate materials. The cellulose filter, Whatman #41, has been found most suitable (refs. 32 and 33) though not perfect because of its hygroscopic properties. This is not insurmountable if proper precautions are taken. Our own studies in Cleveland have shown that there is no statistically significant difference between Whatman 41 and the standard glass fiber filter for determining total mass concentration of particulates with hi-vol samplers (ref. 34). For routine network operation protection of the high purity filters is required to avoid inadvertent contamination by field personnel. A filter holder has been described which adequately solves the problem (ref. 35). Impaction techniques have used coated polymer films with success (ref. 36). Typical elemental concentrations in urban ambient particulates are shown in Figure 1. Trace metals from fuel combustion, incineration and industrial emission sources have been discussed by Lee and von Lehmden (ref. 37). They also emphasize the source implications of particle size. Those elements (e.g., Pb, V) associated with submicron sizes derive from the combustion of organically associated materials (petroleum products) while those associated with larger sizes (e.g., Fe and Mg) are from inorganic constituents of coal or industrial materials. Much of the recent literature emphasizes the use of trace elements for source identification purposes. Friedlander (ref. 38) formulated and applied the theory of source-receptor chemical element balance to the sources of Pasadena, California aerosol. The basic assumption is that each type of source emits a characteristic series of elements. The characteristic series were determined from an inventory of source emissions. Comparisons were made with ambient data. Sea salt, soil, auto exhaust, fuel oil, fly ash and cement dust contributions were estimated and accounted for >72% of the aerosol. Winchester's review of his own group's pioneering contributions
(ref. 39) cites similar approaches in the Chicago area using multielement data. They defined coal, coke and oil combustion (As, Cr, Sn, Ti, Ni, V) and iron and steel industry (Fe, Mn, Cu, Zn) components and found that Pb and Cd could be accounted for from gasoline and coal. Weslowski and his colleagues (ref. 40) examined the San Francisco Bay area aerosol and separated it into a marine aerosol group, a soil derived group (Fe, Sc, Th, Co, Al, Mn, Sm, Ce, Cr, Eu, La) which accounted for most of the aerosol, unlike the Chicago area, and a pollutant group (Sb, Se, Zn, Br, Hg). This was based on the correlation approach used by Winchester's group for the study of Gary, Indiana (ref. 41). The Gary study examined pairwise correlations and found Sc-Co and Zn-Sb to be in the same proportion at all sampling stations and that other pairs, e.g., Fe-Co and Ca-Mg, had single station anomalies. Correlation clusters were defined suggesting that Zn-Sb-Cr and Sc-Co-Th-Ce could be used to trace aerosols for long range transport. Other studies of a similar type were conducted in the Ivory Coast (ref. 42), in Paris (ref. 43), and we, together with the City's Division of Air Pollution Control, have conducted extensive studies in Cleveland (refs. 44 and 45).

In our study 750 samples were taken at 18 sites (Fig. 2) over a period of 15 months. This included some mass size distributions (ref. 46) and also some trace element size distribution measurements not yet interpreted. Since Cleveland is typical of Great Lakes industrial cities and the large size of our data base enables us to consider the influence of wind direction, the results will be discussed in some detail.

In the data interpretation use was made of standard statistical techniques such as pairwise correlation analysis, in addition to cluster analysis algorithms, and pollution-rose plots. Examples are given of the appli-
cation of these techniques to the sorting of elements into related groupings as well as to the effort to distinguish geographic sites by their specific elemental characteristics augmented by enrichment factors in the ambient relative to earth crustal distributions typifying this region for the same elements.

Our first attempts are based on 6 selected days listed in Table II. Days 1, 2, and 6 are typical of the region. Days 3, 4, and 5 were less typical in that it was part of a weather system which was almost unchanged for 8 days. Except for Be, Si, Cd, Bi, and Pb which were done by emission spectroscopy, analyses were made by Instrumental Neutron Activation Analysis. Most of the work was done at the LeRC Plum Brook reactor (ref. 31). Thirty-four elements are considered: Na, Al, Si, Cl, Sc, Ti, V, Mn, Fe, Co, Cu, Zn, As, Se, Br, Ag, Cd, Sn, Sb, Cs, La, Sm, Eu, Tb, Dy, Yb, Lu, Hf, Hg, Pb, Bi, and Th, and cumulative uncertainties at the one sigma level are estimated at about ±25%.

Table III gives estimated pairwise correlation coefficients for 5 pairs of elements. Co-Sc, Sc-Th and Zn-Sb are tabulated because they had the highest values in the study by Dams, et al. (ref. 41) in their study of Gary. Cs-Rb is included because it had the highest value in Cleveland on March 1, 1972. The final pair, Br-Pb, has received attention recently as an indicator of the impact of the automobile on air quality (ref. 47 and 48). It should be emphasized that there are differences in the data bases used for estimating the correlations. The first 2 columns are correlations across all the monitoring sites in Cleveland for the days shown (see Table II). The Gary values (ref. 41) are for 25 sites on one day while the San Francisco values (ref. 40) are for 9 sites on one day.
Table II. Meteorology for the 6 days

<table>
<thead>
<tr>
<th></th>
<th>3/1/72</th>
<th>5/15/72</th>
<th>5/18/72</th>
<th>5/21/72</th>
<th>5/24/72</th>
<th>6/8/72</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average temperature (°F)</td>
<td>60</td>
<td>59</td>
<td>61</td>
<td>64</td>
<td>65</td>
<td>66</td>
</tr>
<tr>
<td>Visibility</td>
<td>fog,haze</td>
<td>fog,haze</td>
<td>fog,haze</td>
<td>fog,haze</td>
<td>fog,haze</td>
<td>clear</td>
</tr>
<tr>
<td>Precipitation (in.)</td>
<td>0.20</td>
<td>0.66</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ave. pressure (in.)a</td>
<td>28.84</td>
<td>28.88</td>
<td>29.18</td>
<td>29.06</td>
<td>29.09</td>
<td>29.07</td>
</tr>
<tr>
<td>Resultant direc. (N = 0°)</td>
<td>190°</td>
<td>200°</td>
<td>0°</td>
<td>0°</td>
<td>20°</td>
<td>230°</td>
</tr>
<tr>
<td>Resultant speed (mph)b</td>
<td>15.3</td>
<td>6.7</td>
<td>3.8</td>
<td>3.1</td>
<td>4.8</td>
<td>10.2</td>
</tr>
<tr>
<td>Average speed (mph)b</td>
<td>15.5</td>
<td>9.1</td>
<td>5.6</td>
<td>6.3</td>
<td>7.5</td>
<td>10.5</td>
</tr>
<tr>
<td>Missing sample (Site #)</td>
<td>10</td>
<td>1</td>
<td>--</td>
<td>--</td>
<td>14</td>
<td>4</td>
</tr>
</tbody>
</table>

a Station elevation - 805 feet mean sea level
b Resultant speed = vectorial averaged amplitude; average speed = scalar average

Table III. Pairwise correlation coefficients

<table>
<thead>
<tr>
<th></th>
<th>Cleve</th>
<th>Cleve</th>
<th>Gary</th>
<th>S.F.</th>
<th>Ivory</th>
<th>Paris</th>
<th>Cleve</th>
<th>Cleve</th>
<th>Cleve</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3/1</td>
<td>5/21</td>
<td>(ref 41)</td>
<td>(ref 40)</td>
<td>(ref 42)</td>
<td>(ref 43)</td>
<td>Site 1</td>
<td>Site 12</td>
<td>Site 20</td>
</tr>
<tr>
<td>Co-Sc</td>
<td>.20</td>
<td>.08</td>
<td>.96</td>
<td>.98</td>
<td>.25</td>
<td>.92</td>
<td>.59</td>
<td>.07</td>
<td>.95</td>
</tr>
<tr>
<td>Sc-Th</td>
<td>.89</td>
<td>.86</td>
<td>.92</td>
<td>.97</td>
<td>.91</td>
<td>.45</td>
<td>.89</td>
<td>.77</td>
<td>.99</td>
</tr>
<tr>
<td>Zn-Sb</td>
<td>-.03</td>
<td>-.09</td>
<td>.91</td>
<td>-</td>
<td>-</td>
<td>-.30</td>
<td>.54</td>
<td>.02</td>
<td>.21</td>
</tr>
<tr>
<td>Cs-Rb</td>
<td>.98</td>
<td>-.52</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-.83</td>
<td>.87</td>
<td>.99</td>
<td></td>
</tr>
<tr>
<td>Br-Pb</td>
<td>.60</td>
<td>.87</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-.81</td>
<td>.67</td>
<td>.76</td>
<td></td>
</tr>
</tbody>
</table>

Table IV. Enrichment factors relative to earth crustal abundances normalized to Si

<table>
<thead>
<tr>
<th></th>
<th>Na</th>
<th>V</th>
<th>Zn</th>
<th>Cd</th>
<th>Eu</th>
<th>Hg</th>
<th>Al</th>
<th>Cr</th>
<th>As</th>
<th>Sn</th>
<th>Tb</th>
<th>Pb</th>
<th>Si</th>
<th>Mn</th>
<th>Se</th>
<th>Sb</th>
<th>Dy</th>
<th>Bi</th>
<th>Hf</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>200</td>
<td>400</td>
<td>2</td>
<td>200</td>
<td>1</td>
<td>6</td>
<td>200</td>
<td>200</td>
<td>0.2</td>
<td>2000</td>
<td>1</td>
<td>4</td>
<td>4000</td>
<td>1500</td>
<td>2</td>
<td>2000</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>6</td>
<td>200</td>
<td>200</td>
<td>0.2</td>
<td>2000</td>
<td>1</td>
<td>4</td>
<td>4000</td>
<td>1500</td>
<td>2</td>
<td>2000</td>
<td>1</td>
<td>4</td>
<td>4000</td>
<td>1500</td>
<td>2</td>
<td>2000</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>100</td>
<td>500</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Ivory Coast study was performed at 20 sites over two 5-day sampling periods while the Paris study was conducted over 8 separate days at 5 sites. The final 3 columns are for 3 specific sites in Cleveland. Site 1 is just east of the industrial valley, site 12 is in a residential neighborhood with some industry to the northeast and site 20 is at the shore of Lake Erie with some light industry to the south. These correlation values are for a single site over an entire year with about 30 to 40 values entering each calculation. Table II is included because this seems to be the easiest and most direct method of reducing data from different sources to a common format which then allows comparisons to be made. However, in light of the relatively small size of the data sets coupled with the rather large percent errors in each measurement, we wish to remind the readers of the necessary caution regarding the interpretation of such information as has been noted by Bogen (ref. 49).

Viewing pairwise plots of the data can often be instructive. Figure 3 shows such a sample plot of lead vs. bromine taken from site 17 for the entire study period. With the exception of 2 points, the linearity of the data is rather good considering the uncertainty in each measurement. The mean ratio is 0.31 which is well within the range of values reported for Br/Pb arising from automotive emissions. Another way of looking at this data is to plot all the values for the 6 days listed in Table I while retaining the information concerning the location at which the data was generated. This is shown in Figure 4. The numbers on the plot are the sites corresponding to Figure 2. It becomes fairly apparent that site 13 which is the only site at ground level and is to the south (generally upwind) from the nearest source of traffic has consistently low values. On the other hand, site 17 which is on the roof of a fire station (could the Pb and Br be coming from inside the building?) and
at the intersection of 2 heavily travelled commercial streets has consistently high values.

A number of efforts have been made to extract groupings (clusters) from trace element data sets in order to differentiate among various classes of sources (e.g., earth's crust, industrial activity) or to further differentiate between specific sources (e.g., auto emissions, powerplants).

One approach is to view this as a problem in pattern recognition amenable to the techniques of cluster analysis. This conceptual approach has been tried by at least 3 other authors. The logic used by Belot et al. (ref. 43) is very similar to the logic of our analyses. However, the work by the other two investigators (refs. 40 and 41) was performed in the early stages of the development of clustering theory. Unfortunately, they chose an algorithm developed for the social sciences which is not applicable to the problem at hand. Thus, we will restrict our comments in this section to a comparison with Belot's results for Paris.

The computer software routines were developed at Lawrence Radiation Laboratory (ref. 50). Figure 5 is a mapping or representation in two-dimensions of the spatial interrelationships amongst the elements for March 1, 1972 as a function of all the measurements (15 dimensions) taken that day throughout the city. Such a visual display is helpful, but the information presented is incomplete and may be distorted. While it is true that the quantitative techniques of cluster analysis may not generate unique results, repeated analyses of the data for each of the 6 days listed in Table II showed the following consistent features: (1) Sb, Hg and Cu are each unique in the spatial variation of their concentrations. The Cu values are suspect because of the possibility of sample contamination from the hi-vol motors (ref. 51). The distinguishability of Sb and Hg was observed by Belot in Paris. (2) Fe, Cr, and Co are
generally grouped together. This is quite reasonable in a location such as Cleveland, where there are large iron and steel works. Belot in Paris found this same grouping but with Sc as a 4th constituent. (3) Sc, Al, Si, Ti, V, La, Sm, Eu, Dy, Th, Pb and Br are always found grouped together. In addition, these elements have relatively little variation from site to site in their concentrations relative to TSP. This latter feature was noted by Dams et al. in their study of Gary. (4) Pb and Br, while not separable quantitatively by numerical clustering techniques from the other 10 elements listed above, are seen by visual inspection to always be a bit apart from the others and relatively close to each other as in Figure 5.

An instructive complement to the clustering approach has been applied by a number of other workers (ref. 52) and consists of looking at the relative enrichments of the elements as compared with their relative abundances in the local rock and soil. For this purpose we chose to use the table of abundances of Mason (ref. 53) and normalize by setting Si = 1. Table IV contains the average enrichment factors for March 1, 1972. We note that the 3 singular elements cited above (Sb, Hg, Cu) are all significantly enriched. This is consistent with their being related to strong localized sources. The set Fe-Cr-Co indicates some possible enrichment. The uncertainty of the data makes this finding non-significant by itself, but is consistent with our earlier speculation associating them with large industrial sources which would tend to be sources of Si as well. Within the large grouping of 12 elements discussed previously, the Pb and Br stand out very distinctively with their very large and similar enrichments. The other 10 elements are found at essentially the same relative abundances as in the earth's crust. This could well indicate the superposition of a natural background and the emissions from numerous coal burning sources.
Figure 6 is the same type of mapping as in Figure 5, but is directed to grouping sites based on the concentrations of 34 elements at each site. While the mapping does seem to group sites, as for example all the U's (site 21) are together, this separation was not borne out when a quantitative cluster algorithm was applied. It is possible that the sites may not be distinguishable on the basis of this limited set of data or else the clustering algorithm may not be appropriate.

With our large data base it is possible to simultaneously look at both space and time distributions. If we assume that the major variation from one sampling day to another is meteorology induced, we can plot all the data for a single element (over 750 values) on a single map as a function of the wind. Figure 7 shows 16 log-polar plots for scandium. Each plot is centered at the site at which the data was generated. Each radial line is proportional to the average concentration observed when the vector resultant wind was inward along the radial line. The data for sites along the lake front was sorted on the basis of wind information from a lakefront airport while the rest of the data was sorted using Hopkins Airport wind data. The main drawback in attempting emission source identification by triangulation with this data set is the 24-hour sampling time, which is long relative to the duration of the wind's directional stability. We have recently developed new sampling instrumentation to overcome this problem (ref. 54) but have not yet had the time to use this new equipment to do large-scale monitoring. It is still possible, however, to draw some inferences from the spatial distribution of an element with even as little personality as scandium. For example, at the most westerly site the concentrations are relatively low and fairly uniform with respect to wind, but still perceptibly larger when the wind is from the easterly direction. This pattern is more pronounced as we look at sites closer to the region of heavy
industry along the river. Thus this soil-derived element is also present in industrial emissions. Figure 8 shows a similar rose for antimony which is uniquely related to industrial operations especially at sites 6, 12, and 13.

We have indicated an approach to analysis of a large elemental concentration data set taken at numerous sites over an extended period of time. Examples presented have emphasized the use of multiple approaches including pairwise correlation statistics, selected plotting of the data, cluster analysis, comparative enrichment studies and the use of wind transport information. Despite the large variability within the data and the considerable accuracy and precision errors it is still possible to extract useful information. We have shown that Br and Pb, as a pair, have fairly uniform distributions, are significantly enriched in TSP relative to Si when compared with their earth crustal abundances and can be used to distinguish between sites. The rare earths generally exhibit the same distributional pattern as Si, Al, and V and do not show any particular enrichment. Fe, Cr, and Co show some slight indication of enrichment and have their own distributional pattern. Sb, Hg, and Cu are each unique in their distributional patterns. We have also shown how a pollution-rose can be used to visualize a large body of data. At the present time we are extending these methods to the entire data set to form the basis for a practical air pollution source identification system useable by enforcement agencies.

(3) **Source Identification Systems**

Trace element measurements in the ambient air provide a practical approach for enforcement agencies to use in quantitatively identifying sources on a routine basis.

The State of California has established a network of 14 stations in 14 cities around the state (ref. 36) specifically designed to sample aerosols with
two-stage Lundgren impactors. Particle size ranges of 20 to 3.6 micrometers (µm) to 0.65µm, and 0.65 to 0.1µm are measured. It is now being operated routinely using Charged Particle X-ray Fluorescence (CPXF) analysis which is completely automated, rapid (1000 samples/day) and inexpensive. The main emphasis in this operation is to separate natural from man-made sources on a regional scale and to give detailed information on the precise source of the man-made contribution. Two-hour intervals can be resolved on the coated Mylar impactor sample strips allowing the diurnal characteristics to be studied (ref. 55) as required. Most attention to date has been placed on separating automotive from non-automotive pollutants.

NASA Lewis together with Cleveland's Division of Air Pollution Control has been developing a source identification system for use within industrial cities with complex source distributions. A comprehensive operational test and evaluation with user groups is now being launched. Figure 9 is a diagram of the system showing all its various components and their interrelationships. The key to the system is a sampler which we call Air Scout which automatically collects suspended particulate matter as a function of both wind direction and time, shown in Figure 10 (ref. 54). Filtration onto Whatman-41 mounted in 2 x 2 slides is utilized for the directional samples and glass fiber is used for the total suspended particulate measurement. Automated CPXF analysis of the exposed slides will be used and the data interpreted by the techniques discussed in the preceding section together with source signatures (training sets) using interactive graphics.

The Regional Air Pollution Study (RAPS) in St. Louis, which is a major program of the EPA for validating air quality models, studying sources, fates and atmospheric chemistry of pollutants, includes trace element monitoring for source identification purposes. Dzubay and Stevens (ref. 56) have described the
"virtual impactor" sampling device which collects in 2 size ranges (2-10μm and <2μm) and the use of the rapid automated X-ray induced X-ray fluorescence analysis system being used in the work.

IV. CONCLUDING REMARKS

A great deal of research has been done on the characterization of air pollutants and identifying the sources with which they are associated. It is significant that these research results are now being developed into operational systems that air pollution regulatory agencies can use in their efforts to define and control air quality. Unfortunately, their use is not widespread and most local agencies are relying on inadequate monitoring techniques.

The needs for further research and development can be considered in four major categories: (1) pollutant chemistry; (2) monitoring technology; (3) modelling; and (4) multimedia implications.

In the area of pollutant chemistry, the various reactions taking place in the atmosphere once pollutants are emitted are poorly understood, particularly the interaction of the gaseous pollutants with the particulates, redox processes, etc. In the Great Lakes region the major emphasis should be on coal burning and industrial sources rather than the photochemical problems dominant in areas blessed with more sunshine. However, long-range transport of oxidants does appear to impact the region and their origin needs better definition. Little is known about the inorganic compounds present in particulate matter and how they are distributed with regard to particle size and even less is understood about the organic constituents. This is significant in evaluating potential health effects.

In the area of monitoring technology there has been a substantial effort on instrumentation for various constituents and significant improvement in
specificity and reliability particularly for the gases. For the particulates the state of the art is improving only slowly and there is opportunity for innovative approaches. It is unfortunate that local agencies are not attempting to monitor trace elements routinely using the inexpensive automated non-destructive techniques now available, especially since the data obtained contains much information on source identity and would complement more conventional surveillance methods and may be particularly useful in defining non-point source problems.

A nagging problem needing attention is the definition of the relationship between the point measurement commonly used in urban networks and the relationship of that point measurement to the area to which the data is applied. The presence of buildings and the specific micrometeorology of the site often perturb measurements. For these reasons the continued development of long path and laser ranging monitoring devices to provide spatially integrated measurement is necessary. The continued development and application/evaluation of remote sensing techniques should be encouraged since their potential contribution in area-wide monitoring is great.

Monitoring technology must take a systems approach and meet cost effectiveness criteria if advances are to be adopted. We must look beyond just the instrumentation to its deployment and how its data will be put into a useable form as information.

In the area of modelling, continuing efforts must be made to improve the models relating source strengths to ambient levels since they are the key to effective abatement strategies and can, in principle, be used to improve the understanding of point measurements and reduce surveillance costs. In the Great Lakes area the lake land interface provides significant complications due to
the Lake Breeze circulation which may make shore line areas far less attractive from an air quality standpoint for coal burning power plant siting in spite of the accessibility of cooling water. Models are also needed which provide a means for analyzing observed air quality trend data in terms of the influence of random meteorological effects (precipitation, wind speed, mixing height, etc.) as opposed to the effect of source emission reduction. This is necessary if the true effectiveness of abatement programs are to be assessed.

The multimedia implications of air pollution sources need much greater attention. As has been documented in a few studies, inputs to the Great Lakes from various types of sources in this region may be significant to water quality through fallout and rainout of trace elements and nutrients. This is especially important if the so-called intermittent control strategies for coal burning power plants are going to win acceptance.

Many of the above problems are already receiving attention and inter-agency cooperation continues to be an important mechanism for progress.
References


Figure Captions

Figure 1. Typical elemental concentration of urban ambient particulates (Cleveland 1971-1972), geometric mean ±1 standard deviation. Note concentrations are on a logarithmic scale.

Figure 2. Air Monitoring Network. The Cuyahoga River (industrial valley) is in the center, flowing north (from bottom to top) into Lake Erie. The circles indicate air quality monitoring sites. Trace element analyses were performed on samples collected at the 16 numbered sites.

Figure 3. Bromine vs. lead data from site No. 17 for the entire year. The mean Br/Pb = 0.314 and the correlation coefficient = 0.454.

Figure 4. Bromine vs. lead data from all sites for the 6 days listed in Table II. Each data point is represented by a number (corresponding to the site numbers of Figure 2) which identifies where it was generated.

Figure 5. Representation of the data by Non-Linear-Mapping—the elemental interrelationships are represented in 2-space as a function of the spatial distribution of their concentrations on 3/1/72 (1st day in Table II).

Figure 6. Representation of the data by Non-Linear-Mapping—the interrelationships of all the air quality samples for the 6 days listed in Table II as a function of their elemental distributions. The sites are coded by alphabetical sequence (i.e., A = 1, B = 2, ... U = 21) and the days by numerical correspondence to their listings in Table II (i.e., C2 = site No. 3 on 5/15/72).

Figure 7. Pollution-Roses for Scandium. The entire data set is used to display the 16 mean concentrations at each site as a function of the resultant (vector average) wind direction averaged over the sampling period. The graphs are polar-logarithmic with the outer scale = 10 ng/m³ and the inner scale = 0.10 ng/m³. The wind direction is toward the center.

Figure 8. Pollution-Roses for Antimony obtained from the data set as in Figure 7. Outer scale = 10⁴ ng/m³ and the inner scale = 10² ng/m³.

Figure 9. Diagram of Source Identification System.

Figure 10. Air Scout arrangement and modular chassis concept.
SOURCE IDENTIFICATION SYSTEM

SOURCE SAMPLING
- NEARBY
- IN STACK

INVENTORIES, PUBLISHED DATA

PRELIMINARY SIGNATURES

CATALOG OF SOURCE SIGNATURES
- NATURAL
- AUTO
- AIRCRAFT
- FLY ASH (OIL, COAL)
- STEEL
- SECONDARY SMELTER
- PHOSPHOR PLANT
- CATALYST MFG

LAB ANALYSIS DATA SOFTWARE

PRESENT DATA
PRELIMINARY SIGNATURES

SOURCE IMPACT FROM WIND DIRECTION AND TIME RESOLVED DATA

AMBIENT AT A SINGLE STATION
A
B
C

AMBIENT MONITORING NETWORK

BIG OR HOT SPOT IMPACTS FROM LUMPED NETWORK

DISPERSION MODEL AND METEOROLOGY

SOURCE IDENTITY
- ENFORCEMENT
- SPECIFIC IMPACTS (HEALTH OR WELFARE HAZARDS)
- GENERAL IMPACTS (e.g., AUTO vs. STATIONARY SOURCES)

Fig. 9.
Fig. 10.