

JONATHAN BORAK & COMPANY, INC.

Specialists in Occupational & Environmental Health

September 15, 2006

Ms. Denise Kennedy
Holland & Hart
555 17th Street, Suite 3200
Denver, CO 80202

Dear Denise:

Last week, you asked me to review the EPA's *Provisional Assessment of Recent Studies on Health effects of Particulate Matter Exposure* to determine whether there were additional studies that were specifically relevant to your clients' concerns. In so doing, I identified eight papers that are discussed below. Following the discussion of those individual papers, I have provided a brief Discussion.

Peel JL et al: Ambient air pollution and respiratory emergency department visits. *Epidemiology* 16:164-173, 2005.

Sinclair and Tolsma: Associations and lags between air pollution and acute respiratory visits in an ambulatory care setting: 25-month results from the Aerosol Research and Inhalation Epidemiological Study. *J Air Waste Manage Assoc* 54:1212-1218, 2004.

These two studies from Atlanta used the identical exposure database and reached opposite conclusions. They are discussed on page 21 of the *Provisional Assessment*, and the study by Peel et al. is included in Figure 3, page 18. The exposure data in both studies are from a 25-month study (8/01/98-8/31/00) from a single monitoring station in central Atlanta. A variety of particulate and other exposures were monitored daily, including PM_{10-2.5} measured with a dichotomous sampler^[1]. It should be noted that the PM_{2.5} levels in Atlanta were relatively high (mean = 19.2 µg/m³, SD = 8.9) while PM_{10-2.5} levels were relatively low (mean = 9.7 µg/m³, SD = 4.7).

^[1] Details of the exposure assessment methods are described in van Loy et al: The Aerosol Research and Inhalation Epidemiology Study (ARIES): PM_{2.5} mass and aerosol component concentrations and sampler intercomparisons. *J Air Waste Manage Assoc* 50: 1446-1458, 2000.

Ms. Denise Kennedy
September 15, 2006
Page 2.

Analyses were performed using single-pollutant models with both GAM and GLM. As described in the two papers, these two studies used identical analytical models.

The Peel study evaluated data for emergency department (ED) visits at 31 Atlanta hospitals. Cases were those with a diagnosis of a) asthma; b) pneumonia; c) upper respiratory infections (URI); d) COPD; and e) all respiratory disease. A total of 183,160 ED visits were categorized as respiratory disease. The Sinclair study evaluated out-patient visits at 10 Kaiser Permanente HMOs during the same time period. Visits were grouped as a) adult asthma; b) childhood asthma; c) URI; and d) lower respiratory infections (LRI). A total of 232,350 visits were categorized as respiratory disease. Based on the ICD-9 codes assigned to these groups, there were small differences between the categories that were included in the two studies; COPD was not considered by Sinclair.

The Peel study found no statistically significant association with $PM_{10-2.5}$; a significant association was noted for coarse PM and asthma, but only with a 6-day lag. On the other hand, associations were positive between URI and COPD for PM_{10} , NO_2 , and CO. The Sinclair study, by contrast, found significant positive associations between $PM_{10-2.5}$ and childhood asthma (0-2 day lag), URI (3-5 day lag), and LRI (3-5 day lag).

Sinclair noted that the "magnitudes of the significant risk ratios were notably weak" and Peel pointed out that "single-pollutant results are likely confounded, at least in part, by correlated pollutants" and that use of a centrally located monitor is likely to result in measurement errors. However, it is difficult to explain why such errors or biasing would differentially impact ED visits vs. HMO visits. Both studies excluded repeat visits on the same day by the same patient, but not repeated visits for the same illness. It might be expected that such "follow-up" visits would have been more common for HMO than ED patients.

In summary, these two studies seem to be "off-setting", but the differences between them are difficult to explain. Because their data reflect only 25 months, the total numbers of cases are relatively limited. Several methodological issues cloud the findings, notably the use of a single central monitor for coarse PM measurements and the reliance on single-pollutant models. I assume that Sinclair was not included in Figure 3 of the *Provisional Assessment* because it did not literally address either "hospital admissions" or "emergency department visits".

Metzger KB, et al: Ambient air pollution and cardiovascular emergency department visits. *Epidemiology* 15:46-56, 2004.

This study from Atlanta used the same exposure database described above for the Peel et al. and Sinclair & Tolsma studies to study cardiovascular diseases (CVD). The analyses were essentially identical to those in the Peel and Sinclair studies; multiple-pollutant models were used for a limited number of comparisons, but these did not include coarse PM. The study is discussed on page 21 of the *Provisional Assessment*.

The study evaluated data for emergency department (ED) visits at 31 Atlanta hospitals. Cases were those with a diagnosis of a) ischaemic heart disease (IHD); b) acute myocardial infarction; c) cardiac dysrhythmia; d) cardiac arrest; e) congestive heart failure; f) peripheral vascular and cerebrovascular disease, g) atherosclerosis; and h) stroke. There was an average of 55 ED visits per day for all CVD categories. Because the total number of cases was small, results for categories b), d), g) and h) were not reported.

No significant association was found between any or all of the CVD and PM_{10-2.5}. On the other hand, CVD was positively associated with PM_{2.5}, NO₂, CO, organic carbon, elemental carbon and oxygenated hydrocarbons. As seen in Figure 1 of the report, levels of PM_{10-2.5} had a negative, but not significant impact on CVD for lags of 1-7 days.

This study suffers from the same methodological limitations as those discussed for the Peel and Sinclair studies. The small number of cases included during the time that coarse PM was measured is of particular importance. Nevertheless, these data argue that in Atlanta, coarse PM has little or no adverse impact on CVD, in clear contrast to fine PM and other combustion-related particulates (e.g., EC and OC).

Slaughter JC et al: Association between particulate matter and emergency room visits, hospital admissions and mortality in Spokane, Washington. *J Expo Anal Environ Epidemiol* 15:153-159, 2005.

Slaughter et al. described an analysis of emergency department (ED) visits for respiratory conditions, hospitalizations for respiratory and cardiac conditions, and non-accidental mortality in Spokane from 1/1/95-12/31/99. Cardiac and COPD was tabulated for only individuals > 15 years old, while other endpoints apparently included all age groups. Over the 1825 days of the study, respiratory

Ms. Denise Kennedy
September 15, 2006
Page 4.

ED visits averaged 12.2 per day, respiratory hospitalizations averaged 7.3 per day, cardiac hospitalizations averaged 7.3 per day and mortality averaged 9 per day. The study is discussed on page 16 of the *Provisional Assessment*.

Exposure data included PM₁₀, PM_{2.5} and PM₁ which were measured hourly by means of Tapered Element Oscillating Microbalances (TEOM) ^[2] at a single monitoring site. PM_{10-2.5} was calculated by the difference method. CO was measured at five different sites, with data averaged and "standardized" to yield a single daily average that served as a marker of combustion-derived pollutants. Data were missing for 18.5% of PM₁ determinations, 6.7% of PM_{2.5} determinations and 6.9% of PM₁₀ determinations; missing data were estimated using "multiple imputations", but the specific imputation calculations were not described.

Data were analyzed using GAM and GLM models. There were 9 days (0.5%) when PM₁₀ concentrations were >100 µg/m³ due to dust storms, and analyses were performed with and without those days; no differences were noted.

There were essentially no associations between health outcomes and exposures to PM_{10-2.5} or PM₁₀: "No trend in the strength of association was seen for PM₁₀ or PM_{10-2.5} and all of the relative risks estimates were near unity". By contrast, stronger associations were found for CO, PM_{2.5} and PM₁, although none of the PM associations were statistically significant.

This study has particular significance because the airborne dust of semi-arid Spokane is expected to be largely crustal. Thus, the finding of a lack of association of dust fractions with respiratory and cardiac health outcomes seems especially favorable, especially because the only consistent association was with CO, regarded as a marker for combustion-derived pollutants. The limitations of this study include the use of a single monitoring station for PM measurements, the need for imputation to address missing data (it is not possible to determine the number of PM_{10-2.5} data that were imputed, but it was between 6.9% and 13.6%), and the relatively small numbers of cases considered in the analyses.

² As discussed in my letter of August 27, there is evidence that TEOM systematically understates levels of PM as compared to the Harvard impactor, a finding that suggests that studies utilizing TEOM for such studies likely leads to overestimation of the potency of measured particulate fractions. This is detailed in: Cyrus J et al: PM_{2.5} measurements in ambient aerosol: comparison between Harvard impactor (HI) and the tapered element oscillating microbalance (TEOM) system. *Sci Total Environ* 278:191-197, 2001.

Mar TF et al: An analysis of the association between respiratory symptoms in subjects with asthma and daily air pollution in Spokane, Washington. *Inhal Toxicol* 16: 809-815, 2004.

A second, smaller study was conducted in Spokane by Mar et al. Sixteen asthmatic adults and nine asthmatic children with asthma were followed for up to 1 year with daily diaries of symptoms and medication use, but not physical exams. No information was provided regarding their severity of disease. Diary records were correlated to exposure measures. This study is discussed on page 21 of the *Provisional Assessment*.

PM₁₀, PM_{2.5} and PM₁ were measured by TEOM as described in Slaughter et al (see above) and PM_{10-2.5} was calculated. No gaseous pollutants were considered.

For the adults, no significant associations were found between symptoms and any of the measures of PM, except 'runny nose' which was negatively related to PM₁₀. No associations were found between any of the PM measurements and lower respiratory symptoms.

For the children, a variety of positive associations were found. Cough was significantly associated with all of the PM measures for one or more lag days. Sputum was significantly associated with coarse PM with no lag, but not with lags of 1 or 2 days. No significant association was found between 'wheeze' or 'trouble breathing' and any of the PM measures. The least strong associations were with coarse PM: "When the relationship between any symptoms ... and PM metrics were analyzed ... the strongest associations were found with PM₁₀, PM_{2.5} and PM_{1.0}", not coarse particulate.

Surprisingly, the authors state that the associations between asthma aggravation and coarse particles "adds to the growing literature suggesting an association between this particle size and asthma aggravation". However, no significant associations were noted between coarse PM and asthma aggravation, i.e., there were no significant associations with wheeze, shortness of breath, or lower respiratory symptoms.

The authors note that their findings "differ from the results of Schwartz et al." ^[3]. They are also clearly different from the study by Rabinovitch et al, in which severely asthmatic minority children at National Jewish Hospital were directly

³ Schwartz J and Neas LM: Fine particles are more strongly associated than coarse particles with acute respiratory health effects in school children. *Epidemiology* 11:6-10, 2000.

Ms. Denise Kennedy
September 15, 2006
Page 6.

monitored for three years. That study, discussed in my earlier comments on the proposed NAAQS ^[4], has thus far been ignored by EPA. It is far more rigorous methodologically than the Mar et al. study in terms of both the exposure data and the health outcome data.

In summary, this is a very small study that ignores co-pollutants, relies on self-reported symptoms rather than direct examination of subjects, and yields results that are contradictory to those of key studies on children, asthma and air pollution.

Staniswallis JG et al: Temporal analysis of airborne particulate matter reveals a dose-rate effect on mortality in El Paso: Indications of differential toxicity for different particle mixtures. *J Air Waste Manage Assoc* 55: 893-902, 2005.

The study by Stanwallis et al. describes a mathematical reinterpretation of data from El Paso undertaken to test two hypotheses: 1) hourly average PM levels would better predict mortality rates than 24-hour average PM levels; 2) lacking data on PM composition, it is possible to distinguish fine vs. coarse PM exposures on the basis of wind speed at the time of PM measurements. The

⁴ For your convenience, I provide my earlier comments on this important study (Rabinovitch N et al: Effects of wintertime ambient air pollutants on asthma exacerbations in urban minority children with moderate to severe disease. *J Allergy Clin Immunol* 114:1131-1137, 2004:

"The children were students at a special school, operated at the National Jewish Hospital in Denver, which specifically enrolled children with chronic diseases including asthma. The school was located in a community where PM₁₀ is dominated by coarse particulate; during the study period, coarse particulate on average comprised 61.2% of PM₁₀.

"For two years, exposure data (including PM₁₀ and PM_{2.5}) were obtained from EPA monitors located 100 meters from the school. During the third year, particulate data were obtained from a community monitoring station located 2.8 miles from the school. Children were monitored for asthma symptoms, asthma exacerbations, twice-daily FEV₁ and peak flows, use of asthma medications, and URI events. School activities were not modified in response to pollution alerts 'so as not to bias any potential pollution effects'.

"Associations between air pollutants and asthma outcomes were found in simple models, but not in complex modeling that included all pollutants and time-dependent covariates such as URI events. Using the more complex model, no significant associations were observed between pulmonary function and PM₁₀. Asthma symptoms were significantly associated with ozone levels, but not PM₁₀ and no significant associations were noted between asthma exacerbations and PM₁₀. By contrast, URI symptoms were strongly associated with decreased pulmonary function, increased medication usage, asthma symptoms, and asthma exacerbations."

Ms. Denise Kennedy
September 15, 2006
Page 7.

study itself is highly mathematical; even in the context of statistically-demanding air pollution studies, this study is mathematically dense. It is discussed on page 17 of the *Provisional Assessment*.

The study was based on ambient exposure data from an area where: 1) composition of PM₁₀ is known to vary greatly from hour to hour; 2) PM₁₀ levels peak evenings, especially winter evenings when they can increase up to 4-fold; 3) on average, PM_{2.5} comprises only 25% of PM₁₀, but most PM₁₀ peaks are associated with PM_{2.5}. The daily average PM₁₀ levels ranged from 0.2-133.4 µg/m³.

The study utilized data from 1992-1995, a period with almost complete hourly PM₁₀ data from a centrally located monitoring station; measurement methods were not described. Data were not available for PM_{2.5} or PM_{10-2.5}, so the authors used wind speed as a surrogate: low wind speed (e.g., <6 mph, 2.7 m/sec) was assumed to indicate mainly fine PM, while high wind speed (e.g., >17 mph, 7.6 m/sec) were assumed to indicate predominantly coarse PM. Principal Component Analysis (PCA) was used to evaluate the association between PM levels and mortality.

The analytical results indicate that hourly PM₁₀ averages, rather than daily averages, were significantly associated with daily mortality rates lagged 3 days. Exposures during periods of high-wind speed were associated with a 10% lower risk of mortality compared to comparable levels of PM₁₀ exposure during periods of mid- or low-wind speed. These results argue that wind-blown dust (presumably coarse PM) is less harmful than urban dust (presumably fine PM).

This study is limited in several ways. It lacks direct measurement of PM₁₀ components, thus its conclusions are necessarily uncertain. Also the study size is relatively small. The authors stress this point: "This analysis cannot be used to form firm conclusions, because it uses a very small data set (one location, one monitoring site, small mortality counts, and only 4 years of exposure data). Based on these findings, however, they argue that PCA should be included in future studies as a means of better characterizing pollution-related health risks.

Villeneuve PJ et al: A time-series study of air pollution, socioeconomic status, and mortality in Vancouver, Canada. *J of Expo Anal Environ Epidemiol* 13:427-435, 2003.

The study of Villeneuve et al. is a complicated time-series analysis from Vancouver that uses different exposure data than the data set utilized by the

Ms. Denise Kennedy
September 15, 2006
Page 8.

same research group (Chen et al. 2004; Chen et al 2005; Yang et al. 2004) that were described in my August 27 report.

Exposure data, collected from 1/1/86-12/31/98, included daily $PM_{2.5}$ and $PM_{10-2.5}$ measured by TEOM, and PM_{10} calculated as the sum of fine and coarse PM (1995-1998) and $PM_{2.5}$, $PM_{10-2.5}$, PM_{10} , TSP and SO_4 measured every 6th day by dichotomous sampler (1986-1995). Gaseous pollutants were measured hourly and coefficient of haze was measured every 2 hours. These data were analyzed to determine associations with daily mortality rates in people >65 years old for all non-accidental deaths; cardiovascular disease; respiratory disease; and, cancer. In addition, individuals who died were categorized according to socioeconomic status (SES) based on mean family income levels for the Census Enumeration Areas in which they lived, and these categories were included in the analyses.

Analyses were performed with time-series models and step-wise regression was employed to determine the effects of co-pollutants. The percent changes in mortality rates were determined across SES categories.

Results of the analysis indicated a significant association between coarse PM levels and cardiovascular deaths, but only without any lag; no significant associations were found for cardiovascular deaths with lags of 1-3 days and no associations were seen for other death categories. For all causes mortality, there were no associations with $PM_{2.5}$, $PM_{10-2.5}$, or PM_{10} ; a significant positive association seen for TSP lagged 2 days. Increased mortality was associated more generally with NO_2 , CO and SO_2 . There were essentially no associations between air pollution measures and death due to respiratory disease.

The *Provisional Assessment* discusses this study on page 16 as providing support for an association between coarse PM and mortality. However, the authors disagree; they generally viewed their study as negative for particulates:

“The absence of observed effects for $PM_{2.5}$, PM_{10} , and $PM_{10-2.5}$ in Vancouver may in part be due to low ambient levels, and the availability of only 3 years of data with daily sampling.”

It is difficult to understand how particulate would be more important without a lag, given that most studies that found significant effects of exposure indicate that effects are greater only after specific lag times. In addition, the findings of an important predictive role for NO_2 , which confirms the findings of Burnett, Thurston and others discussed in my earlier comments to EPA, reinforces the argument

that single-pollutant models are not adequate for the analysis of PM-related health effects.

Lipfert FW et al.: Traffic density as a surrogate measure of environmental exposures in studies of air pollution health effects: Long-term mortality in a cohort of US veterans. *Atmospheric Environment* 40:154-169, 2006.

Lipfert et al describe an on-going study of mortality in a large group of US veterans who were first diagnosed with hypertension during the 1970s^[5]. This specific report is primarily concerned with the contributions to overall mortality of exposure to motor vehicle emissions, not PM. It is further complicated by the fact that 81% of the subjects were current or former smokers. This study is discussed on page 10 of the *Provisional Assessment*.

I have serious concerns about the exposure data used in this study. Although the current study does not describe measurement methods, previous reports by these authors defined coarse particulate as PM_{15-2.5}, not PM_{10-2.5}^[6] and in the current report, coarse PM levels (defined as PM_{10-2.5}) are reported for only seven years (1989-1996) of a 25-year study (1976- 2001). Thus most of this study lacks appropriate coarse PM data. In addition, the exposure data used in these analyses were annual averages of pollutants for individual counties, reflecting the county of residence of each subject at the time of his entry into the study^[7]; subjects were assumed to have resided at the same location for the duration of the study.

⁵ The nature of the cohort population is described slightly differently in different reports from this study. A 2000 report described "a national cohort of about 50,000 US veterans who were diagnosed as hypertensive" [Lipfert FW et al: The Washington University-EPRI veterans' cohort mortality study: Preliminary results. *Inhalation Toxicology* 12(Suppl 4):41-73, 2000], but the current study describes a "cohort of about 70,000 male US veterans (the Washington University – EPRI Veterans Cohort)... who were recruited in 1975 for a study of hypertension."

⁶ See: Lipfert FW et al: The Washington University-EPRI veterans' cohort mortality study: Preliminary results. *Inhalation Toxicology* 12(Suppl 4):41-73, 2000; and, Lipfert FW & Morris SC: Temporal and spatial relations between age specific mortality and ambient air quality in the United States: regression results for counties, 1960-97. *Occupational and Environmental Medicine* 59:156-174, 2002.

⁷ The study methods are not well described in the current report. In 2000, however, they were detailed as follows: "More detailed exposure information was thus obtained by averaging air quality data by year for each county of residence at the time of entry to the study... No data were available on personal exposure to air pollution; the assumption is made that the temporal patterns in countywide average ambient data based on residence at entry to the study are representative throughout each subject's life." Lipfert FW et al: The Washington University-EPRI veterans' cohort mortality study: Preliminary results. *Inhalation Toxicology* 12(Suppl 4) at p. 46-7.

Ms. Denise Kennedy
September 15, 2006
Page 10.

The significant finding highlighted in the *Provisional Assessment* was for PM_{10-2.5} in a single pollutant model; no significance was found for a multi-pollutant model and the association became negative (but not significant) when vehicle-derived exposures (measured in terms of "traffic density") were included. (Use of multi-pollutant models also eliminated the apparent significance of PM_{2.5} in this study).

In short, this report lacks methodological details, but companion reports provide evidence that the study analysis relied on crude exposures measurements (i.e., annual county averages) and appropriate coarse PM measurements were lacking for most of the study.

Discussion

The additional studies included in the *Provisional Assessment* provide essentially no additional insights about the toxicity of PM_{10-2.5}. The additional epidemiological studies are generally small and methodologically problematic. Most relied on central monitoring stations that would have failed to document the expected local variations of coarse PM, thus providing exposure estimates of only limited accuracy. A number included coarse PM data that were measured only every sixth day, that required unspecified "imputation" to address missing data, or that were available for only a small proportion of the study period. One study used exposure data that were annualized average levels, thereby ignoring the important effects of day-to-day variations and peak exposures. By contrast, the potential errors associated with that sort of averaging were described in a study from El Paso that demonstrated important differences between hourly averages (which were significantly associated with health outcomes) and daily averages that were not. It should be obvious that annual averages are even less informative and useful than daily averages.

In short, many of these additional studies suffer from important limitations of exposure data. As I wrote in my original report on the Proposed NAAQS, positive associations between PM_{10-2.5} and specific health outcomes have most often been reported in studies with weak or flawed exposure assessments. That view is merely enforced by these recent studies.

I am also struck by the tendency to 'cherry pick' studies and study data for presentation in both the *Provisional Assessment* and the PM_{10-2.5} NAAQS documentation. For example, the inclusion of the Mar report, which described a symptom diary study for a small number of asthmatics in a community with only limited exposure data, as contrasted to the failure to include the study by Rabinovitch et al., which performed detailed functional medical assessments

Ms. Denise Kennedy
September 15, 2006
Page 11.

over a three-year period for a larger number of severe asthmatics in a setting with extensive, rigorous exposure data.

Finally, and contrary to statements in the *Provisional Assessment*, these additional studies do not support the view that reported associations were "generally robust to alternative modeling strategies or consideration of potential confounding by co-pollutants". To the contrary, to the extent that multiple-pollutant models were used, the reported effects were decreased and/or became not significant in virtually every case.

In short, it remains my professional and scientific opinion that there is not sufficient scientific data to justify the proposed PM_{10-2.5} NAAQS.

Many thanks for this opportunity to be of assistance to you.

Yours truly,

Jonathan Borak, MD, DABT

Associate Clinical Professor of Epidemiology and Medicine
Yale University

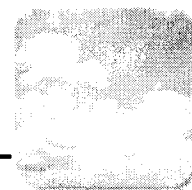
http://www.epa.gov/nerl/goals/pm/air_pm_receptor.html#intro

Last updated on Thursday, July 6th, 2006.



Exposure Research

Source Receptor Modeling NERL PM Research



- [Introduction¹](#)
- [Current Research²](#)
- [Recent Accomplishments³](#)
- [Related Web Sites⁴](#)

Introduction

Source Receptor models use measurement data on PM mass and composition to determine which sources are responsible for PM mass and the relative contributions of each source to the total mass. NERL is conducting research to develop measurement methods and modeling tools for source receptor analyses. NERL source receptor research also includes the application of measurements and models to understand relative source contributions to ambient PM levels. In this way, the NERL source receptor research program supports implementation actions by verifying emissions source inventories and identifying those sources that have a substantial impact on ambient PM levels so that effective control strategies can be developed. We are also extending source receptor techniques to personal exposure data to determine which sources are responsible for the PM that people actually breathe. This new application could provide important inputs for future NAAQS development and implementation activities.

Current Research

Tasks to Accomplish this work:

Task 9572: PM Source Apportionment/Receptor Modeling - PM Toxic Compounds (FY02-05). This task builds on previous PM source apportionment research with increased emphasis on receptor model development and application to micro-environments and human exposures to PM. Work continues on the application of radiocarbon sample analysis for estimating biogenic emissions contribution to ambient PM concentrations. The modeling products of this task address the goal of providing tools that are demonstrably suited to quantifying the impact of pollutant sources contributing to human exposure including CMB, UNMIX and PMF.

PI: Charles Lewis, Gary Norris

Task 15099: Source Apportionment/Other Data Analysis (FY03-05). This task is mainly concerned with the development of source apportionment models and their evaluation by means of application to PM ambient data from EPA's Supersites. Models of current interest are Chemical Mass Balance (CMB), and the second generation multivariate

models Unmix and Positive Matrix Factorization (PMF). The objective of this task is To deliver improved, documented, and tested receptor models for use by State and local air pollution staff, as tools for SIP development beginning 2005.

PI: Gary Norris

New Proposed Task: PM Source Apportionment -- Lab Analysis Support (FY05-07). The primary focus of this task is to provide laboratory analysis of field samples needed to perform source apportionment calculations. Possible analyses include XRF, IC, radiocarbon, computer-controlled scanning electron microscopy combined with individual-particle X-ray analysis (CCSEM or SEM/EDX) and organic marker measurements. A secondary objective is to provide the source apportionment calculations themselves, especially for studies of modest size that are mainly in-house efforts or not covered by separate tasks. This task is similar to the previous task 9572. The laboratory analyses to be performed under the task are crucial input to HEASD's Human Exposure Measurements Program, as well as to PM Implementation work of interest to OAQPS. The latter include QA audit and organic analysis support for the PM2.5 Speciation Trends Network.

PI: Charles Lewis

Modeling PM on Regional and Urban Scales. This research is designed to develop and evaluate (operationally and diagnostically) a PM modeling capability within EPA's Models-3 Community Multiscale Air Quality (CMAQ) modeling system. The goal is to produce an Eulerian framework state-of-the-science, emissions-based, regional-to-urban scale modeling system to address current and projected atmospheric loading of PM of varying size distributions, composition, and chemistry across varying spatial and temporal scales. This air quality model is one of the key tools needed by the States to study and develop control strategies for implementing the requirements of the NAAQS. The Models-3/CMAQ system can be applied to PM, ozone, and other air toxics. Because of this diversity, tasks are structured relative to the models framework as opposed to specific pollutant application.

You will need Adobe Acrobat Reader to view the Adobe PDF files on this page. See EPA's PDF page⁵ for more information about getting and using the free Acrobat Reader.

Recent Accomplishments

- [A Probabilistic Exposure Model Capable of Estimating Population Exposures to PM Components⁶](#) (PDF⁷, 3 pp., 24 KB)
- [Development of the Unmix Receptor Model for Calculating the Composition and Contributions of Particulate Matter Sources⁸](#) (PDF⁹, 4 pp., 20 KB)
- [Guidelines for the Application of SEM/EDX Analytical Techniques to Particulate Matter Samples¹⁰](#) (PDF¹¹, 3 pp., 60 KB)
- [New PM2.5 Profiles for Mobile Source Emissions¹²](#) (PDF¹³, 2 pp., 9 KB)
- [New Receptor Models for PM2.5 Source Apportionment¹⁴](#) (PDF¹⁵, 3 pp., 11 KB)

Related Web Sites

- [HEASD Particulate Matter¹⁶](#) - reports from recent studies
- [HEDS¹⁷](#) - Human Exposure Database System. HEDS is an integrated database system that contains chemical measurements, questionnaire responses, documents, and other information related to EPA research studies of the exposure of people to Environmental contaminants.
- [CHAD¹⁸](#) - Consolidated Human Activity Database (CHAD) contains data obtained from pre-existing human activity studies that were collected at city, state, and national levels. CHAD is intended to be an input file for exposure/intake dose modeling and/or statistical analysis.

URLs Provided for your Reference

1. http://www.epa.gov/nerl/goals/pm/air_pm_receptor.html#intro
2. http://www.epa.gov/nerl/goals/pm/air_pm_receptor.html#current
3. http://www.epa.gov/nerl/goals/pm/air_pm_receptor.html#ra
4. http://www.epa.gov/nerl/goals/pm/air_pm_receptor.html#web
5. <http://www.epa.gov/epahome/pdf.html>
6. <http://www.epa.gov/nerl/research/2004/g1-3.html>
7. <http://www.epa.gov/nerl/research/2004/g1-3.pdf>
8. <http://www.epa.gov/nerl/research/2004/g1-6.html>
9. <http://www.epa.gov/nerl/research/2004/g1-6.pdf>
10. <http://www.epa.gov/nerl/research/2002/g1-5.html>
11. <http://www.epa.gov/nerl/research/2002/pdf/g1-5.pdf>
12. <http://www.epa.gov/nerl/research/2000/html/g1-09.html>
13. http://www.epa.gov/nerl/research/2000/pdf_conversions/G1-09.pdf
14. <http://www.epa.gov/nerl/research/2000/html/g1-11.html>
15. http://www.epa.gov/nerl/research/2000/pdf_conversions/G1-11.pdf
16. <http://www.epa.gov/headweb/pm/pm.htm>
17. <http://www.epa.gov/heds/>
18. <http://www.epa.gov/chadnet1/>

Agriculture and Mining Exclusion Language

1. The indicator definition in proposed section 50.13(a)(2)(B)(ii) should read as follows:

"The standard for PM10-2.5 (1) applies to any ambient mix of PM10-2.5 that is dominated by resuspended dust from high-density traffic on paved roads and PM generated by industrial sources and urban construction sources, and (2) does not apply to PM10-2.5 that is (a) non-urban windblown dust and soils, (b) crustal materials, and (c) soils and fugitive PM generated by agricultural, mining, and other similar sources."

2. The following definition for agriculture sources should be fully explained in the Preamble, and included in the regulations themselves, as follows:

"Agricultural sources" as used in this regulation refer to all activities ordinarily and customarily occurring on farms and ranches, and livestock, poultry and equine pasturing and animal feeding operations, including tilling of soils, planting of seeds or transplanted seedlings, application of fertilizer, the action of livestock, poultry and equine hooves or feet on pen and soil surfaces, soil conditioning and reclamation, farm equipment and vehicle operation, crop harvesting and mechanical post-harvest plant residue management, and the collection and application of fertilizer and manure from feedlots and other livestock poultry and equine operations to croplands or its composting and handling for use as fertilizer, and similar activities that cause soil, seed, fertilizer and manure to become airborne. Such categories would also include, but not be limited to, Division A SIC code categories, which are separated as follows: Group 01: Agriculture Production Crops; Group 02: Agriculture Production Livestock and Animal Specialties; and Group 03: Agriculture Services.

3. The following definition for "mining sources" should be fully explained in the Preamble and included in the regulations themselves, as follows:

"Mining sources" as used in this regulation refers to all activities ordinarily and customarily conducted in and around mining operations that involve the removal of overburden, and the extraction and/or beneficiation of ore or materials from the earth, or its transportation, movement or storage at such operations."

Chemical Mass Balance Technique

Overview

A chemical mass balance (CMB) receptor modeling approach is used to quantify the source contributions to fallout particulate at a receptor. This method is based on direct measurement of the chemical composition of fallout particles present in the area of interest. The relative apportionment of these chemical species between potential sources is based on a statistical comparison of a chemical profile or “fingerprint” of each source with the chemical profile of an ambient fallout particle sample.

With this “fingerprinting” approach, impacts are based on retrospective measurements of samples selected from a specific period of potential maximum impact. Results represent the most probable quantitative source impacts for each specific sample selected.

The Chemical Mass Balance Method

The relationship between particulate emissions and ambient fallout concentrations measured at a receptor (pollutant sampler) site distant from an emitting source is a complicated one. Many variables, primarily meteorological, make the direct correlation between source emissions and ambient concentrations a poor one. Each of these variables is random in nature, will vary with space and time, and may combine with other variables in a nonlinear manner. Thus, any estimation of source contribution to fallout particles based on emissions and meteorology is approximate at best. However, the chemical mass balance (CMB) receptor-oriented model is a comparatively simple “model” based on physical principles which can be used to determine the average contribution of specific sources categories to particulate fallout. This model is based on the conservation of relative aerosol chemistry from the time a chemical species is emitted from its source to the time it is measured at a receptor. That is, if p sources are emitting M_j mass of particles, where m is the total mass of the particulate collected on a fallout tray at a receptor site, the model assumes the mass on the fallout tray is a linear combination of the mass contributed from each of the sources.

The mass of a specific chemical species, m_i , is given by the following:

where M_{ij} is the mass of element I from source j and FN_{ij} is the fraction of chemical species I in the mass from source j collected at the receptor. It is usually assumed that:

where F_{ij} is the fraction of chemical I emitted by source j as measured at the source. The degree of validity in this assumption depends on the chemical and physical properties of the species and its potential for atmospheric modifications such as condensation, volatilization, chemical reactions, sedimentation, etc.

If we accept this equation, however, and divide both sides of Equation 1 by the total mass of the deposit collected at the receptor site, it follows that:

$$(3)$$

or,

where C_i is the concentration of the chemical component i measured at the receptor and S_j is the source contribution, i.e., the ratio of the mass contributed from source j to the total mass collected at the receptor site. In practice, it is this fraction of particulate pollution measured at a receptor due to source j , S_j , which is of primary interest in receptor modeling calculations.

If the C_i and the F_{ij} at the receptor for all p of the source types suspected of affecting the receptor are known, and $p < n$ (n = number of chemical species), a set of n simultaneous equations exists from which the source type contributions S_j may be calculated by least squares methods.

Application of the CMB Modeling Method

In a typical chemical mass balance application, EPA's Version 7.0 CMB model (EPA, 1990) is applied to selected ambient samples. The CMB receptor modeling is performed in a manner consistent with EPA's *Protocol for Applying and Validating the CMB Model* (EPA, 1987).

The CMB procedure begins with a set of linear equations which expresses the ambient concentrations of chemical species measured at an ambient receptor site as the sum of products of source compositions and source contributions. This set of equations is over-determined (more than one possible solution) because the number of chemical species exceeds the number of contributing source types. The source contributions are the unknowns in these equations. However, a unique solution cannot be found for this set of equations because measurement uncertainty precludes determination of exact values for source and receptor data. When these uncertainties are estimated for both source and receptor measurements, additional physical constraints are applied which yield a most probable solution. This solution minimizes the difference between calculated and measured receptor concentrations by using an effective variance weighting scheme. The weighting has a physical significance in that it is derived from the measurement uncertainties of both source and receptor chemical species. (Species with higher relative concentration uncertainties carry less weight in the regression than species with lower relative uncertainties.) Although the CMB solution is identical to some statistical inference methods, it is not dependent on statistical principles. The basic model equations

which represent the source receptor relationship, the effective variance weighting, and the error propagation are all based on physical principles.

The CMB provides a source contribution estimate (SCE) and associated standard error uncertainty (STD ERR) for each source category. The model produces these estimates by making an effective variance weighted least squares fit between the chemical composition of the ambient sample and the composition of the sources. It estimates what amounts of each source (the SCEs) will collectively best explain the chemical composition of the ambient sample.

There are five basic data types necessary for CMB modeling:

- Source category names;
- Chemical composition or profile to be associated with each source category;
- Uncertainty in the chemical composition of each source category;
- Chemical composition of the fallout particles sampled at a receptor; and
- Uncertainty in the receptor chemical composition.

The ability of the CMB model to achieve a proposed set of apportionment goals is determined before the data is input into the computer. In other words, the chemical composition of the source profiles and ambient aerosol are established before the model is applied. At the time of data input, the only options available are the selection of source profiles and the source category names to associate with the profiles.

There are four major steps involved in applying the CMB receptor model to an existing database:

- Determine the appropriateness of the application;
- Form the input data files;
- Select the optimum model solution for each receptor sample; and
- Validate the model results.

The appropriateness of a data set for CMB modeling must be determined before the CMB model is applied. There are no quantitative rules that can be used. However, the EPA suggests using the following criteria as a guide (EPA, 1987):

- Although the model can be applied to a single sample, an adequate number of samples need to be available and included to represent the area or time period for which conclusions are to be drawn.
 - Species appropriate to the problem must be included in the database and with precision and accuracy's adequate to achieve source apportionment goals.
 - Source categories must not be collinear and their chemical compositions must represent the range of variability expected from a number of individual emitters in the same source type category.
-

- Source profiles must be representative of the emissions as they would arrive at the receptor.
- The number of source categories in a single application must be less than the number of species included in the regression.

Once it is determined that application of the CMB model is appropriate, it can be applied at varying levels of complexity. The EPA arbitrarily separates these into three levels. Level I uses existing data or data that can easily be obtained from analyses of existing samples. Level II involves additional analyses on existing samples or the acquisition of additional samples. Level III is a comprehensive CMB analysis and includes the acquisition of new data from both ambient and source sampling.

The process of CMB analysis consists of selecting the optimum solution to the effective variance least squares regression using the following seven steps:

- Assessment of the general applicability of the CMB model to the situation under study;
- Configuration of the model with appropriate sources, source profiles, and chemical species concentrations at receptor sites;
- Examination of model statistics and diagnostics;
- Determination of agreement with model assumptions;
- Identification of problems, changing the model configuration, and rerunning;
- Testing of the consistency and stability of model results; and
- Evaluation of the validity of model results.

Although there is a degree of subjectivity in this selection process, much of the subjectivity is removed if the fitting protocols and goodness-of-fit statistical criteria recommended by the EPA are used. The first step is to include all the sources or representatives of all source categories and all defined key species in the initial CMB analysis. Examination of the statistical goodness-of-fit criteria resulting from this initial analysis is used to evaluate the quality of the source contribution estimates. Based on this examination, a different set of sources and species is selected and evaluated. This stepwise procedure continues until, based on the following criteria, an optimum fit is obtained:

- Percent mass explained is close to 100%;
- R-square is close to 1;
- Chi-square is minimized;
- T-statistic is greater than 2;
- Source uncertainty clusters are minimized;
- Calculated-to-measured species ratios are close to 1;
- Ratios of R/U are close to 0; and
- Degrees of freedom are maximized. These criteria are defined and described in Table 1.

Output/Statistic	Abbreviation	EPA Target	Explanation
Std. Error	STD ERR	<< SCE	The standard error of the SCE.
T-statistic	T-STAT	> 2.0	The ratio of the value of the SCE to the uncertainty in the SCE. A T-STAT greater than 2 means that the SCE has a relative uncertainty of less than 50%. T-STAT = SCE/STD ERR
R-square	R-SQUARE	0.80 to 1.00	A measure of the variance of the ambient concentration explained by the calculated concentration. The target range is 0.8 to 1.0, where an r-square of 1.0 is perfect.
Chi-square	CHI-SQUARE	0.0 to 4.0	A term that compares the difference between the calculated and measured ambient concentrations to the uncertainty of the difference. A perfect fit has a chi-square of 0, and a chi-square less than 2 usually indicates a good fit. The target range is 0.0 to 4.0.
Percent Mass Explained	% MASS	100% " 20%	The ratio of the total calculated to measured mass. The target range is 80% to 120%. % MASS = $M_c/M_m \times 100$
Degrees of Freedom	DF	> 5	The difference between the number of fitting species and the number of fitting sources. This value must exceed 1 and should be greater than 5.
Uncertainty/Similarity Clusters	U/S CLUSTERS	None	A list of sources that were not sufficiently resolved by the CMB analysis. No clustering is preferred.
Ratio of Calculated to Measured	RATIO C/M	0.5 to 2.0	The ratio of the calculated to measured concentration of an ambient species. Ideally, this value should be 1.0, but the target range is 0.5 to 2.0. RATIO C/M = C_i/M_i for each species I.
Ratio of Residual to Uncertainty	RATIO R/U	B2.0 to 2.0	The ratio of the residual (calculated minus measured) to the uncertainty of the residual (square root of the sum of squares of the uncertainties). Target range is -2.0 to 2.0.

The model provides three primary outputs: the contribution estimates to ambient concentrations of the sources or source Categories which are included in the fit (SCE), the standard errors of these source contribution estimates (STD ERR), and the species concentrations calculated from the fit (CALC).

The model provides three statistical measures which can be used to evaluate how well the model's calculated species concentrations match the ambient measurements for these species. These statistics are the percent of total mass explained by the fit (% MASS), R-SQUARE, and CHI-SQUARE. It is generally desirable to obtain a good fit of the data based on these three measures while obtaining SCEs with low STD ERR relative to the size of the SCE.

The model provides four diagnostics to help identify data responsible for a poor fit so that improved data might be obtained or included to rectify the situation. These are the uncertainty/similarity clusters (U/S CLUSTERS), the ratio of calculated to measured species concentrations (RATIO C/M), the ratio of the residual (calculated minus measured) to the uncertainty of this difference (RATIO R/U), and the portion of a calculated species concentration that is attributed by the model to each source (SSCONT). The latter diagnostic is not included on the standard CMB printout.

There are four main error categories that can impact model performance: incorrect ambient data, incorrect source profiles, incorrect source list, and profile uncertainty/incorrect collinearity. The existence of these errors can be inferred from the diagnostics and indicators listed above. Possible corrective actions include evaluating ambient and source data, reanalyzing samples, including different sources in the source list, deleting sources from the source list, compositing collinear source profiles, analyzing samples for additional species, etc. After corrective action has been taken, the fit of the measured species data is reevaluated.

When statistically sound and physically reasonable fits have been obtained for the ambient samples of interest, the stability of the CMB model results are assessed. This includes the evaluation of the sensitivity of the model's results to errors in the sources, source profiles, and the ambient data. The final step in the application of the CMB model is validation. In this step, the model results are evaluated for their consistency with available related data (e.g. meteorological, spatial, emissions, and particle size data). Comparisons are made with the results of other receptor and/or dispersion models, if available.

When the summary statistics and diagnostics are generally within target ranges, when there are no significant deviations from model assumptions, when the sensitivity tests uncover no unacceptable instability or consistency problems, and when the results are consistent with available related data, the CMB analysis is considered complete and valid.

Using the fitting parameters in Table 1 and the EPA guidelines, this modeling procedure will generally result in optimized source contributions. The resulting fit is only one of many possible solutions, but it should be the most probable solution. The existence of several different solutions with similar fitting parameters suggests similar probabilities of correctness for each set of source contributions. In such a case, the SCEs of the major sources will likely be quite similar.
