UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON D.C. 20460

November *xx*, 2005

EPA-SAB-CASAC-06-001

OFFICE OF THE ADMINISTRATOR
SCIENCE ADVISORY BOARD

Honorable Stephen L. Johnson
Administrator
U.S. Environmental Protection Agency
1200 Pennsylvania Avenue, NW
Washington, DC 20460

Subject: Clean Air Scientific Advisory Committee (CASAC) Peer Review of the Federal Reference Method (FRM) for Coarse Particulate Matter ($PM_{10-2.5}$)

Dear Administrator Johnson:

The Clean Air Scientific Advisory Committee's (CASAC or "Committee") Ambient Air Monitoring and Methods (AAMM) Subcommittee ("Subcommittee") met in a public meeting on September 21-22, 2005 in Durham, NC to conduct a peer review of EPA's proposed Federal Reference Method (FRM) for coarse particulate matter ($PM_{10-2.5}$) and a consultation on various particulate matter (PM) monitoring-related issues.

The CASAC hereby forwards this letter to you as the Committee's consensus report on this subject. The current roster of the seven-member Clean Air Scientific Advisory Committee — three members of which are also members of the Subcommittee — is attached as Appendix A to this report, and the CASAC AAMM Subcommittee roster is contained in Appendix B. EPA's charge to the Subcommittee is found in Appendix C of this report, and Subcommittee members' individual review comments are provided in Appendix D.

The Committee was extremely impressed by the continuing high quality of technical work evident in the field methods evaluation studies. In general, the CASAC agrees that there are several important scientific or operational strengths of the proposed difference method $PM_{10-2.5}$ FRM, while noting that there are several prominent weaknesses as well. Despite these weaknesses, no other better, currently available candidate FRM method has been identified at the present time. If the Agency adopts the proposed $PM_{10-2.5}$ difference method as the exclusive FRM, it should be done with the clear understanding that it is not intended for extensive field deployment, and would be employed primarily as a benchmark for evaluating performance of other continuous or dichotomous FEM instruments. The Agency should also consider the possibility of specifying more than one FRM for $PM_{10-2.5}$ (as it has for PM_{10}). There is clearly a need for the Agency to continue to work diligently toward developing more direct coarse-

particle-only sampling methods and an associated need to devote more resources to support the needed research and development in this important area.

1. Background

The CASAC, which comprises seven members appointed by the EPA Administrator, was established under section 109(d)(2) of the Clean Air Act (CAA or Act) (42 U.S.C. 7409) as an independent scientific advisory committee, in part to provide advice, information and recommendations on the scientific and technical aspects of issues related to air quality criteria and national ambient air quality standards (NAAQS) under sections 108 and 109 of the Act. The CASAC, which is administratively located under EPA's Science Advisory Board (SAB) Staff Office, is a Federal advisory committee chartered under the Federal Advisory Committee Act (FACA), as amended, 5 U.S.C., App. The SAB Staff Office established the CASAC AAMM Subcommittee in early 2004 as a standing subcommittee to provide the EPA Administrator, through the CASAC, with advice and recommendations, as necessary, on topical areas related to ambient air monitoring, methods and networks. The CASAC and the Subcommittee comply with the provisions of FACA and all appropriate SAB Staff Office procedural policies.

Under section 108 of the CAA, the Agency is required to establish NAAQS for each pollutant for which EPA has issued criteria, including particulate matter (PM). Section 109(d)(1) of the CAA requires that EPA carry out a periodic review and revision, where appropriate, of the air quality criteria, to reflect advances in scientific knowledge on the effects of the pollutant on public health and welfare, and the NAAQS for "criteria" air pollutants such as PM. EPA is currently reviewing the NAAQS for PM. As part of this review, the Agency is considering potential NAAQS for coarse particulate matter ($PM_{10-2.5}$).

In conjunction with the review of the NAAQS for PM, EPA is evaluating potential monitoring methods for measurement of $PM_{10-2.5}$. The Agency's Office of Air Quality Planning and Standards (OAQPS), within EPA's Office of Air and Radiation (OAR), requested that the CASAC conduct a peer review of the proposed Federal Reference Method (FRM) for $PM_{10-2.5}$, to provide independent scientific advice on the appropriateness of this method as a basis of comparison in approving Federal Equivalent Method (FEM) coarse-particle monitors, which provide better temporal (continuous) information, or provide coarse-only filter samples (dichotomous) more amenable to chemical analyses. The FRM for $PM_{10-2.5}$ will establish the basis for approval of FEM monitoring methods in a performance-based measurement system process.

In addition, OAQPS asked the CASAC to conduct a consultation with the Agency on: fine particle ($PM_{2.5}$) FRM optimization and equivalency criteria for continuous monitors; and $PM_{10-2.5}$ methods evaluation, network data quality objectives (DQOs), and equivalency criteria for continuous monitors. (This consultation portion of the Subcommittee's September 21-22 meeting is not covered by this report, although it is addressed by individual reviewers in Appendix D.) The CASAC AAMM Subcommittee previously provided advice and recommendations for this ongoing work at a July 22, 2004 consultative meeting on $PM_{10-2.5}$ methods and DQOs. Prior to this meeting, OAQPS posted all relevant written review materials on the "CASAC File Area" page of the Agency's Ambient Monitoring Technology Information Center (AMTIC) Web site at URL: http://www.epa.gov/ttn/amtic/casacinf.html.

2. CASAC Peer Review of the FRM for $PM_{10-2.5}$

With respect to the first charge question (Appendix C), CASAC AAMM Subcommittee members generally agree that there are several important scientific or operational strengths of the proposed difference method $PM_{10-2.5}$ FRM. These include:

- Direct gravimetric measurement of mass by proven and available technology;
- Use of existing FRM equipment minimizes equipment and training costs;
- Measurements can be highly precise, even when mass concentrations are low;
- Low face velocities may reduce evaporative losses of some semi-volatile species;
- The particle size separation properties (sampling effectiveness curves) of the samplers are better characterized and documented than other candidate methods;
- The use of reference method filters, filter handling procedures and inlets for $PM_{2.5}$ and PM_{10} makes the method "accurate" by definition (although this doesn't necessarily provide an accurate depiction of coarse particles in the ambient air).
- No need for air-conditioned shelters;
- Consistency with historical database of mass measurements avoids the need for expensive field comparisons (although historical PM_{10} measurements are primarily by high-volume methods);
- Presence of $PM_{2.5}$ µm particles causes coarse particles to adhere to filter (avoiding mass losses that may affect dichotomous coarse-only filter samples);
- Filter-based samples may allow for chemical speciation (although the quality of speciation measurements by difference methods is questionable and requires further evaluation).

The Subcommittee also noted several weaknesses of the proposed method, including:

- Accuracy of the proposed filter difference method is unknown and difficult to establish under all relevant conditions.
- Marginal suitability for speciation analysis (by subtraction), especially for species not predominantly in the coarse mode (dichotomous or impactor samples would be more suitable, if sufficient sample material can be collected).
- Possible sampling artifacts from losses of volatile material (*i.e.*, nitrate, organic compounds) during sampling.
- Likelihood of different sampling artifacts for the $PM_{2.5}$ and PM_{10} filters, because reactions between fine and coarse PM species may reduce the volatility of nitrates and other compounds, and/or because evaporative losses on $PM_{2.5}$ may exceed evaporative losses on PM_{10} due to the pressure drop provided by the WINS or cyclone. Positive artifacts — such as from reactions between acidic gases and coarse alkaline crustal material might also occur more frequently on PM_{10} filters.
- Positive or negative sampling artifacts lead to inaccuracies that cannot be quantified.
- Poor time resolution: only suitable for determining compliance with 24-hour standards; completely unsuitable for use in Air Quality Index (AQI) reports/forecasting or investigating associations between short-term (e.g., hourly average or maximum) concentrations and health endpoints.

- Expensive, labor-intensive, manual sample collection and laboratory analysis, requiring great care in all aspects of method operation.

If the Agency adopts the proposed $PM_{10-2.5}$ difference method as the exclusive FRM, it should be done with the clear understanding that it is not intended for extensive field deployment, and would be employed primarily as a benchmark for evaluating performance of other continuous or dichotomous FEM instruments. Continued development and evaluation of methods based on virtual impaction to collect samples of coarse-particles-only should be given a high priority. The Agency should also consider the possibility of specifying more than one FRM for $PM_{10-2.5}$ (as it has for PM_{10}), if one or more of the current or evolving dichotomous sampler designs shows reasonable agreement with the difference method (assuming filter-handling procedures can be developed to minimize losses of coarse-only particles prior to weighing). Work should continue on development of *accurate* techniques for measurements of coarse particle mass concentrations (and on methods to directly quantify the accuracy). In addition, Subcommittee members expressed their desire to see automated, time-resolved (hourly), real-time samplers for lower operating costs, forecasting and AQI support, health studies, *etc.*

Since the proposed difference method FRM would be used as a basis for approval of other methods, some of the Subcommittee members were concerned that this may close the door for the new methods that more accurately measure the ambient PM mass (including semi-volatile species) than the proposed difference method. Precision, historical continuity, and compatibility between different PM definitions are all desirable data quality objectives, but these objectives should not take precedence over the "science quality objective" of providing an accurate characterization of coarse particle concentration and composition in the ambient air. EPA should explore options to certify alternative Federal Equivalent Methods (or alternative FRM methods) that can demonstrate superior accuracy to the difference method FRM.

Many Subcommittee members felt that more thought should be given to comparing the responses of alternative samplers with the FRM using laboratory generated-aerosols of known composition and size or size distribution. Such work could include calibrated generation and sampling of known semi-volatile compounds, such as ammonium nitrate and selected organic compounds. While this methodology might not be applicable to equivalency determinations as specified by law, it could take us a long way towards an understanding of measurement accuracy. The laboratory tests would enable unambiguous testing of sampler performance to particles having known physical and chemical properties. This approach would help to improve our understanding of measurement accuracy, and would lead to the design of improved samplers in the future.

Some members questioned EPA's claim that a difference method FRM would provide a sound basis for chemical analysis (i.e. coarse chemical composition by subtraction). This is an important issue in that the "urban" focus of EPA's proposed "$UPM_{10-2.5}$" indicator is based on assumed (but not routinely measured) differences in the chemical (and/or biological) composition of coarse particles in urban vs. rural locations. Clearly, separate PM_{10} and $PM_{2.5}$ samples can be collected and analyzed, but the much higher uncertainty associated with the chemical analyses of these speciation samples, and the probability of different, chemical or size-specific sampling and analytical artifacts on PM_{10} and $PM_{2.5}$ filters are factors that make the

quality of such "speciation by difference" data highly uncertain. Perhaps it will yield acceptable data, perhaps not; further studies to examine the practicality of the difference method for speciation are needed to demonstrate its utility. EPA should address some of the questions about speciation by analyzing the already-collected speciation data from the field studies. The need for speciation data, however, is inescapable, and the virtual impactor has some significant advantages for providing samples for speciation analysis, although its incomplete separation of $PM_{2.5}$ and $PM_{10-2.5}$ is problematic, and the relatively small total coarse mass sample size may not allow for detailed chemical analyses (for example, for carbonaceous species or for biological content). Conversely, the state of the virtual impactor design has advanced significantly during the last 30 years (since the original Loo and Cork design — used with no or minimal modifications in all candidate dichotomous instruments tested to date), and it is likely that a much-improved virtual impactor design could be developed if support for such research were made available.

Regarding the second charge question (Appendix C), the Subcommittee was extremely impressed by the continuing high quality of technical work evident in the field methods evaluation studies, in particular, by EPA's National Exposure Research Laboratory (NERL), within the Office of Research and Development (ORD); equipment vendors; and the Jefferson County [AL] Department of Health. It is essential that such work be continued and that the resources necessary to sustain it are maintained or increased. A majority of the Subcommittee members expressed the opinion that the demonstrated data quality of the $PM_{10-2.5}$ difference method in the EPA field studies performed to date supports it being proposed as a PM coarse FRM. However, due to many weaknesses of the difference method, EPA should emphasize deployment of dichotomous or continuous FEMs in the network and use the difference method FRMs primarily to evaluate the performance of these alternative methods. Such performance evaluations need to be conducted over a range of locations and seasons, but there should not be requirements for a high proportion of collocated difference method FRM samplers in state local and tribal monitoring networks. It is also doubtful that a difference method FRM would be useful as a routine field audit device. Application of rigorous criteria for designation of equivalence for $PM_{10-2.5}$ methods, such as were presented in Attachment 4 (for $PM_{2.5}$) of the review materials, may be quite cumbersome and problematic. By definition, it would seem that only the FRM difference method could be expected to produce fully equivalent results. Perhaps a "functional equivalency", based on practical considerations and the above-noted accuracy issues, would be more useful.

It has been also proposed that since the number of $PM_{10-2.5}$ [urban] non-attainment areas is expected to be much smaller than for $PM_{2.5}$ — and many of these areas have existing PM_{10} compliance and Toxics monitoring programs — EPA could limit deployment of the cumbersome difference method for $PM_{10-2.5}$ if it allowed PM_{10} monitors to be used to demonstrate attainment. U.S. EPA and State, local and Tribal (SLT) agencies have already invested large resources into the current regulatory PM_{10}, Toxics PM_{10} and $PM_{2.5}$ monitoring networks. Several states (*e.g.*, California) have State ambient air quality standards for PM_{10} and do not plan to follow U.S. EPA in adopting a coarse particle standard. It stands to reason that, if a site meets the $PM_{10-2.5}$ standard with PM_{10} monitoring data (uncorrected for the inclusion of $PM_{2.5}$), then there is no need to deploy a $PM_{10-2.5}$-specific monitor at the site for compliance determination. Sampling for PM-coarse in urban areas should be primarily at locations that reflect the urban scale PM-

coarse signal for use in health effect studies; a limited number of micro/mid or even some neighborhood scale sites is of limited use in health studies.

Although the quality of data obtained by the Jefferson County Department of Health is very impressive, some members questioned if the same data quality could be obtained under the resource-constrained routine compliance monitoring network conditions, typically found at many State and local monitoring agencies. It can be noted for example that the precision of the $PM_{2.5}$ measurements in the EPA and Jefferson County field studies is substantially tighter than that which has been observed for $PM_{2.5}$ nationwide. These field studies indicate that precise $PM_{10-2.5}$ data can be obtained by careful, expert personnel, but not necessarily that such precise data will be obtained in more routine field operations. It was also noted that problems with the Tapered Element Oscillating Microbalance (TEOM™), dichotomous sampler, Aerodynamic Particle Sizer (APS™), and other units were only discovered during the ORD inter-comparison study because multiple units were carefully collocated and operated by U.S. EPA and monitoring industry experts. If the units were operating by themselves in an SLT agency monitoring network, it is unlikely that the multiple instrument problems observed in the EPA study would have been detected in a timely manner. Without the ability to challenge a PM analyzer with a known concentration of PM, all we have to verify proper operation of an analyzer is the "due diligence" of the site technician and highly skilled data review, including level 2 data validation, both which occur months after data collection – maximizing the potential for data loss. Once again, this reflects on a need for resources — preferably by an EPA "Science to Achieve Results" (STAR) grant — to develop traceable standards for PM. This would obviously be a very challenging undertaking, but without such standards, PM in any size range will always be a pollutant defined by how it is sampled, rather than by what is in the ambient air and what may potentially deposit onto the human respiratory tract. Research priorities should also be placed on development of an improved virtual impactor (for both PM fine and coarse), and for more clearly identifying (and eliminating) PM coarse filter handling or shipping losses.

In summary, although the proposed difference method has many flaws, no other better, currently available candidate FRM method has been identified at the present time. There is clearly a need to work diligently toward identifying more direct coarse-particle-only sampling methods and an associated need to devote more resources to support the needed research and development in this important area. The Clean Air Scientific Advisory Committee and the CASAC AAMM Subcommittee were pleased to advise the Agency in this extremely important task. As always, we wish EPA well as the Agency continues this process.

Sincerely,

Dr. Rogene Henderson, Chair
Clean Air Scientific Advisory Committee

Appendix A – Roster of the Clean Air Scientific Advisory Committee

Appendix B – Roster of the CASAC AAMM Subcommittee

Appendix C – Charge to the CASAC AAMM Subcommittee

Appendix D – Review Comments from Individual CASAC AAMM Subcommittee Members

Appendix A – Roster of the Clean Air Scientific Advisory Committee

---

U.S. Environmental Protection Agency
Science Advisory Board (SAB) Staff Office
Clean Air Scientific Advisory Committee (CASAC)

CHAIR

Dr. Rogene Henderson, Scientist Emeritus, Lovelace Respiratory Research Institute, Albuquerque, NM

MEMBERS

Dr. Ellis Cowling, University Distinguished Professor-at-Large, North Carolina State University, Colleges of Natural Resources and Agriculture and Life Sciences, North Carolina State University, Raleigh, NC

Dr. James D. Crapo, Professor, Department of Medicine, Biomedical Research and Patient Care, National Jewish Medical and Research Center, Denver, CO

Dr. Frederick J. Miller, Consultant, Cary, NC

Mr. Richard L. Poirot, Environmental Analyst, Air Pollution Control Division, Department of Environmental Conservation, Vermont Agency of Natural Resources, Waterbury, VT

Dr. Frank Speizer, Edward Kass Professor of Medicine, Channing Laboratory, Harvard Medical School, Boston, MA

Dr. Barbara Zielinska, Research Professor, Division of Atmospheric Science, Desert Research Institute, Reno, NV

SCIENCE ADVISORY BOARD STAFF

Mr. Fred Butterfield, CASAC Designated Federal Officer, 1200 Pennsylvania Avenue, N.W., Washington, DC, 20460, Phone: 202-343-9994, Fax: 202-233-0643 (butterfield.fred@epa.gov) (Physical/Courier/FedEx Address: Fred A. Butterfield, III, EPA Science Advisory Board Staff Office (Mail Code 1400F), Woodies Building, 1025 F Street, N.W., Room 3604, Washington, DC 20004, Telephone: 202-343-9994)

Appendix B – Roster of the CASAC AAMM Subcommittee

U.S. Environmental Protection Agency
Science Advisory Board (SAB) Staff Office
Clean Air Scientific Advisory Committee (CASAC)
CASAC Ambient Air Monitoring & Methods (AAMM) Subcommittee

CHAIRS

Mr. Richard L. Poirot* (Chair – Monitoring), Environmental Analyst, Air Pollution Control Division, Department of Environmental Conservation, Vermont Agency of Natural Resources, Waterbury, VT

Dr. Barbara Zielinska* (Chair – Methods), Research Professor, Division of Atmospheric Science, Desert Research Institute, Reno, NV

SUBCOMMITTEE MEMBERS

Mr. George Allen, Senior Scientist, Northeast States for Coordinated Air Use Management (NESCAUM), Boston, MA

Dr. Judith Chow, Research Professor, Desert Research Institute, Air Resources Laboratory, University of Nevada, Reno, NV

Dr. Ellis Cowling*, University Distinguished Professor-at-Large, North Carolina State University, Colleges of Natural Resources and Agriculture and Life Sciences, North Carolina State University, Raleigh, NC

Mr. Bart Croes, Chief, Research Division, California Air Resources Board, Sacramento, CA

Dr. Kenneth Demerjian, Professor and Director, Atmospheric Sciences Research Center, State University of New York, Albany, NY

Dr. Delbert Eatough, Professor of Chemistry, Chemistry and Biochemistry Department, Brigham Young University, Provo, UT

Mr. Eric Edgerton, President, Atmospheric Research & Analysis, Inc., Cary, NC

Mr. Henry (Dirk) Felton, Research Scientist, Division of Air Resources, Bureau of Air Quality Surveillance, New York State Department of Environmental Conservation, Albany, NY

Dr. Philip Hopke, Bayard D. Clarkson Distinguished Professor, Department of Chemical Engineering, Clarkson University, Potsdam, NY

Dr. Rudolf Husar, Professor, Mechanical Engineering, Engineering and Applied Science, Washington University, St. Louis, MO

Dr. Kazuhiko Ito, Assistant Professor, Environmental Medicine, School of Medicine, New York University, Tuxedo, NY

Dr. Donna Kenski, Data Analyst, Lake Michigan Air Directors Consortium, Des Plaines, IL

Dr. Thomas Lumley, Associate Professor, Biostatistics, School of Public Health and Community Medicine, University of Washington, Seattle, WA

Dr. Peter McMurry, Professor and Head, Department of Mechanical Engineering, Institute of Technology, University of Minnesota, Minneapolis, MN

Dr. Kimberly Prather, Professor, Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA

Dr. Armistead (Ted) Russell, Georgia Power Distinguished Professor of Environmental Engineering, Environmental Engineering Group, School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA

Dr. Jay Turner, Visiting Professor, Crocker Nuclear Laboratory, University of California - Davis, Davis, CA

Dr. Warren H. White, Research Professor, Crocker Nuclear Laboratory, University of California - Davis, Davis, CA

Dr. Yousheng Zeng, Air Quality Services Director, Providence Engineering & Environmental Group LLC, Baton Rouge, LA

SCIENCE ADVISORY BOARD STAFF

Mr. Fred Butterfield, CASAC Designated Federal Officer, 1200 Pennsylvania Avenue, N.W., Washington, DC, 20460, Phone: (202) 343-9994, Fax: (202) 233-0643 (butterfield.fred@epa.gov) [Physical/Courier/FedEx Address: Fred A. Butterfield, III, EPA Science Advisory Board Staff Office (Mail Code 1400F), Woodies Building, 1025 F Street, N.W., Room 3604, Washington, DC 20004, Telephone: (202) 343-9994]

* Members of the statutory Clean Air Scientific Advisory Committee (CASAC) appointed by the EPA Administrator

Appendix C – Charge to the CASAC AAMM Subcommittee

Peer Review Questions:

Questions associated with Attachment 1 – Selection and technical summary of $PM_{10-2.5}$ FRM:

1. What are the scientific and operational strengths and weaknesses of the $PM_{10-2.5}$ difference method relative to other options for a proposed FRM, especially when used as the basis for approval of other methods?

2. Based on the field study report as well as any other available data, *e.g.*, data from State and local agencies, how does the demonstrated data quality of the $PM_{10-2.5}$ difference method support or detract from it being proposed as a FRM?

Consultation Questions:

Question associated with Attachment 2 – EPA's Multi-Site Evaluations of Candidate Methodologies for Determining Coarse Particulate Matter ($PM_{10-2.5}$) Concentrations: August 2005 Updated Report Regarding Second-Generation and New $PM_{10-2.5}$ Samplers:

1. Based upon the latest available field study data, which $PM_{10-2.5}$ methods have both sufficient utility to meet one or more important monitoring objectives and appropriate data quality to be considered for deployment as Federal Equivalent Methods (FEMs) or speciation samplers in a potential $PM_{10-2.5}$ monitoring network?

Questions associated with Attachment 3 – Memo to PM NAAQS Review Docket (OAR-2001-0017) – Potential changes being evaluated for the $PM_{2.5}$ Federal Reference Method

2. What are the Subcommittee's views on the Very Sharp Cut Cyclone (VSCC) being approved as an alternative second-stage impactor to the Well Impactor Ninety-Six (WINS) for use on a $PM_{2.5}$ FRM?

3. To what extent are the stated advantages of relaxing existing requirements identified for the $PM_{2.5}$ FRM supported by the information cited in Attachment 3, available literature, or good field and laboratory practices? Does the Subcommittee have additional recommendations for the $PM_{2.5}$ FRM that would be neutral with respect to bias, but would improve the performance and minimize the burden on agencies conducting the sampling?

Questions associated with Attachment 4 – Criteria for Designation of Equivalence Methods for Continuous Surveillance of $PM_{2.5}$ Ambient Air Quality

4. Considering the statistical measures of precision, correlation, multiplicative bias, and additive bias identified for approval of $PM_{2.5}$ continuous methods, what are the

Subcommittee's views on the usefulness of each measure to ensure that approved or equivalent methods meet the monitoring network data quality objectives?

5. What are the advantages and disadvantages of using sampler precision and sample population to help determine the minimum correlation requirement for the approval of $PM_{2.5}$ continuous methods?

6. What are the Subcommittee's views on using a $PM_{2.5}$ continuous monitor approved as a FEM, being applicable for use as part of a potential $PM_{2.5}$ secondary standard for visibility?

Question associated with Attachment 5 – Sensitivity of the $PM_{10-2.5}$ Data Quality Objectives to Spatially Related Uncertainties

7. To what extent have the assessments of spatial variability and the sensitivity of the DQO process to a variety of population distributions been appropriately addressed?

Question associated with Attachment 6 – $PM_{10-2.5}$ Method Equivalency Development

8. What are the Subcommittee's views on the approach identified for the development of criteria to approve continuous $PM_{10-2.5}$ equivalent methods?

Appendix D – Review Comments from Individual CASAC AAMM Subcommittee Members

This appendix contains the preliminary and/or final written review comments of the individual members of the Clean Air Scientific Advisory Committee (CASAC) Ambient Air Monitoring & Methods (AAMM) Subcommittee who submitted such comments electronically. The comments are included here to provide both a full perspective and a range of individual views expressed by Subcommittee members during the review process. These comments do not represent the views of the CASAC AAMM Subcommittee, the CASAC, the EPA Science Advisory Board, or the EPA itself. The views of the CASAC AAMM Subcommittee and the CASAC as a whole are contained in the text of the report to which this appendix is attached. Subcommittee members providing review comments are listed on the next page, and their individual comments follow.

Panelist **Page #**

Mr. George Allen ……D-3

Dr. Judith Chow ……D-10

Dr. Ellis Cowling ……D-20

Mr. Bart Croes ……D-25

Dr. Kenneth Demerjian ……D-34

Dr. Delbert Eatough ……D-38

Mr. Dirk Felton ……D-51

Dr. Philip Hopke ……D-57

Dr. Rudolf Husar ……D-60

Dr. Kazuhiko Ito ……D-62

Dr. Donna Kenski ……D-68

Dr. Thomas Lumley ……D-70

Dr. Peter McMurry ……D-74

Dr. Kimberly Prather ……D-78

Dr. Armistead (Ted) Russell ……D-83

Dr. Jay Turner ……D-85

Dr. Warren H. White ……D-89

Dr. Yousheng Zeng ……D-96

Mr. George Allen

To: Fred Butterfield, Designated Federal Officer
EPA SAB, Clean Air Scientific Advisory Committee (CASAC)
Ambient Air Monitoring and Methods Subcommittee

From: George Allen, AAMM subcommittee member, September 30, 2005

The following are revised written comments based on discussions during the September 21-22, 2005 meeting on a peer review of the PM-coarse FRM. A copy of these comments is being sent to Dr. Barbara Zielinska and Mr. Rich Poirot, CASAC AAMM Subcommittee Co-Chairs. These comments address the Charge Questions in the EPA OAQPS memo to the SAB dated August 19, 2005.

Questions associated with Attachment 1 – Selection and technical summary of PM10-2.5 FRM:
1. What are the scientific and operational strengths and weaknesses of the PM10-2.5 difference method relative to other options for a proposed FRM, especially when used as the basis for approval of other methods?
2. Based on the field study report as well as any other available data, e.g., data from State and local agencies, how does the demonstrated data quality of the PM10-2.5 difference method support or detract from it being proposed as an FRM?

Attachment 1 is a clear, concise, and thorough summary of the issues involved with use of the difference method for PM-coarse (PM10-PM2.5 or $PM_{10-2.5}$ as used in this document). This summary is well written and technically complete. My comments on these two questions based on this document follow.

<u>Charge Question #1.</u>

I agree with EPA that the difference method for PM-coarse is clearly the most defendable approach for a reference measurement method. When exactly matched pairs of samplers and protocols are run by highly skilled staff in a well-controlled research environment, the difference method is the most definitive PM-coarse measurement with the fewest data quality ambiguities. As noted in the writeup, identical hardware for both systems must be used (with the exception of the lack of a PM2.5 impactor or cyclone in the "PM10c" sampler. Specifically, this excludes use of a PM10 Hi-Vol sampler for this purpose, even when it is designated as an FRM.

The EPA summary recommends that any sampler pair that has been designated as a PM2.5 FRM can be used as a PM-coarse FRM. This reviewer suggests that for PM-coarse, the acceptable average field blank value of 30 µg for the PM2.5 FRM is too high and could lead to degraded precision of PM-coarse data where PM2.5 is substantially greater than PM-coarse (much of the eastern U.S.). There is a wide range in both mean and variation of field blanks across different PM2.5 FRMs (here, a field blank is defined as a filter loaded into the sampler and left in place for the same duration as a normal sample -- typically 48 hours or more). Some methods routinely achieve mean field blanks in the range of 5 to 10 µg; it is recommended that this specification be

reduced to a maximum of 10 µg field blank mean for sampler pairs used for PM-coarse measurements.

I do not agree with the proposal to also accept (as class I equivalent methods) pairs of dissimilar models of FRMs that have been designated as PM2.5 FRMs without further testing. There are potential subtle differences between models that may degrade the PM-coarse measurement. This is especially true of sequential samplers. Issues here include the potential for different field blank values among different sampler models, and in the case of a sequential and manual FRM pair, different post-sample times (and thus different semi-volatile PM losses) before field collection. Another potential variable is "effective filter face velocity", determined not only by flow and "apparent" exposed filter surface area, but also the "effective" exposed area, determined by the actual open area (holes) in the filter support screen. This latter parameter can vary between different models of PM2.5 FRMs unless the same support screens are used; the resulting difference in "effective filter face velocity" could affect loss of semivolatile PM species. In summary, all of these parameters (physical and temporal) must be kept identical between the two samplers in a PM-coarse pair to avoid potential biases.

This summary suggests on page 5 that the difference method for PM-coarse can be rapidly deployed into existing compliance monitoring networks with minimal training or pilot operation periods. Although the difference method for PM-coarse has performed very well in the EPA tests (a highly controlled research-grade environment), that performance is likely to degrade substantially in the real-world of resource constrained SLT monitoring programs. This reviewer does not recommend wide deployment of this PM-coarse method without substantial further field tests in the SLT environment across a range of agencies.

Related to this issue of SLT deployment of the difference method, as I understand it the officially designated FRM method for PM-coarse must be used by SLTs for audit purposes. This poses a potential problem; many agencies may have difficulty generating sufficiently precise pm-coarse data with the difference method to meaningfully evaluate the performance of the routine (nondifference) method, rendering the audit process useless at best and misleading at worst.

Page 5 also suggests that the difference method can provide speciated analyses of coarse mode aerosols. This is only true in a practical sense for species that are present primarily only in the coarse mode (typically crustal elements); for PM species that are dominant in the fine mode, the data from this approach will usually have severely degraded coarse mode precision. The dichotomous sampler method is much better suited for this kind of analysis. Ideally a medium flow dichot design would be used for coarse mass speciation; sample flows would be at least 6 times those in the "classic" dichot sampler, or at least 100 LPM inlet flow. An example of such an approach would be the custom dichot system used in the St. Louis supersite, with an inlet flow of 113 LPM and a coarse channel flow of 11.3 LPM. These flows are almost seven times greater than the standard dichot sampler, with corresponding increases in sample material on the coarse channel filter.

This summary correctly reflects the concerns related to use of a virtual impactor-based (dichotomous) sampler. This reviewer agrees with the EPA summary that the dichotomous sampler is less suitable as a PM-coarse (mass) reference method. The loss of large particles from

the coarse mode filter during shipment and the potential for excessive amounts of coarse particles being carried over to the fine mode filter are very real concerns.

With regard to the first dichot issue above, a meaningful dichot coarse filter mass shipping loss test has still not yet been performed; given that the existing literature shows losses (for PM-15) in the range of 30 to 50% (Spengler and Thurston, *JAPCA* December 1983, **33**:12; and Dzubay and Barbour, *JAPCA* August 1983, **33**:7), this is a critical test.

The second issue is one of virtual impactor design. All the systems tested (except one version of the Kimoto dichot) use essentially the same virtual impactor design - the Loo and Cork design from the mid 1970's, described in Loo and Cork, *Aerosol Science and Technology* **9**:167-176 (1988). The state of the virtual impactor aerosol science has advanced dramatically over the last 30 years, and it is likely that a much improved virtual impactor design (minimizing the problems observed with the current design) could be developed if support for such research were made available.

It is worth noting here that any potential legal constraints against using a difference method for the PM-coarse FRM would not only prohibit the PM10-PM2.5 method, but may also limit the use of the dichotomous sampler, thus leaving no practical method for a filter-based (with gravimetric analysis) PM-coarse FRM. The dichot is also a difference method, although in a somewhat different way from paired samplers. PM-coarse for the dichot sampler is calculated as follows:

[a] calculate PM10 using the net mass from the sum of both the fine and coarse channel filters and the total sampler (inlet) flow;
[b] calculate PM2.5 using the net mass from the fine channel filter and the fine channel flow;
[c] calculate PM-coarse by subtracting PM2.5 in [b] from PM10 in [a].

The final step [c] is clearly a calculation of PM-coarse by difference. This dichot PM-coarse calculation description is not the one normally used in instrument manuals or network data reduction procedures, but is mathematically identical to those calculation procedures and is fundamentally clearer and simpler to implement. Any future documentation on how to calculate PM-coarse from dichot samplers should use this simpler approach.

There is potential for a PM-coarse method that avoids the limitations of both the difference and dichot methods that has not been evaluated. Collection by impaction is in principle a simple and direct measurement of PM-coarse. An example of such a method would be the existing PM2.5 FRM, with the WINS impactor substrate replaced with a suitable (weigh-able) filter media that retains the coarse-mode particles (e.g., does not have a significant “coarse particle bounce” problem). This is worth evaluating, since the approach has several advantages over present methods:

1. It uses existing technology already deployed in state/local networks.
2. Collection by impaction minimizes chemical artifacts relative to filtration methods.
3. It is the only way to get a direct measurement of PM-coarse.

The largest concern in an impaction-based PM-coarse method is particle bounce (loss) from the impaction substrate. The test results from an FRM with crystalized Dow 704 impactor oil as the

WINS surface (Vanderpool et al., *Aerosol Science and Technology*, **34**(5): 465-476, May 2001) imply that a suitable filter substrate (perhaps Fluoropore) is practical and worth evaluating. The Vanderpool tests with crystalized Dow oil showed that "...no large particle bounce from the crystallized oil surface was observed" and "...there appears to be no adverse effect of crystallized oil on the overall performance of the WINS separator nor on the PM2.5 concentration measurement." This reviewer strongly encourages that this approach be considered and properly evaluated as a candidate for routine field deployment.

Charge Question #2.

The demonstrated data quality of the PM10-2.5 difference method in the EPA field studies performed to date clearly supports it being proposed as a PM-coarse FRM; other than the greater than typical resources required to generate data of the high quality of this study, there are no detractions to this approach for PM-coarse measurement. There are limited studies done by state or local air agencies; the work done in the Birmingham AL area by the Jefferson County Dept. of Health is the only one available to this reviewer. This study included some sites (the Providence site for example, with ratios of mean PM2.5 to mean PM10 of 0.7) where the PM2.5 to PM-coarse ratios were substantially greater than one and mean PM-coarse concentrations were under 10 $\mu g/m^3$. These concentrations are typical of much of the eastern U.S., even in neighborhood urban sampling locations. This scenario is a tougher test of the PM-coarse difference method than the sites used in the EPA field test studies, but is essential in understanding the performance of the method under this common condition.

The Jefferson County tests were clearly performed under carefully controlled operating conditions; the mean CV for both collocated PM2.5 and PM10 samples was 1.4% for a single sampler (this needs clarification in the report; the CV reported there is from a pair of samplers, not a single sampler), an excellent result for the mean PM2.5 and PM10 concentrations sampled (15 and 24 $\mu g/m^3$ respectively). Despite the excellent PM2.5 and PM10 sampler precision as indicated by the reported CVs, the single-sampler PM-coarse CV was much higher at 5.7%. While still acceptable, this degradation in PM-coarse precision (relative to the PM2.5 and PM10 data used to generate the PM-coarse data) at sites with higher PM2.5 than PM-coarse is expected with the difference method. This effect is explained in detail in Allen et al. *(J. Air & Waste Manage. Assoc.*, **49**:PM, September 1999, pages 133-141), available at: http://www.awma.org/journal/special/Sept99/allen.pdf

As noted above, there has not been any assessment of the difference method for PM-coarse under the resource-constrained routine compliance monitoring network environments typically found at many State and local monitoring agencies. The reported PM-coarse difference method precision (CV) of 5.7% is likely to be much higher in these network environments, especially where PM2.5 dominates the PM10 concentrations.

During the panel discussions on other candidates for PM-coarse FRMs, the R&P coarse TEOM method was mentioned. Assuming proprietary methods are not candidates for designation for FRMs, that rules out this approach. One panelist commented that patents can always be worked around. In this case, that is not likely. The claims in US patent # 6,829,919 issued December

14, 2004 to Costas Sioutas and Paul Solomon for a "High-quality continuous particulate matter monitor" cover inlet flows over the range of 5 LPM to 100 LPM and a virtual impactor flow ratio of 2 to 50 (an inlet flow of 50 LPM and a flow ratio of 25 is used in the commercial version of the method). These claims are sufficiently broad to make it essentially impossible to create a public domain version of this method that does not infringe on this patent. As such, this method should not be considered as a candidate FRM.

To: Fred Butterfield, Designated Federal Officer
EPA SAB, Clean Air Scientific Advisory Committee (CASAC)
Ambient Air Monitoring and Methods Subcommittee

From: George Allen, AAMM subcommittee member, September 30, 2005

The following are revised written comments based on discussions during the September 21-22, 2005 meeting on a consultation on the development of criteria for a PM-coarse FEM. A copy of these comments is being sent to Dr. Barbara Zielinska and Mr. Rich Poirot, CASAC AAMM Subcommittee Co-Chairs. As requested in the memo from the AAMM co-chairs dated August 25, 2005, these comments address Question 8 in the EPA OAQPS memo to the SAB dated August 19, 2005. An additional comment on an approach to quantitatively defining "urban" PM-coarse is included at the end of these comments.

Consultation Question 8, associated with Attachment 6:
"What are the Subcommittee's views on the approach identified for the development of criteria to approve continuous $PM_{10-2.5}$ equivalent methods?"

Attachment 6 is titled "Technical Report on $PM_{10-2.5}$ Method Equivalency Development", by OAQPS and Battelle, dated August 19, 2005. As noted in the executive summary, the objective is to develop standards or criteria to compare candidate PM-coarse methods (continuous or "direct") to the PM-coarse FRM. The goal of this comparison is to insure that the quality of decisions made with regard to compliance with a PM-coarse NAAQS are as good as if the measurements were made with the FRM.

The process presented in this report is based on and is parallel to the process used to develop equivalency requirements for PM2.5 continuous FEMs (presently available in draft form and expected to be published for comment on December 20, 2005). My comments on this approach for the use of this technique for PM-coarse FEMs follow.

Section 2.1, Data Assumptions. This section details the collocated data set needed to determine equivalency. It appropriately requires multiple sites (the number not yet established), but does not discuss the need for a range of aerosol size distributions or concentrations for those sites. Only one site would be required to have tests performed during more than a single season; this may not be sufficient unless seasons are identified in advance and chosen to provide the most challenging aerosol for the candidate method. An example is Phoenix AZ; if a single season is

chosen it would have to be summer, not winter. For each "season" test period, samplers will be run for 30 days, with 75% (23 days) of valid data required (23 days where data from at least 2 of the three collocated pairs are valid). An acceptable concentration range of collected data will be set; this may further reduce the available number of days for analysis below 23. This reviewer recommends a target sample period of 40 collection days, not 30. That is more likely to produce a valid and useful comparison data set of 30 or more days, resulting in a more robust and meaningful statistical analysis.

General comments:

Among other metrics, the DQO acceptance method relies on the compensating effect of additive bias (regression intercept) and multiplicative bias (regression slope), resulting in an example acceptance range "window" for regression slopes and intercepts for a given FEM candidate as shown in figure 4. While this approach has some merit for qualifying FEMs for measurements to determine compliance with the PM2.5 annual mean standard metric, it does not work as well for a daily standard form such as the expected PM-coarse NAAQS. This is explained in the following paragraphs.

A 3-year annual mean metric is composed of up to 1095 daily samples with a population that has a large dynamic range -- samples close to zero and samples 4 to 5 times the "bright line" that is the NAAQS standard. The additive and multiplicative biases are combined in ways that tend to average out the method's errors with respect to the 3-year annual mean NAAQS metric (currently 15 $\mu g/m^3$).

A daily PM-coarse standard similar to that proposed in the final EPA PM staff paper (the 3-year mean of the 98th or 99th percentile annual values) is based on 3 samples that are typically clustered in a range of concentrations near the standard's value; as such, the sample population is very small and has a very limited dynamic range. In this situation, the only performance metric that is important in the context of NAAQS compliance is the method's response for samples slightly above or below the standard -- typically in the range of 80 to 90 $\mu g/m^3$ -- a relatively small dynamic range. Unlike an annual standard, how the method performs at levels well below the standard has no meaning in this daily standard compliance context. The question boils down to "how accurately does the FEM measure 85 $\mu g/m^3$".

To demonstrate the limitations of the example in Figure 4, if the PM-coarse standard were 85 (the upper range of the staff paper), a method that gave values ranging from 65 to 107 when the actual value was 85 would fit within the box and be an acceptable FEM. 107 ÷ 65 is 1.65 -- a rather wide range when determination of compliance with the NAAQS is the goal. It is recognized that the example given in Figure 4 is not necessarily the EPA's recommendation for PM-coarse FEM performance, but the underlying concerns remain.

Of course, determining compliance with NAAQS is not the only goal of PM-coarse measurements. To be useful to the health-effects community for future PM-health studies and for modeling purposes, a PM-coarse method must produce reasonably accurate data over the entire range of ambient values, and do this ideally on a time-frame much shorter than 24-hours. This is not the goal of the currently proposed FEM evaluation process, and beyond the scope of this

charge question. Still, this additional data use objective must be kept in mind when determining how to characterize what level of performance is adequate for designation of a method as a PM-coarse FEM. If we end up with a network of PM-coarse FEMs that do not generate data of sufficient quality at the relatively low (with respect to a likely standard) ambient concentration levels frequently observed in the eastern U.S., then we limit future progress in better understanding the health effects of PM-coarse.

There was some discussion during the second day of the meeting regarding how EPA might define "urban" PM-coarse for regulatory purposes. This reviewer would like to offer an approach to quantitatively define "urban" PM-coarse that was not discussed at the meeting. One of the underlying assumptions behind having PM-coarse be an urban-only standard is that urban PM-coarse particles are more harmful than a simple wind-blown "clean dust" because they have a mobile-source related chemical composition on the particle surface. A simple and quantitative assessment of the extent of how "urban" (mobile-source influenced) a coarse mode aerosol is can be made based on the color of the coarse mode aerosol. An example of this is readily observed by looking at the coarse-mode particles collected in the PM2.5 FRM WINS impactor. In core urban areas, those particles are black (very black, not just grey). In rural areas (areas without substantial local mobile source influence), the color is somewhere between a sandy color and a greyish-sandy color. There are optical reflectance methods that can quantify how "black" a particle deposit is. Thus, a quantitative assessment of how "urban" a site is can be made if sufficient samples are analyzed in this manner.

Dr. Judith Chow

September 29, 2005

To: Fred Butterfield, Designated Federal Officer
Clean Air Scientific Advisory Committee (CASAC)
Rich Poirot, Co-Chair – Monitoring
Barbara Zielinska, Co-Chair – Methods
CASAC Ambient Air Monitoring and Methods (AAMM) Subcommittee
From: Judith C. Chow, CASAC AAMM Subcommittee Member

Subject: CASAC Review of the Particle Methods and Data Quality Objectives

This memo addresses the questions on which the Subcommittee members were asked to comment regarding Attachment 1 ("Summary and Rationale for the $PM_{10-2.5}$ FRM") and Attachment 3 ["Memo to PM NAAQS Review Docket (OAR-2001-0017): Potential Changes being Evaluated for the $PM_{2.5}$ FRM"].

Questions on Attachment 1 (Summary and Rationale for the $PM_{10-2.5}$ FRM)

Question 1: What are the scientific and operational strengths and weaknesses of the $PM_{10-2.5}$ difference method relative to other options for a proposed FRM, especially when used as the basis for approval of other methods?

The $PM_{10-2.5}$ FRM is selected based on: 1) its ability to provide credible, reliable, and validated monitoring data for NAAQS attainment, and 2) its practicality in comparison with alternative methods to determine their qualifications as Federal Equivalent Methods (FEMs).

I agree with the statements in the "Summary and Rationale for the $PM_{10-2.5}$ FRM" (Attachment1) in that: 1) the difference method provides maximum comparability (e.g., filter medium, sample collection, gravimetric analysis, quality assurance procedures) to new or existing PM data sets, 2) $PM_{2.5}$ and PM_{10} sampling inlets have been wind tunnel tested to sustain a wide range of wind speeds and directions, 3) the $PM_{2.5}$ and PM_{10c} (PM_{10} reference sampler with replacement of the $PM_{2.5}$ FRM WINS impactor with straight downtube adaptor)

FRM samplers have been laboratory and field tested and evaluated, and 4) they are commercially available and network operators are familiar with them.

The statement (second paragraph on page 4) that: "An inherent advantage of a difference method is that some (additive) biases may be eliminated or substantially reduced by the subtraction" is not entirely true. There are several aspects of the proposed difference method that need to be considered:

- Potential variations in the changes of flow rates during the 24-hour sampling duration. Even though each $PM_{2.5}$ and PM_{10} sampler is equipped with a volumetric

flow control, as particle loading increases, the flow rate may alter differently with $PM_{2.5}$ and PM_{10} particles. The alteration of flow rates may not affect the cut points by very much (unless the flow is restricted as happens under certain pollution episodes), but it may bias the $PM_{10-2.5}$ measurements.

- Filter equilibration conditions need to be reconciled. For PM_{10} FRM weighing, filters are equilibrated at a set temperature between 15 °C and 30 °C with a variability not more than ± 3 °C and a set relative humidity (RH) between 20% and 45% with a variability not more than ± 5% for 24 hours prior to weighing. For $PM_{2.5}$ FRM weighing, filters are equilibrated for 24 hours at a set temperature between 20 °C and 23°C with a variability not more than ± 2 °C and a set RH between 30% and 40% with a variability not more than ± 5% (U.S. EPA, 1998). To minimize volatilization and water retention, a temperature of < 20 °C and RH < 20% would be preferable, although the RH may be difficult to control in humid environments. To ensure comparability, the weighing conditions for $PM_{10-2.5}$ need to conform to those of the $PM_{2.5}$ FRM.

- Sampling and reporting conditions need to be consistent. Currently, PM_{10} is adjusted to sea level pressure and a 25 °C temperature, whereas $PM_{2.5}$ sample volumes are intended to represent sampling conditions. PM_{10} samplers are intended to operate in actual conditions, but this isn't always the case because the calibrator conditions (usually standard) need to be adjusted to the actual conditions (which vary by season), then reconverted to the sea level conditions. Try to figure out the three temperature and pressure set-points for the PM_{10} TEOM sampler and see how often it isn't correct. This doesn't make a big difference in moderate coastal areas, but it is significant in mountainous areas with seasonal (and even diurnal) temperature extremes.

- Separate timers and pumps need to be synchronized. A single timer, temperature, and pressure sensor could be used to start, stop, and perform temperature and pressure corrections for both parallel units. A retrofit for existing samplers might be possible.

- The use of two samplers requires increased initial capital investment and perhaps site modification to accommodate the additional sampler. Also, it requires more of the operator's time for sampling and maintenance. Newer designs might combine both channels in a single unit with a common set of controls.

- If chemical analysis, such as X-ray fluorescence (XRF), is to be performed on these samples, it will add additional cost and uncertainties, as large-particle corrections (Dzubay and Nelson, 1975) only apply to the $PM_{10-2.5}$ fraction.

- There is less volatilized nitrate on PM_{10} samples than on $PM_{2.5}$ samples, probably owing to the adherence of nitric acid to alkaline soil particles and sea salt (Wu and Okada, 1994, Galy-Lacaux, 2001, Underwood et al., 2001). The different amounts of evaporation will make the coarse mass appear higher than it is in the atmosphere.

- At times and places where $PM_{2.5}$ constitutes most of the PM_{10} mass, negative values for $PM_{10-2.5}$ are possible, even though they may be within measurement error. The uncertainty of the difference [i.e., $F_{PM10-PM2.5} = (F_{PM10}^2 + F_{PM2.5}^2)^{1/2}$, Bevington, 1969] should be estimated and compared with the difference to determine its significance.

- To overcome the poor time-resolution of integrated 24-hour samples, EPA is encouraged to examine the difference between PM_{25} and PM_{10} BAM and TEOM mass and their equivalence or comparability with filter-based $PM_{10-2.5}$, Since many state and local agencies already own a $PM_{2.5}$ or PM_{10} BAMs or TEOMs, additional units can be added to acquire hourly $PM_{10-2.5}$ mass. For areas or seasons for which coarse particle volatization is not significant, the difference between the PM_{10} or $PM_{2.5}$ BAM or TEOM may be considered as a candidate for the equivalency method.

Question 2: Based on the field study report as well as any other available data e.g., data from State and local agencies, how does the demonstrated data quality of the $PM_{10-2.5}$ difference method support or detract from it being proposed as an FRM?

Data presented in the report on "Network Operations of the $PM_{10-2.5}$ Difference Method" by Vanderpool and Dillard (2004) demonstrate that when procedures are followed good precision can be obtained with the difference method for the Jefferson County, AL, aerosol. It would be useful to report the $PM_{10-2.5}$ mass concentrations with the propagated precisions, as described above, to confirm that measurements are identical within the measurement uncertainties. These tests represent seven sites in Jefferson County, AL, using BGI PQ200 samplers. PM in Jefferson County shows a high sulfate and moderate crustal aerosol. More tests are needed in dry, high crustal environments (e.g., Phoenix, Las Vegas) and at sites that have large fractions of nitrate and crustal material (e.g., Rubidoux, Fresno).

Questions on Attachment 3 [Memo to PM NAAQS Review Docket (OAR-2001-0017): Potential Changes being Evaluated for the $PM_{2.5}$ FRM]

Consultation Question 2: What are the Subcommittee's views on the Very Sharp Cut Cyclone (VSCC) being approved as an alternate, second-stage impactor to the Well Impactor Ninety-Six (WINS) for use on a $PM_{2.5}$ FRM?

Field experiments (e.g., Kenny et al., 2004) show that WINS impactors in $PM_{2.5}$ FRMs need frequent cleaning to retain their cut-points. The oil can also freeze at low temperatures, although alternatives are available (Hunike, 2000, Vanderpool et al., 2004). The oil can also get on the filters (Pitchford et al., 1997). The WINS was not intended for continuous $PM_{2.5}$ monitors, and may be impractical when used with them. Cyclone inlets don't use oil, have a high loading capacity, and can be easily cleaned (Chan and Lippmann, 1977; Gussman et al., 2002; Kenny and Gussman, 1997, 2000; John and Reischl, 1980; Kenny et al, 2000, 2004; Peters et al., 2001a). Sharp-cut cyclones (Kenny et al., 2000) have sampling effectiveness curves only slightly flatter than the WINS. The very sharp cut cyclone (VSCC, Kenny et al., 2004) has a sharper effectiveness curve and retains its D_{50} and effectiveness curve even under conditions with heavy loadings of 150 $\mu g/m^3$ (Kenny and Thorpe, 2001; Kenny et al., 2004).

Given the nature of the WINS impactor and the burden of frequent cleaning and oiling in the field (typically once every five runs) as compared to VSCC (once every 30 runs), it is reasonable to substitute the VSCC for the WINS on the FRM. More frequent cleaning than once every 30 runs (e.g., every 10 to 15 runs) for the VSCC should be required to ensure data quality. Replacement of WINS with VSCC in $PM_{2.5}$ FRM is approved as a Class II FEM (Federal Register, 2002). A slight loosening of the inlet effectiveness requirement would allow sharp cut cyclones to be used as well, probably with a negligible effect on the $PM_{2.5}$ mass measurement.

Past experience shows that some of the EPA-designated methods may perform poorly if the sampler is not well-maintained and the inlet is not frequently cleaned (Chow, 1995). While it is important to maintain consistency and data quality, the EPA should encourage rather than discourage vendors to apply for the Class III equivalent method designation for in-situ continuous monitors.

Consultation Question 3: To what extent are the stated advantages of relaxing existing requirements identified for the $PM_{2.5}$ FRM supported by the information cited in Attachment 3, available literature, or good field and laboratory practices? Does the Subcommittee have additional recommendations for the $PM_{2.5}$ FRM that would be neutral with respect to bias, but would improve the performance and minimize the burden on agencies conducting the sampling?

There are several published comparison studies of $PM_{2.5}$ FRMs among themselves and with other $PM_{2.5}$ samplers (Allen et al., 1997: Babich et al., 2000; Bardsley and Dal Sasso, 2005; Chung et al., 2001; Kenny et al., 2000; Lee et al., 2005a, 2005b; Long et al., 2003; Motallebi et al., 2003; Peters et al., 2001a, 2001b; Pitchford et al., 1997; Poor et al., 2002; Rizzo et al., 2003; Russell et al., 2004; Tanner and Parkhurst, 2000; Tropp et al., 1998; Yanosky et al., 2002). As one might expect, some show better agreement than others. Comparisons are poorer for environments with: 1) low concentrations, 2) a larger fraction of coarse particles, and/or 3) plentiful ammonium nitrate.

FRM prototype field tests in Birmingham, AL (Pitchford et al., 1997) showed oil drops on the filter due to the oil used in WINS splashing out of the impactor well. In addition, accumulation of particles in the impactor produced a cone-shaped deposit at a point just below the impactor jet; a slender needle developed at the top of the cone that extended above the oil surface. When the needle broke off, it contaminated downstream filter measurements. The alternative oil, diocty sebacate (DOS) seems to perform better than the previously specified diffusion oil, tetramethyltetraphenyltrisiloxane, CAS3982-82-9 (i.e., DOW 704) that overcomes crystallization under extreme atmospheric conditions (Vanderpool et al., 2004). I agree with the recommendation to use WINS with DOS oil as an approved equivalent method. This should be an independent FEM, irrespective of the FRM status of VSCC.

The recommendation of changing filter recovery time from 96 hours to 177 hours is favorable and will ease the burden of site visits from nearly twice to once per week. Field tests by Papp et al. (2002) did not report field blanks. Field blanks should be acquired and evaluated for passive deposition and gas adsorption over the different time periods in the sampler and in storage. Passive deposition may differ by sampler type. Papp et al. (2002) sampled different aerosols at

Seattle, Rubidoux, Austin, Athens, Augusta, and RTP during different seasons. However, five days per calendar quarter does not necessarily address the issues of positive and negative organic artifact (e.g., Eatough et al., 1990; McDow and Huntzicker, 1990; Gundel et al., 1995; Chow et al., 2005a) or nitrate volatilization (e.g., Hering and Cass, 1999; Chow et al., 2005b) with the additional 81 hours in the field. Chow et al. (2002) showed that nitrate volatilization is approximately 20% for 24 hours of filter recovery time after sampling, and approximately 44% for 72 hours of filter recovery time in Mexico City. For areas with high nitrate concentrations, such as Rubidoux, Fresno, and Bakersfield, CA, more tests during fall and winter are needed. With regard to the "Filter Transport Temperature and Post Sampling Recovery Time," EPA recommended the extension of post-sampling gravimetric analysis up to 30 days after the end of the sample period (i.e., assume the day of sample recovery), provided that samples are maintained at < 4 °C during transport from the field to the laboratory. The current equation (Mobley, 2002) to calculate the elapsed time for post-weighing is:

$$D(\text{Number of Days}) = 34 - T_{ave}\,(\text{Average Temperature in °C}) \qquad (1)$$

where:

$$T_{ave} = (T_{max} + T_{min})/2 \qquad (2)$$

Equation 2 is inexact since the guidance does not specify the method for measuring T_{max} or T_{min} when field operators pack the $PM_{2.5}$ FRM samples and record the temperatures. Our past experience with Texas' FRM $PM_{2.5}$ samples shows that T_{max} represents ambient temperature of the open cooler (i.e., T_{max} is recorded immediately after the thermometer is turned on and before the top layers of packed ice have covered the thermometer), rather than the cooler temperature at the time that field operator is packing the cooler (Tropp et al., 2003). This issue can be resolved if the procedure specifies that the thermometer is to be conditioned for 15 minutes before the T_{max} reading is taken.

During the past three years, on many occasions when T_{max} was 20 °C and T_{min} was -2 °C as recorded by the field operator before shipping, the cooler temperature remained at approximately 3 °C when the cooler was opened in our laboratory. According to Equations 1 and 2, the allocated days to perform gravimetric analysis would be 25 days instead of 31 days. Specific procedures to record T_{max} and T_{min} should be clarified to implement such a calculation.

Under the current procedure, one can retrieve, pack, and ship the filters within one day, and have them arrive at the laboratory the next day at a temperature of < 4 °C. One can then unload these filters from the cooler, and equilibrate the filters in the weighing laboratory (at approximately 21 °C) for up to 28 days before weighing, since the regulations only state that they must be conditioned at certain environmental conditions for at least 24 hours prior to weighing.

Additional tests in our laboratory during past years have shown that as long as filters are sealed (airtight), and stored at < 4 °C, post-gravimetric weights remained within the tolerance of microbalance precision even after two years of storage.

Additional Recommendations for the $PM_{2.5}$ FRM

- The current guidelines do not specify the chain-of-custody as field data (e.g., temperature, flow rate) are downloaded. To prevent alteration of the raw data file, the EPA is recommended to issue guidance for field data retrieval similar to that in U.S. EPA (1995).

- With respect to the Teflon-membrane filter, we have found the Pall Sciences (Ann Arbor, MI) PTFE (polytetrafluoroethylene) Teflon-membrane with PMP (polymethylpentene) support ring to be more versatile than the Whatman (Hillsboro, OR) PTFE Teflon-membrane with polypropylene support ring (Catalogue # 7592-104). Whatman Teflon filters (40 µm thickness) are 60% thicker than those from Pall Sciences (25 µm thickness), which results in poorer minimum detection limits (MDLs) for elements by X-Ray Fluorescence (XRF) analysis. Sometimes the FRM samples are used for elemental analysis. The advantage of lower MDLs outweighs the cost difference between the two types of Teflon-membrane filters. In the $PM_{10-2.5}$ fraction, elements will be the dominant component. Sampling on thinner filters for XRF analysis is desirable.

- For the Rupprecht and Patashnick (R&P) 2025 Partisol-Plus Sequential Filter Sampler (Rupprecht and Patashnick, Albany, NY), which is widely used in the $PM_{2.5}$ FRM network, the filter cassettes and magazines should be pre-labeled to minimize filter switching. This will be more of an issue when two samplers are used for $PM_{10-2.5}$, where misplacing the filter cassette or magazine can result in erroneous $PM_{10-2.5}$ masses. We have also observed that filter cassettes can flip over within the magazine if they are jostled too much. The FRM support grid is also an effective filter, as little mass gain was observed when air was drawn through the holder in the wrong direction.

References

Allen, G. A., Sioutas, C., Koutrakis, P., Reiss, R., Lurmann, F. W., and Roberts, P. T. (1997). Evaluation of the TEOM method for measurement of ambient particulate mass in urban areas. Journal of the Air & Waste Management Association 47(6), 682-689.

Babich, P., Davey, M., Allen, G., and Koutrakis, P. (2000). Method comparisons for particulate nitrate, elemental carbon, and $PM_{2.5}$ mass in seven U.S. cities. Journal of the Air & Waste Management Association 50(7), 1095-1105.

Bardsley, T. B. and Dal Sasso, N. A. (2005). Field trial of a TEOM®, FDMS TEOM® and manual gravimetric reference method for determination of $PM_{2.5}$ mass concentration. 1-5. 2005. Hobart, Tasmania. Proceedings, 17th Annual Meeting of the Clean Air Society of Australia and New Zealand. 5-3-2005.

Bevington, P. R. (1969). Data Reduction and Error Analysis for the Physical Sciences, McGraw Hill, New York, NY.

Chan, T. and Lippmann, M. (1977). Particle collection efficiencies of sampling cyclones: An empirical theory. Enivron. Sci. Technol. 11(4), 377-386.

Chow, J. C. (1995). Critical review: Measurement methods to determine compliance with ambient air quality standards for suspended particles. Journal of the Air & Waste Management Association 45(5), 320-382.

Chow, J. C., Watson, J. G., Edgerton, S. A., Vega, E., and Ortiz, E. (2002). Spatial differences in outdoor PM_{10} mass and aerosol composition in Mexico City.Journal of the Air & Waste Management Association 52(4), 423-434.

Chow, J. C., Watson, J. G., Lowenthal, D. H., Chen, L.-W. A., and Magliano, K. (2005a). Particulate carbon measurements in California's San Joaquin Valley. Chemosphere: in press.

Chow, J. C., Watson, J. G., Lowenthal, D. H., and Magliano, K. (2005b). Loss of $PM_{2.5}$ nitrate from filter samples in central California. Journal of the Air & Waste Management Association 55(8), 1158-1168.

Chung, A., Chang, D. P. Y., Kleeman, M. J., Perry, K. D., Cahill, T. A., Dutcher, D., McDougall, E. M., and Stroud, K. (2001). Comparison of real-time instruments used to monitor airborne particulate matter. Journal of the Air & Waste Management Association 51(1), 109-120.

Dzubay, T. G. and Nelson, R.O. (1975). Self absorption corrections for x-ray fluorescence analysis of aerosols, in *Advances in X-Ray Analysis, Vol. 18*, edited by W. L. Pickles, C. S. Barrett, J. B. Newkirk, and C. O. Rund, pp. 619-631, Plenum Publishing Corporation, New York, NY.

Eatough, D. J., Aghdaie, N., Cottam, M., Gammon, T., Hansen, L. D., Lewis, E. A., and Farber, R. J. (1990). Loss of semi-volatile organic compounds from particles during sampling on filters, in *Transactions, Visibility and Fine Particles*, edited by C. V. Mathai, pp. 146-156, Air & Waste Management Association, Pittsburgh, PA.

Federal Register. (2002). Final Rule: Consolidated emissions reporting, 40 CFR Part 51. Federal Register 67(111), 39602-39616. 6-10-2002.

Galy-Lacaux, C., Carmichael, G. R., Song, C. H., Lacaux, J. P., Al Ourabi, H., and Modi, A. I. (2001). Heterogeneous processes involving nitrogenous compounds and Saharan dust inferred from measurements and model calculations. Journal of Geophysical Research 106(D12), 12559-12578.

Gundel, L. A., Stevens, R. K., Daisey, J. M., Lee, V. C., Mahanama, K. R. R., and Cancel-Velez, H. G. (1995). Direct determination of the phase distributions of semivolatile polycyclic aromatic hydrocarbons using annular denuders. Atmospheric Environment 29(14), 1719-1733.

Gussman, R. A., Kenny, L. C., Labickas, M., and Norton, P. (2002). Design, calibration, and field test of a cyclone for PM_1 ambient air sampling. Aerosol Science & Technology 36(3), 361-365.

Hering, S. V. and Cass, G. R. (1999). The magnitude of bias in the measurement of $PM_{2.5}$ arising from volatilization of particulate nitrate from Teflon filters. Journal of the Air & Waste Management Association 49(6), 725-733.

Hunike, E. (2000). Memorandum: Alternative WINS oil. U.S. Environmental Protection Agency, Research Triangle Park, NC.

John, W. and Reischl, G. (1980). A cyclone for size-selective sampling of ambient air. Journal of the Air Pollution Control Association 30(8), 872-876.

Kenny, L. C. and Gussman, R. A. (1997). Characterization and modeling of a family of cyclone aerosol preseparators. J. Aerosol Sci. 28(4), 677-688.

Kenny, L. C. and Gussman, R. A. (2000). A direct approach to the design of cyclones for aerosol-monitoring applications. J. Aerosol Sci. 31(12), 1407-1420.

Kenny, L. C., Gussman, R. A., and Meyer, M. B. (2000). Development of a sharp-cut cyclone for ambient aerosol monitoring applications. Aerosol Science & Technology 32(4), 338-358.

Kenny, L. C. and Thorpe, A. (2001). Evaluation of VSCC cyclones. IR/L/EXM/01/01. Health and Safety Laboratory, Sheffield, UK.

Kenny, L. C., Merrifield, T., Mark, D., Gussman, R., and Thorpe, A. (2004). The development and designation testing of a new USEPA-approved fine particle inlet: A study of the USEPA designation process. Aerosol Science & Technology 38(Suppl.2), 15-22.

Lee, J. H., Hopke, P. K., Holsen, T. M., Polissar, A. V., Lee, D. W., Edgerton, E. S., Ondov, J. M., and Allen, G. (2005a). Measurements of fine particle mass concentrations using continuous and integrated monitors in eastern US cities. Aerosol Science & Technology 39(3), 261-275.

Lee, J. H., Hopke, P. K., Holsen, T. M., Lee, D. W., Jaques, P. A., Sioutas, C., and Ambs, J. R. L. (2005b). Performance evaluation of continuous $PM_{2.5}$ mass concentration monitors. J. Aerosol Sci. 36(1), 95-109.

Long, R. W., Eatough, N. L., Mangelson, N. F., Thompson, W., Fiet, K., Smith, S., Smith, R., Eatough, D. J., Pope, C. A., and Wilson, W. E. (2003). The measurement of $PM_{2.5}$, including semi-volatile components, in the EMPACT program: Results from the Salt Lake City Study. Atmospheric Environment 37(31), 4407-4417.

McDow, S. R. and Huntzicker, J. J. (1990). Vapor adsorption artifact in the sampling of organic aerosol: Face velocity effects. Atmospheric Environment 24A(10), 2563-2571.

Mobley, J. D. (2002). Memorandum: Extension of filter retrieval time period for $PM_{2.5}$ samples. U.S. Environmental Protection Agency, Research Triangle Park, NC. Motallebi, N., Taylor, C. A., Turkiewicz, K., and Croes, B. E. (2003). Particulate matter in California: Part 1 - Intercomparison of several $PM_{2.5}$, $PM_{10-2.5}$, and PM_{10} monitoring networks. Journal of the Air & Waste Management Association 53(12), 1509-1516.

Papp, M., Eberly, S. I., Hanley, T., Watkins, N., Barden, H., Noah, G., Bermudez, R., Eden, R., Franks, B., Johnson, A., Marriner, R., and Michel, E. (2002). Evaluation of filter recovery period for the determination of fine particulate matter as $PM_{2.5}$ in the atmosphere. U.S. Environmental Protection Agency, Research Triangle Park, NC.

Peters, T. M., Gussman, R. A., Kenny, L. C., and Vanderpool, R. W. (2001a). Evaluation of $PM_{2.5}$ size selectors used in speciation samplers. Aerosol Science & Technology 34(5), 422-429.

Peters, T. M., Vanderpool, R. W., and Wiener, R. W. (2001b). Design and calibration of the EPA $PM_{2.5}$ well impactor ninety-six (WINS). Aerosol Science & Technology 34(5), 389-397.

Pitchford, M. L., Chow, J. C., Watson, J. G., Moore, C. T., Campbell, D. E., Eldred, R. A., Vanderpool, R. W., Ouchida, P., Hering, S. V., and Frank, N. H. (1997). Prototype $PM_{2.5}$ federal reference method field studies report -An EPA staff report. U.S. Environmental Protection Agency, Las Vegas, NV.

Poor, N., Clark, T., Nye, L., Tamanini, T., Tate, K., Stevens, R., and Atkeson, T. (2002). Field performance of dichotomous sequential PM air samplers. Atmospheric Environment 36(20), 3289-3298.

Rizzo, M., Scheff, P. A., and Kaldy, W. (2003). Adjusting tapered element oscillating microbalance data for comparison with Federal Reference Method $PM_{2.5}$ measurements in Region 5. Journal of the Air & Waste Management Association 53(5), 596-607.

Russell, M., Allen, D. T., Collins, D. R., and Fraser, M. P. (2004). Daily, seasonal, and spatial trends in $PM_{2.5}$ mass and composition in Southeast Texas. Aerosol Science & Technology 38(Suppl. 1), 14-26.

Tanner, R. L. and Parkhurst, W. J. (2000). Chemical composition of fine particles in the Tennessee Valley region. Journal of the Air & Waste Management Association 50(8), 1299-1307.

Tropp, R. J., Jones, K., Kuhn, G., and Berg, Jr., N. J. (1998). Comparison of $PM_{2.5}$ saturation samplers with prototype $PM_{2.5}$ Federal Reference Method Samplers, in *Proceedings, $PM_{2.5}$: A Fine Particle Standard*, edited by J. C. Chow and P. Koutrakis, pp. 215-225, Air & Waste Management Association, Pittsburgh, PA.

Tropp, R. J., Engelbrecht, J. P., Lowenthal, D. H., Kohl, S. D., Dickerson, A. L., DuBois, D. W., Chow, J. C., Watson, J. G., Countess, R. J., and Countess, S. J. (2003). Assessment of $PM_{2.5}$ chemical speciation results for Texas: Part 1. 6470-665-2925.1D1. Desert Research Institute, Reno, NV.

U.S. EPA. (1995). Good automated laboratory practices: Principles and guidance to regulations for ensuring data integrity in automated laboratory operations. U.S. Environmental Protection Agency, Research Triangle Park, NC.

U.S. EPA. (1998). Quality assurance guidance document 2.12: Monitoring $PM_{2.5}$ in ambient air using designated reference or class I equivalent methods. U.S. Environmental Protection Agency, Research Triangle Park, NC.

Underwood, G. M., Song, C. H., Phadnis, M., Carmichael, G. R., and Grassian, V. H. (2001). Heterogeneous reactions of NO_2 and HNO_3 on oxides and mineral dust: A combined laboratory and modeling study. Journal of Geophysical Research 106(D16), 18055-18066.

Vanderpool, R. W., Byrd, L., Wiener, R. W., Hunike, E., Labickas, M., Leston, A., Tolocka, M. P., McElroy, F. C., Murdoch, R. W., Natarajan, S., and Noble, C. A. (2004). Laboratory and field evaluation of crystallized Dow 704 oil on the performance of the WINS $PM_{2.5}$ fractionator, in *Proceedings, Symposium on Air Quality Measurement Methods and Technology-2004*, pp. 23-1-23-17, Air and Waste Management Association, Pittsburgh, PA.

Vanderpool, R. W. and Diller, D. J. (2005). Network operation of the $PM_{10-2.5}$ difference method. U.S. Environmental Protection Agency, Research Triangle Park, NC. Wu, P. M. and Okada, K. (1994). Nature of coarse nitrate particles in the atmosphere –A single particle approach. Atmospheric Environment 28(12), 2053-2060.

Yanosky, J. D., Williams, P. L., and MacIntosh, D. L. (2002). A comparison of two directreading aerosol monitors with the Federal Reference Method for $PM_{2.5}$ in indoor air. Atmospheric Environment 36, 107-113.

Dr. Ellis Cowling

Comments by Ellis Cowling in Connection with the CASAC Ambient Air Monitoring and Methods (AAMM) Subcommittee Peer Review of the $PM_{10-2.5}$ Federal Reference Method and Consultation on Field Evaluation of $PM_{10-2.5}$ Methods, and Related Matters as Discussed During the CASAC AAMM Subcommittee Meeting in Durham, NC on September 21-22, 2005

After reviewing the documents provided and listening carefully to the presentations made during this Peer Review and Consultation with the CASAC AAMM Subcommittee, I offer the following two general recommendations, the first of which was mentioned in an earlier CASAC peer review and consultation on EPA's National Air Monitoring Strategy.

Recommendation 1:

EPA should guard against the tendency to give undue emphasis to "Data Quality Objectives" in the selection and evaluation of instruments and subsequent implementation of field monitoring programs to the exclusion of concern about "Science Quality Objectives" and "Policy Relevancy Objectives."

Experience within the Southern Oxidants Study and other large-scale field measurement and monitoring campaigns has demonstrated repeatedly that undue emphasis on "Data Quality Objectives" often leads to:

1) Serious lack of attention to the scientific hypotheses and assumptions that are inherent in any choice of scientific instruments, the appropriateness of the ground-based sites at which the instruments are located, the skills of the instrument operators, the data processing and data-display programs used, and especially the scientific quality of the conclusions and statements of findings that are drawn from analysis and interpretation of the measurements that are made; and
2) Equally serious lack of attention to the policy relevancy of the measurements being made -- relevancy to the general or specific enhancements of environmental protection that are the real reason behind the public health and public welfare concerns that led to the decision to establish a monitoring program in the first place.

A very important example of the need for inclusion of "Science Quality Objectives" together with "Data Quality Objectives" in implementation plans for air quality monitoring programs was provided during the verbal presentation during the meeting in Durham on September 21, 2005. As Tim Hanley pointed out, a very prevalent worry about use of subtraction methods in reporting ambient air quality measurements was the widely believed frequency with which negative numbers show up in reports of air concentrations where subtraction methods are used. When Hanley and his colleagues did very careful analyses of the actual numbers used in making the subtractions that resulted in negative air concentration numbers, it became clear that the data actually used frequently were from instruments that were not co-located and thus were not measuring pollutant concentrations at the same location, or were for measurements made at different times of the day, so that the two numbers being subtracted were not for the same air parcel! Thus Hanley et al concluded that

the prevalent worry about negative numbers was largely a myth – which he suggested (for our enjoyment!) was similar to reports of "observations" of a "Loch Ness Monster" in Scottish fairy tales!

The many years of experience accumulated by the scientists and engineer in the Southern Oxidants Study show that careful analysis, careful interpretation, and careful formulation of statements of findings from air quality measurements is the best quality assurance method of all.

This is why I suggest that any program of ambient air quality measurements (including the program for measurement and reporting of air concentrations of $PM_{10-2.5}$) should have explicitly stated "Science Quality Objectives" and well as explicitly stated "Data Quality Objectives."

In this latter connection, permit me to call attention to the attached "*Guidelines for the Formulation of Scientific Findings to be Used for Policy Purposes.*" These guidelines were developed originally by the NAPAP Oversight Review Board led by Milton Russell, former Assistant Administrator for EPA. Please find attached on page 4 below, an electronic version of these Guidelines which the Southern Oxidants Study adopted and very slightly adapted for use in formulating policy relevant scientific findings from our research in 1988-2005.

The original version of these Guidelines was published as Appendix III of the April 1999 Report titled "*The Experience and Legacy of NAPAP*." This was a Report to the Joint Chairs Council of the Interagency Task Force on Acidic Deposition of the Oversight Review Board (ORB) of the National Acid Precipitation Assessment Program. As indicated in Appendix III:

> "The following guidelines in the form of checklist questions were developed by the Oversight Review Board to assist scientists in formulating presentations of research results to be used in policy decision processes. These guidelines may have broader utility in other programs at the interface of science and public policy and are presented here with that potential use in mind."

Recommendation 2:

This second recommendation derives from the general principle that **"*The words we use often show the quality of our understanding.*" This general principle of communication about public policy matters leads me to join with Peter McMurry and others in the AAMM Subcommittee in recommending that care should be taken in the choice of words that EPA uses (and all the rest of us use!) to describe:**

1) The **$PM_{10-2.5}$ particles** themselves,
2) The **$PM_{10-2.5}$ monitoring network** established to measure and report air concentrations of these particles in various parts of the country, and
3) The **$PM_{10-2.5}$ National Ambient Air Quality Standard(s)** that are established to help decrease human-health and public-welfare risks associated with exposure to $PM_{10-2.5}$ particles.

The following words often (or occasionally) are used more-or-less interchangeably in describing these three different entities even though these various words are not necessarily equivalent and in many cases do not portray the same understanding about these important entities:

1) **Words used to describe the $PM_{10-2.5}$ particles themselves**:
   - coarse particles (CP),
   - coarse aerosol particles (CAP),
   - urban coarse particles (UCP),
   - coarse particulate matter (CPM),
   - urban coarse particulate matter (UCPM),
   - thoracic particles (TP),
   - urban thoracic particles (UTP),
   - urban thoracic coarse particles (UTCP),
   - urban thoracic coarse particulate matter (UTCPM),
   - $PM_{10-2.5}$ particles ($PM_{10-2.5}P$),
   - urban $PM_{10-2.5}$ particulate matter ($U_{10-2.5}PM$);

2) **Words used to describe the $PM_{10-2.5}$ monitoring network**:
   - Coarse Particle Network (CPN),
   - Coarse Aerosol Monitoring Network (CAMN),
   - Urban Coarse Particle Network (UCPMN),
   - Coarse Particulate Matter Network (CPMN),
   - Urban Coarse Particle Network UCPN),
   - Thoracic Particle Network (TPN),
   - Thoracic Coarse Particle Network (TCPN),
   - Urban Thoracic Coarse Particle Network (UTCPN),
   - Urban Thoracic Coarse Particulate Matter Network (UTCPMN),
   - $PM_{10-2.5}$ Network ($PM_{10-2.5}N$),
   - $Urban_{10-2.5}$Particulate Matter Network ($U_{10-2.5}PMN$);

3) **Words used to describe the $PM_{10-2.5}$** National **Ambient Air Quality Standard (NAAQS)**:
   - Coarse Particle Standard (CPS),
   - Coarse Aerosol Standard (CAS),
   - Urban Coarse Particle Standard (UCPS),
   - Coarse Particulate Matter Standard (CPMS),
   - Urban Coarse Particulate Matter Standard (UCPMS),
   - Thoracic Particle Standard (TPS),
   - Urban Thoracic Particle Standard (UTPS),
   - Urban Thoracic Coarse Particle Standard (UTCPS),
   - Urban Thoracic Coarse Particulate Matter Standard (UTCPMS),
   - $PM_{10-2.5}$ Particle Standard ($PM_{10-2.5}PS$),
   - Urban $PM_{10-2.5}$ Particulate Matter Standard ($U_{10-2.5}PMS$).

GUIDELINES FOR THE FORMULATION OF STATEMENTS OF SCIENTIFIC FINDINGS TO BE USED FOR POLICY PURPOSES

The following guidelines in the form of checklist questions were developed by the NAPAP Oversight Review Board to assist scientists and engineers in formulating statements of research findings to be used in policy decision processes.

1) **IS THE STATEMENT SOUND?** Have the central issues been clearly identified? Does each statement contain the distilled essence of present scientific and technical understanding of the phenomenon or process to which it applies? Is the statement consistent with all relevant evidence-evidence developed either through NAPAP [or SOS] research or through analysis of research conducted outside of NAPAP [or SOS]? Is the statement contradicted by any important evidence developed through research inside or outside of NAPAP [or SOS]? Have apparent contradictions or interpretations of available evidence been considered in formulating the statement of principal findings?

2) **IS THE STATEMENT DIRECTIONAL AND, WHERE APPROPRIATE, QUANTITATIVE?** Does the statement correctly quantify both the direction and magnitude of trends and relationships in the phenomenon or process to which the statement is relevant? When possible, is a range of uncertainty given for each quantitative result? Have various sources of uncertainty been identified and quantified, for example, does the statement include or acknowledge errors in actual measurements, standard errors of estimate, possible biases in the availability of data, extrapolation of results beyond the mathematical, geographical, or temporal relevancy of available information, etc. In short, are there numbers in the statement? Are the numbers correct? Are the numbers relevant to the general meaning of the statement?

3) **IS THE DEGREE OF CERTAINTY OR UNCERTAINTY OF THE STATEMENT INDICATED CLEARLY?** Have appropriate statistical tests been applied to the data used in drawing the conclusion set forth in the statement? If the statement is based on a mathematical or novel conceptual model, has the model or concept been validated? Does the statement describe the model or concept on which it is based and the degree of validity of that model or concept?

4) **IS THE STATEMENT CORRECT WITHOUT QUALIFICATION?** Are there limitations of time, space, or other special circumstances in which the statement is true? If the statement is true only in some circumstances, are these limitations described adequately and briefly?

5) **IS THE STATEMENT CLEAR AND UNAMBIGUOUS?** Are the words and phrases used in the statement understandable by the decision makers of our society? Is the statement free of specialized jargon? Will too many people misunderstand its meaning?

6) **IS THE STATEMENT AS CONCISE AS IT CAN BE MADE WITHOUT RISK OF MISUNDERSTANDING?** Are there any excess words, phrases, or ideas in the statement that are not necessary to communicate the meaning of the statement? Are there so many caveats in the statement that the statement itself is trivial, confusing, or ambiguous?

7) **IS THE STATEMENT FREE OF SCIENTIFIC OR OTHER BIASES OR IMPLICATIONS OF SOCIETAL VALUE JUDGMENTS?** Is the statement free of influence by specific schools of scientific thought? Is the statement also free of words, phrases, or concepts that have political, economic, ideological, religious, moral, or other personal-, agency-, or organization-specific values, overtones, or implications? Does the choice of how the statement is expressed rather than its specific words suggest underlying biases or value judgments? Is the tone impartial and free of special pleading? If societal value judgments have been discussed, have these judgments been identified as such and described both clearly and objectively?

8) **HAVE SOCIETAL IMPLICATIONS BEEN DESCRIBED OBJECTIVELY?** Consideration of alternative courses of action and their consequences inherently involves judgments of their feasibility and the importance of effects. For this reason, it is important to ask if a reasonable range of alternative policies or courses of action have been evaluated? Have societal implications of alternative courses of action been stated in the following general form?:

"If this [particular option] were adopted then that [particular outcome] would be expected."

9) **HAVE THE PROFESSIONAL BIASES OF AUTHORS AND REVIEWERS BEEN DESCRIBED OPENLY?** Acknowledgment of potential sources of bias is important so that readers can judge for themselves the credibility of reports and assessments.

Mr. Bart Croes

U.S. EPA's Coarse PM FRM and Other PM Monitoring Issues September 21-22, 2005 Peer Review and Consultation Meeting CASAC AAMM Subcommittee Review Comments, Bart Croes

Overall, the documents the Subcommittee reviewed continue the impressive initiative by U.S. EPA to take a systematic approach towards implementation of a likely coarse particle ($PM_{10-2.5}$) National Ambient Air Quality Standard (NAAQS). I appreciate the opportunity to comment during this near-final stage of the process, about a year after we provided input on intermediate results. The documents provide a good description of the basis for a $PM_{10-2.5}$ Federal Reference Method (FRM), clearly explain the new results from the multi-site evaluation of candidate methods, and provide a reasonable rationale for changes to the $PM_{2.5}$ FRM and development of a $PM_{10-2.5}$ Federal Equivalent Method (FEM) process. I agree with the basic approach taken by U.S. EPA, and offer comments on several aspects that need further attention. My comments address the two peer review and eight consultation questions posed by Phil Lorang in his August 19, 2005 memo to Fred Butterfield. These comments reflect considerable input from California Air Resources Board (CARB) staff responsible for implementing U.S. EPA monitoring requirements and using the data in source apportionment and health studies.

Peer Review Questions:

1. What are the scientific and operational strengths and weaknesses of the $PM_{10-2.5}$ difference method relative to other options for a proposed FRM, especially when used as the basis for approval of other methods?

The difference method has several major strengths over other FRM options; namely that it uses monitoring equipment already deployed as part of the Nation's large investment in the $PM_{2.5}$ FRM. It provides a consistent basis for comparison and forces cutpoint curves in candidate samplers to conform to those of the existing $PM_{2.5}$ FRM. From an operational perspective, the equipment and procedures to operate a PM_{10} low-volume sampler are almost identical to those of the $PM_{2.5}$ FRM filter sampler, therefore the PM_{10} and $PM_{2.5}$ FRM manual filter methods should integrate easily into existing $PM_{2.5}$ monitoring networks. California re-established its PM_{10} ambient air quality standards in 2002 (and set a new annual-average $PM_{2.5}$ standard). The use of the PM_{10} low-volume sampler for the $PM_{10-2.5}$ FRM will enable monitoring agencies such as CARB to maintain compliance with current State PM_{10} monitoring requirements. Another advantage is that filters can be analyzed for particle composition, although it is unclear whether or not errors in the difference method for key tracer species are sufficiently small to allow receptor models to be applied, in contrast to the dichot method where direct chemical analysis of coarse particles is possible.

The major drawbacks of the difference method are the same as those of other filter-based approaches, namely, the lack of 24-hour time resolution and expensive, manual filter handling and analysis. Time-resolved, real-time availability of PM data are necessary for use air quality index (AQI) forecasting and burn allocations, and can lead to a better understanding of emission

sources, transport, background levels, deposition, and health effects of PM, although U.S. EPA is addressing this need through the FEM process. As in previous monitoring programs for airborne particles (i.e., TSP, PM_{10}, $PM_{2.5}$), it will not be advantageous to consider a method that is only useful to answer questions of attainment if it limits one in making other important observations. The difference method will also prove cumbersome and expensive because it requires great care in shipping and filter handling by experienced personnel. This means that widespread deployment of the difference method in field operations is likely to result in less data being collected at fewer sites then for a real-time instrument. The sampler difference method requires high quality data from collocated instruments, and despite the excellent precision results obtained in the multi-site sampler inter-comparison study by U.S. EPA, is subject to more errors deriving from both instrument operations and operator error than single-instrument approaches. It has yet to be demonstrated that the difference method can function in low-PM environments and produce useful data, although this will not be an issue if only a 24-hour (and not an annual-average) $PM_{10-2.5}$ standard is promulgated.

All of these issues are addressable if the FEM process qualifies continuous, real-time methods(s) and perhaps a single-instrument filter sampler for speciation analyses. I am in agreement with U.S. EPA on using the difference method as an FRM. After all, a difference method is already successfully used to determine NO_2 levels. However, since the number of $PM_{10-2.5}$ non-attainment areas is expected to be much smaller than for $PM_{2.5}$ and many of these areas have existing PM_{10} monitoring programs, U.S. EPA could limit deployment of the cumbersome difference method for $PM_{10-2.5}$ if it allowed PM_{10} monitors to be used to determine attainment. U.S. EPA and SLT agencies have already invested huge resources into the current PM_{10} and $PM_{2.5}$ monitoring networks. Several states (i.e., California) have State ambient air quality standards for PM_{10} and do not plan to follow U.S. EPA in adopting a coarse particle standard. Surely if a site meets the $PM_{10-2.5}$ standard with PM_{10} monitoring data (uncorrected), then there is no need to deploy a $PM_{10-2.5}$-specific monitor at the site.

The potential scope of a national $PM_{10-2.5}$ monitoring network should be defined. While U.S. EPA has not yet promulgated a $PM_{10-2.5}$ NAAQS, it has released a Staff Paper with a proposed range of possible standards for $PM_{2.5}$ and $PM_{10-2.5}$. As a first-order estimate, data from the existing PM_{10} monitoring network should be compared to the proposed lower and upper ranges of the coarse particle recommendations to determine if the potential scope of a $PM_{10-2.5}$ monitoring network would be national in scale or restricted to a few states. In these likely non-attainment areas, PM_{10} would primarily consist of the coarse fraction. Sites that have collocated $PM_{2.5}$ and PM_{10} monitors, or SLT agencies that have operated dichot samplers (see Motallebi, et al., 2003ab for California) provide more relevant data. A list and map of sites with PM_{10} only, $PM_{2.5}$ only, and both would be a useful summary.

On a parallel basis, the U.S. EPA should devote resources (preferably a STAR grant) to developing a traceable standard for PM. Problems with the TEOM, APS, and other units were only discovered during the inter-comparison study because multiple units were carefully operated by U.S. EPA and monitoring industry experts. If the units were operating by themselves in an SLT agency monitoring station, it is unlikely that instrument drift and other problems would have been noticed. Without the ability to challenge a PM analyzer with a know concentration of PM, all you have to verify proper operation of an analyzer is the "due diligence"

of the site technician. If U.S. EPA took a dozen of the candidate samplers and sent them to 12 randomly selected SLT agencies, after four months they would get a dozen different regressions and correlations, no matter how consistent the analyzers performed in the controlled, three-city study.

Other continuous, criteria pollutant monitors (O_3, NO_2, CO, SO_2) are challenged with a known concentration each day (the in-station zero and span checks) and at six- and twelve-month intervals (independent transfer standards). For filter samplers the micro-balance used to weigh the filter is similarly "zeroed and spanned" with NIST-traceable standard weights each weighing session. Ozone does not come in a bottle, but accurate and precise quantities are generated on demand to challenge ozone analyzers. Resources should be devoted to research an accurate and precise PM generation system. I realize this would be very difficult to do with PM, but perhaps something similar to an aerosol inhaler (used for administering asthma medication) could be developed.

When simple comparisons of the FRM-like difference method to other approaches is undertaken it is very easy to forget that the FRM does not really report on PM as it is in the real world, rather it represents the result of highly controlled sampling and filter handling that is known to, in some conditions, to over- or under-report ambient PM. It may not serve the reader to include statements such as are in Attachment 1 as follows: "…the proposed $PM_{10-2.5}$ FRM utilizes the same fundamentally sound integrated sample, filter-collection, and mass-based gravimetric measurement technology that has been basic to all previous FRMs…" It is known that organic and nitrate compounds are not retained fully in these "fundamentally sound integrated samples". It is also well known that moisture either on particles or retained by the filter can be extremely influential in PM mass measurements. In general, the text presented on page four of this attachment is far too simplistic.

2. Based on the field study report as well as any other available data, e.g., data from State and local agencies, how does the demonstrated data quality of the $PM_{10-2.5}$ difference method support or detract from it being proposed as a FRM?

For sites operating both PM_{10} and $PM_{2.5}$ FRM samplers, the sampler difference method will be very cost-effective, but it is dependent on successful collection of both samples, thus sample recovery and data completeness would be hostage to the degree of success of operation of the other two samplers. For sites lacking one or the other sampler (e.g., established attainment areas), it may be preferable to operate a "stand-alone" coarse particle sampler, thus adopting a difference-based FRM should be accompanied by adoption of a single-instrument equivalent method.

Of the candidate technologies, only the difference method and the dichot method offer the significant advantage of traceability to primary standards for both the mass and volume (flow × time) components of the $PM_{10-2.5}$ mass concentration**.** However, our experience with dichotomous samplers has shown that they can be more difficult to calibrate, maintain, and audit than the FRM samplers. This is due to the fact that both the $PM_{2.5}$ and $PM_{10-2.5}$ flowrates are interdependent and changes/adjustments to one affects the other. Given this, and the reality that the laboratory resources are identical for both the difference method (two pre-weights, two post

weights) and the dichot method (two pre-weights, two post weights), we support selection of the difference method as the $PM_{10-2.5}$ FRM.

Decisions regarding the deployment of the various continuous $PM_{10-2.5}$ samplers should not be made solely on equivalence to the FRM difference approach. There may be varying valid reasons to select one or more of these methods beyond simple equivalency. For example, the ability to address coarse or fine PM levels on an hourly or sub-hourly time interval could allow experimental investigations of short-term health impacts and chemical/photochemical processes that produce ambient PM. It is probably more important to understand the fundamental reasons why the various alternative methods vary from the FRM than to simply establish a few points of difference. Application of rigorous criteria for designation of equivalence for $PM_{10-2.5}$ methods, such as are presented in Attachment 4 (for $PM_{2.5}$), may be quite cumbersome and problematic. By definition, it would seem that only the FRM difference method can be expected to produce fully equivalent results. Perhaps a functional equivalency based on practical considerations would be more useful. However, discussions of functional equivalency issues are quite sparse in the documents provided.

The attachment contains lengthy considerations related to the need to avoid selection of proprietary sampler technologies. However, if one views the nature of the current $PM_{2.5}$ network, we find that only two or three companies sell FRMs and most of these are sold by R&P. Few companies have the resources to develop, test, and comply with all the factors needed to reach acceptance and those that do consider the probable market when they determine to propose devices. Should a beta gauge technology be chosen, there are no more than three manufacturers worldwide (with Metone being the dominant U.S. vendor), and there is only one for the TEOM or the APS. What this means is that it is likely that a proprietary sampler technology is almost inevitable should any continuous device be selected or certified.

If the difference method is selected as the FRM for $PM_{10-2.5}$, how will negative numbers, however infrequently they occur, be handled? Will negative values be reported, invalidated, corrected to zero, or corrected to detection limit values?

Consultation Questions:

1. Based upon the latest available field study data, which $PM_{10-2.5}$ methods have both sufficient utility to meet one or more important monitoring objectives and appropriate data quality to be considered for deployment as Federal Equivalent Methods (FEMs) or speciation in a potential $PM_{10-2.5}$ monitoring network?

Simply documenting exceedances of PM standards, without collecting a sample that can by subjected to chemical or physical analysis, is a poor allocation of resources. Physical samples need not be analyzed for more than mass on a routine basis, but an archived set of samples available for further analyses can be an invaluable aid to determining causes of past exceedances and as a cost- and time-saving resource for a wide range of atmospheric process studies. Methods that do not collect samples (e.g., TEOM) should not be used in areas of marginal attainment or non-attainment of standards where regulators can reasonably anticipate that source-related analyses will be needed. These considerations all point to a next-generation (i.e., tighter

2.5 μm cutpoint) dichotomous sampler as the most flexible measurement method for attainment determination and source assessment.

Only the manual dichot sampler appears capable of meeting the objective of "high FRM comparability". Though not in itself a speciation method, the dichot is the best method to collect $PM_{10-2.5}$ samples for subsequent laboratory speciation. At this time, there does not appear to be a validated candidate for providing highly time-resolved data to support a $PM_{10-2.5}$ standard, although several methods appear promising and U.S. EPA's efforts to evaluate candidate devices should continue.

The field study document indicated that the next (and final) intercomparsion study would take place at Birmingham, AL. Two alternative sites (in likely nonattainment areas) would present more challenging environments. Riverside, CA has high levels of $PM_{10-2.5}$ and $PM_{2.5}$ nitrate during the late summer and fall, and nitrate volatilization losses could be evaluated for the candidate samplers. The remaining PM Supersite in Fresno, CA also has high $PM_{10-2.5}$ and $PM_{2.5}$ nitrate during the fall harvest season. SLT agencies should be involved to duplicate "real-world" operation. Perhaps the existing dichot monitor and high-volume PM_{10} and low-volume $PM_{2.5}$ monitors (that are already deployed by some SLT agencies) should be included to determine their suitability to determine if an area meets the NAAQS (e.g., they do not have negative biases).

The initial text for this attachment should clearly delineate the objectives for the tests conducted. The question requests input on whether methods tested "…meet one or more important monitoring objectives…to be considered for deployment as Federal Equivalent Methods…" It is somewhat unclear whether the text describes the results of comparison studies conducted to determine the performance of devices that may be considered as $PM_{10-2.5}$ FRMs or as FEM monitors. The document or other text should contain descriptions of these possible monitoring objectives as a part of framing the testing approach and presentation of results. For example, it should describe the data needs of health researchers, air modelers, those performing compliance ascertainment, or others. While the researchers/planners and the U.S. EPA authors of Attachment 2 may have an idea of data needs and how various monitors might meet these needs, this would be useful information for the regulated community and the regulatory agencies charged with carrying out the monitoring program.

2. What are the Subcommittee's views on the Very Sharp Cut Cyclone (VSCC) being approved as an alternate second-stage impactor to the WINS for use on a $PM_{2.5}$ FRM?

Replacing the WINS with the VSCC has strong technical merit. For the BGI, R&P, and Andersen $PM_{2.5}$ FRM, the U.S. EPA has already designated the BGI VSCC $PM_{2.5}$ size inlet modification as Federal Equivalent Methods. The VSCC requires less maintenance than the WINS impactor. The VSCC also provides improved cut point characteristics and does not use any oil. However, mandated network-wide retrofit of installed FRM samplers should be avoided unless new data indicate that the "old" methods pose problems (i.e., let the local operators update at their discretion).

3. To what extent are the stated advantages of relaxing existing requirements identified for the $PM_{2.5}$ FRM supported by the information cited in Attachment 3, available literature, or good field and laboratory practices? Does the Subcommittee have additional recommendations for the $PM_{2.5}$ FRM that would be neutral with respect to bias, but would improve the performance and minimize the burden on agencies conducting the sampling?

The other changes recommended by U.S. EPA staff also have strong technical merit.

WINS Oil: The use of the alternative DOS oil is reasonable, although utilization of the VSCC (see above) eliminates the need for any impaction oil.

Filter Recovery Time: Although the 177-hour recovery time may ease operator time schedules, the effort to pick up exposed filters as soon as possible after sampling should continue. Concerns about greater opportunities for changes on exposed filters should not be over-weighted since most of the artifacts of concern (e.g., nitrate loss, water loss/gain) are not adequately prevented by the present procedures.

Filter Transport Temperature and Post Sampling Recovery Time: PM_{10} and $PM_{2.5}$ filters are equilibrated at 25°C for a minimum of 24 hours prior to post weighing. Given this, is maintaining 4°C from post sampling pickup and transportation necessary? Is there evidence to support the significance of maintaining 4°C or less during transport of filter samples? Shipping costs would lessen if transportation conditions were relaxed to 25°C or below.

4. Considering the statistical measures of precision, correlation, multiplicative bias, and additive bias identified for approval of $PM_{2.5}$ continuous methods, what are the Subcommittee's views on the usefulness of each measure to ensure that approved or equivalent methods meet the monitoring network data quality objectives?

The methods presented in the September 30, 2004 draft are reasonable. However, U.S. EPA should bear in mind that there are three goals of equivalency:

- Equal protection of public health across a wide range of administrative and regulatory settings.
- Equal treatment of the regulated community.
- Flexibility for local regulators to tailor monitoring to local needs.

So long as the uncertainty in sampling is well below the uncertainty in other aspects of the regulatory process, then reasonable goals for "good measurement practice" are sufficient. The DQO software tool approach is a good start in linking monitoring specifications to their regulatory consequences. However, the current criteria for $PM_{2.5}$ FEM designation are significantly more straightforward and much easier to generate and duplicate. Once FRM and candidate data sets are validated and paired into a software program such as Microsoft EXCEL, graphs can be generated quickly, with linear regression results of slope (m), intercept (b) and correlation coefficient (r). These linear regression values are universally understood by most data users.

The proposed criteria call for a range of candidate samplers. This should be a fixed number for all FEM testing. Further, the proposed criteria call for averaging candidate sampler data. This would have the tendency to smooth data, and average out noisy candidate samplers. This criteria does not directly reflect real world applications where most monitoring stations have only one sampler or monitor and data is not averaged.

The minimum daily FRM sample time should remain at 24 hours +/- 1 hour (23 to 25 hours), adhering to current CFR criteria. There is no need to shorten this criterion.

5. What are the advantages and disadvantages of using sampler precision and sample population to help determine the minimum correlation requirement for the approval of $PM_{2.5}$ continuous methods?

The measurement community has long recognized that local conditions can sometimes cause otherwise "acceptable" methods to behave poorly. Specifications based on assumptions about the composition of the atmosphere should be, where practical, replaced by specifications based on actual local conditions, or at least, interpretation of QA data should be sensitive to local variation. See response to Question 4 (above). In this context, the analysis presented should be expanded to show how real data from a range of sites compare to the synthetic data presented in the September 30, 2004 draft.

Moreover, when treating both real and synthetic data, they should be recognized as composed of continuous size distributions (i.e., the tail of the "coarse" overlaps the "fine", and vice-versa). This would permit evaluation of the meaning of discrepancies among samplers, rather than flat specification of the FRM as "right" and the candidate methods being "over" or "under".

If a "loosening" of FEM criteria is desired, then it would be more prudent to simply alter the slope, intercept, and/or correlation coefficient. In the proposed criteria, target values for the correlation coefficient and additive bias remain unknown until all calculations are completed. Maintaining current linear regression calculations is significantly more beneficial when presenting comparison data or meeting government regulations.

With the proposed FEM calculations, the mixing of "population" and "sample" based equations limits the use to manual or macro type calculations, significantly increasing the effort and time involved in generating the proposed results. The proposed FEM method increases the possibility of error while making the final results more difficult to understand. It appears that the proposed FEM criteria is meant to skew the comparison of candidate samplers to FRM samplers so that specific biases may pass while opposite but equal biases may fail.

6. What are the Subcommittee's views on using a $PM_{2.5}$ continuous monitor approved as a FEM, being applicable for use as part of a potential $PM_{2.5}$ secondary standard for visibility?

Continuous $PM_{2.5}$ monitoring is suggested as a reasonable surrogate for anthropogenic visibility degradation. Since $PM_{2.5}$ is generally correlated with reduced visibility, the appeal is obvious. However, the problem with this approach is that there is no accompanying proposed mechanism for making use of short-term $PM_{2.5}$ data as a basis for regulating. The literature of human

perception of haze and "smog" suggests that responses can be highly idiosyncratic. Given that there is no policy guidance on how frequent or persistent low visibility needs to be to have a "deleterious effect" on public welfare, it is unclear how specifying measurements now is useful to U.S. EPA's regulatory process. It may be that visibility protection could be achieved based solely on a program using current and expected health-based PM monitoring data.

7. To what extent have the assessments of spatial variability and the sensitivity of the DQO process to a variety of population distributions been appropriately addressed?

Spatial and temporal variation and autocorrelation need to be treated more realistically. The reality of sampler exposure in urban settings can include strong diurnal cycles that mix particles from one source (e.g. road dust – with very high spatial and temporal variation) with those from entirely different ones (e.g., aged secondary aerosol – a more regionally distributed pollutant) within a time-integrated sample. The spatial and temporal structure of ambient concentrations, size distributions, volatility, hygroscopicity, and other factors can all influence sampler performance and how well sample values represent the ambient environment's real variability. The analysis presented is a good start, but only a start: the next step is to add continuous size distributions and to populate this analysis with data based on a wide range of realistic exposure cases. See response to Question 5 (above).

The results from special studies could be analyzed to determine the spatial distributions of $PM_{10-2.5}$ and provide guidance to the number of monitors that need to be sited in potential $PM_{10-2.5}$ non-attainment areas to properly represent population exposure. The California Air Resources Board and perhaps other SLT agencies have conducted such special studies. One example is a PM saturation study conducted by DRI with mini-vols during 2000 in Corcoran, an agricultural community in the San Joaquin Valley with high dust levels. Similar studies may have been conducted in Las Vegas and Phoenix.

8. What are the Subcommittee's views on the approach identified for the development of criteria to approve continuous $PM_{10-2.5}$ equivalent methods?

The approach laid out in the August 19, 2005 draft is not adequately sensitive to "real-world" variables. Specifically, the requirement for seasonal comparison should apply to a range of sites, just as the "given season" is applied across a range of sites. Think of this as equalizing the sampling matrix, with consequent potential for better statistical characterization not only of inter-sampler agreement, but site- and season- linked variation as well.

(A few examples: winter in Fresno challenges samplers with high nitrate and high humidity; summer in Riverside presents high nitrate and a high potential for sampling artifacts; winter in St. Paul tests sampler resistance to freezing and internal ice formation; summer in Atlanta tests for possible effects of ambient nitric acid and unsaturated aerosol sulfate ion).

References

Motallebi, N., C. A. Taylor, Jr., K. Turkiewicz, and B. E. Croes (2003a) Particulate matter in California: Part 1 – Intercomparison of several $PM_{2.5}$, PM_{10-25}, and PM_{10} monitoring networks. J. Air Waste Manage. Assoc., **53:** 1509-1516.

Motallebi, N., C. A. Taylor, Jr., and B. E. Croes (2003b) Particulate matter in California: Part 2 – Spatial, temporal, and compositional patterns of $PM_{2.5}$, PM_{10-25}, and PM_{10}. J. Air Waste Manage. Assoc., **53:** 1517-1530.

Dr. Kenneth Demerjian

Final Comments by Demerjian. K.L.
September 25, 2005

U.S. EPA Clean Air Scientific Advisory Committee (CASAC)
CASAC Ambient Air Monitoring & Methods (AAMM) Subcommittee
Meeting September 21, 2005 – September 22, 2005
Peer Review of PM10-2.5 Federal Reference Method (FRM); and Consultation on Field Evaluation of PM10-2.5 Methods, Optimization of the PM2.5 FRM, Equivalency Criteria for PM2.5 Continuous Methods, Monitoring Data Quality Objectives for PM10-2.5, and Equivalency Criteria for PM10-2.5 Continuous Methods

Response to charge Questions:

Attachment 1 – Selection and technical summary of $PM_{10-2.5}$ FRM:

What are the scientific and operational strengths and weaknesses of the $PM_{10-2.5}$ difference method relative to other options for a proposed FRM, especially when used as the basis for approval of other methods?

The technique does not address volatile losses and may introduce a systematic bias in mass measurements which are likely source, composition and temperature sensitive. The health consequences of PM exposures are likely size and chemical composition related. There is evidence that FRM methods are subject to seasonal bias due to losses of volatile species. It is also likely that such losses will have diurnal characteristics, emphasizing the importance of higher time resolved measurements.

Based on the field study report as well as any other available data, *e.g.*, data from State and local agencies, how does the demonstrated data quality of the $PM_{10-2.5}$ difference method support or detract from it being proposed as a FRM?

The precision reported in these studies certainly supports EPA's rationale and desired objective to establish $PM_{10-2.5}$ FRM that has a "high degree of fidelity and faithfulness" but unfortunately it may not be the correct realization of PM mass in the atmosphere and therefore may not be representative of inhalation exposures of greatest potential harm. Therefore in developing and defining the FRM, EPA must acknowledge the likelihood of volatile losses biasing the FRM measurement and provide the latitude for embracing emerging instrumentation technologies that are proving to provide more accurate measurement of ambient PM mass.

Attachment 2 – EPA's Multi-Site Evaluations of Candidate Methodologies for Determining Coarse Particulate Matter ($PM_{10-2.5}$) Concentrations: August 2005 Updated Report Regarding Second-Generation and New $PM_{10-2.5}$ Samplers

Based upon the latest available field study data, which $PM_{10-2.5}$ methods have both sufficient utility to meet one or more important monitoring objectives and appropriate data quality to be

considered for deployment as Federal Equivalent Methods (FEMs) or speciation samplers in a potential $PM_{10-2.5}$ monitoring network?

The subject work is a more detailed look at stage one evaluations of PM course methods (2003 – 2004) as presented to the CASAC AAMM in July 22, 2004 and follow-up work on second generation samplers performed in Phoenix, AZ (May & April 2005). As much as I find these results interesting, they seem not to address fundamental questions about $PM_{10-2.5}$ mass measurement and the uncertainties associated with mass losses as a function of chemical composition, source mix, and season. This issue was raised at the July 22, 2004 meeting and seems to have been ignored. The fact is if the $PM_{2.5}$, PM_{10} and $PM_{10-2.5}$ FRMs as mandated may not be the correct realization of PM mass in the atmosphere and therefore may not be representative of inhalation exposures of greatest potential harm. At the July 22, 2004 meeting in discussing the first draft, the suggestion was made that the chemical content of the samples should be studied for the various sites and the potential for volatile losses assessed for the respective sites. The current report provides no information on the chemical composition of PM at these sites. Why not?

I understand EPA's rationale for wanting to establish $PM_{10-2.5}$ FRM (attachment I) because of its "high degree of fidelity and faithfulness", but I am concerned that its regulatory mandate has put blinders on its monitoring strategy. The health consequences of PM exposures are likely size and chemical composition related. There is evidence that FRM methods are subject to bias due to losses of volatile species. It is also likely that such losses will have diurnal characteristics, emphasizing the importance of higher time resolved measurements. Some of the second generation techniques tested may more effectively capture PM masses as compared to the FRMs. I don't doubt that differences identified in the subject study with regard to size cutoff issues for some second generation techniques evaluated are important and real, but the study has discounted volatile losses and made no attempt to test it presence/contribution to the observed differences experienced. Progress in determining the health effects of PM (and likely important chemical constituents) will remain hampered if our only main monitoring resource is 24-hr FRM mass measurements.

Based on the performance evaluations reported in the subject document and assuming the inlet and engineering issues identified are effectively resolved and robust operational requirements met, I would considered the R&P Model 2025 sequential dichotomous sampler, the R&P continuous Coarse TEOM monitor, the R&P single event dichotomous sampler, the Sierra-Anderson 241 dichotomous sampler, and the R&P dichotomous TEOM sampler for deployment as Federal Equivalent Methods (FEMs). Further testing and resolution of issues associated with the Kimoto SPM-613D and the BGI Omni samplers are required prior to their consideration. Given their dependencies of aerosol density and shape factors, it is not clear that the TSI Model 3321 APS or the GRIMM ET Model 1.107 can achieve FEM status. Further analysis of the Kimoto SPM-613D and GRIMM ET Model 1.107 - FRM comparisons should be considered in conjunction with available chemical speciation data. Both methods suggest a high bias that may reflect FRM species volatility issues.

Additional comments on next steps

Although one can provide some rationale for the sites selected in the preliminary evaluations, the proposed follow-on study in Birmingham, AL does not seem justified. Why not consider deployments in major cities like Houston or New York where PM population exposures are high and the sites represent diverse source mixes and climate differences that might well accentuate the FRM bias issues raised concerning PM volatility.

Attachment 3 – Memo to PM NAAQS Review Docket (OAR-2001-0017) – Potential changes being evaluated for the $PM_{2.5}$ Federal Reference Method

What are the Subcommittee's views on the Very Sharp Cut Cyclone (VSCC) being approved as an alternative second-stage impactor to the Well Impactor Ninety-Six (WINS) for use on a $PM_{2.5}$ FRM?

Technically this is an acceptable alternative, but one does wonder if the minimal testing that has occurred under these studies has identified all sources of error that might result in more representative multi-site, year around operational testing.

To what extent are the stated advantages of relaxing existing requirements identified for the $PM_{2.5}$ FRM supported by the information cited in Attachment 3, available literature, or good field and laboratory practices? Does the Subcommittee have additional recommendations for the $PM_{2.5}$ FRM that would be neutral with respect to bias, but would improve the performance and minimize the burden on agencies conducting the sampling?

The three suggested changes in FRM procedures (in addition to the use of the Very Sharp Cut Cyclone), i.e. 1) the use of alternative DOS WINS oil, 2) extension in filter recovery time to 177 hrs, and 3) maintenance of filter transport temperature and post sampling recovery temperature, are acceptable changes that should not effect the precision and accuracy of the PM2.5 FRM and will facilitate field operations, improving efficiencies in technician support.

Attachment 4 – Criteria for Designation of Equivalence Methods for Continuous Surveillance of $PM_{2.5}$ Ambient Air Quality

Considering the statistical measures of precision, correlation, multiplicative bias, and additive bias identified for approval of $PM_{2.5}$ continuous methods, what are the Subcommittee's views on the usefulness of each measure to ensure that approved or equivalent methods meet the monitoring network data quality objectives?

Each statistical measure contributes to basic understanding of the performance of equivalent method under study. It should be noted that observed bias in comparisons may not be an indictment of the equivalent method, but that of the FRM. Given the expected bias due to FRM volatile losses, the statistical measure with respect to additive bias should be adjusted accordingly.

What are the advantages and disadvantages of using sampler precision and sample population to help determine the minimum correlation requirement for the approval of $PM_{2.5}$ continuous methods? *Sampler precision and sample populations are one element in developing approval of*

PM2.5 continuous methods. Establishment of absolute accuracy of the FRM method must be the ultimate goal and if the techniques to address the additive bias as discussed in the question above are accomplished, these techniques will prove a useful tool in evaluating PM2.5 continuous methods.

What are the Subcommittee's views on using a $PM_{2.5}$ continuous monitor approved as a FEM, being applicable for use as part of a potential $PM_{2.5}$ secondary standard for visibility?
PM2.5 continuous mass monitors can provide viable data for assessing compliance of a secondary visibility standard if the water content of PM can be adequately accounted for. The R&P FDMS $PM_{2.5}$ continuous mass monitor shows great promise in the measurement of ambient PM with its associated water.

Attachment 5 – Sensitivity of the $PM_{10-2.5}$ Data Quality Objectives to Spatially Related Uncertainties

To what extent have the assessments of spatial variability and the sensitivity of the DQO process to a variety of population distributions been appropriately addressed?

The techniques applied to address the effects of spatial variability and multi-modal distributions on PM10-2.5 DQO process are reasonable. The assessment findings indicated that for the daily standard, the performance curves were most sensitive to sampling frequency, followed by data completeness and population. The effect of multi-modal distributions was observed to be very small as was the effect of the spatial variability. The question remains as to whether or not PM volatility and spatial gradients issues have been adequately address in the subject sensitivity issues.

Question associated with Attachment 6 – $PM_{10-2.5}$ Method Equivalency Development

What are the Subcommittee's views on the approach identified for the development of criteria to approve continuous $PM_{10-2.5}$ equivalent methods?

The proposed approach is reasonable.

Dr. Delbert Eatough

Comments on Summary and Rationale for the $PM_{10-2.5}$ FRM

Delbert J. Eatough
15 September 2005

A. Overall.

The rational for the choice of the proposed $PM_{10-2.5}$ FRM method is well laid out. Assumptions which are made, and a summary of what the proposed method will and will not measure are given with reasonable thought. A few comments on assumptions which may not be valid are given in the next sections.

B. Semi-volatile Components.

First, the summary does recognize that semi-volatile species will be lost from the collected particles with the proposed FRM. The recognition that semi-volatile species will not be properly sampled is a step forward in the science. The assumption is made that semi-volatile losses will be the same for both the PM_{10} and $PM_{2.5}$ portions of the FRM. The assumption is also made that losses will be greater with the proposed FRM than with an alternate virtual impactor sampler because of the higher flow rate of the FRM sampler. There is no data that I am aware of that would validate or invalidate the first statement. It is presently untested. There are data which indicate that the second assumption is not correct.

With respect to the losses of semi-volatile material from a PM_{10} as compared to a $PM_{2.5}$ sampler, this assumption has not been directly tested. The great body of data which exist on sampling for semi-volatile species are limited to the study of the sampling of fine particles. There is a limited body of older data which suggest that the loss of ammonium nitrate will be substantially less for the PM_{10} sampler than for the $PM_{2.5}$ sampler in some environments. Hence, for these environments $PM_{10-2.5}$ would be overestimated with the FRM. This phenomenon is apparently associated, at least in part, with the reduction in the loss of ammonium nitrate in fine aerosols due to the neutralization of some acidic species during the collection of PM_{10}, $PM_{10-2.5}$ usually being less acidic than $PM_{2.5}$ in an urban environment. The reduction in acidity reduces the loss of ammonium nitrate due to reactions such as ;

$$NH_4NO_3(s) + NH_4HSO_4(s) \rightarrow (NH_4)_2SO_4(s) + HNO_3(g),$$

as fine particles of different compositions are mixed on the filter during the sampling process. This mixing is exacerbated by 24-h sampling, with the possible changes in aerosol composition over the source of a day. Such chemistry has been suggested, for example, in studies in the Los Angeles Basin. Similar effects could be observed for semi-volatile organic material where the volatility of a given compound in fine particles is altered by mixing with absorptive coarse particle material during sampling and volatilization is reduced. While this is a very feasible mechanism for reduction in the loss of semi-volatile organic material on a PM_{10}, compared to a $PM_{2.5}$ sample, I

am not aware of any data examining this question.

The assumption is also made that the loss of semi-volatile organic material and ammonium nitrate from a $PM_{2.5}$ collecting filter sampled at a flow rate of 17 L/min (the FRM) will be higher than for a flow rate of 2 L/min (a virtual impactor) for the same air mass and collection filter. This will probably not be the case based on published data. The question is one of relative kinetics and whether or not the losses occurring over a 24-h period are a function of flow rate. There is some data indicating that losses are a function of flow rate for relatively short sample collection time periods (hour or less). But we have conducted many studies comparing undenuded and denuded quartz and Teflon filters over time periods of about an hour up to 24-h. In all cases, the data indicate losses are about the same. This is even a more stressful test than the comparison of different sampling rates since the gas phase species are significantly reduced in concentration. It appears that the loss of semi-volatile organic material is fast enough that for multi-hour sampling periods, the end result is not dependent on the flow rate (or vapor pressure of the gas phase compound), e.g, the kinetics of loss are much faster than the sampling time period. This results is also consistent with rates of loss of organic compounds from denuded particles measured in chamber experiments by Kamens. If this is the case, semi-volatile losses will not be sensitive to the filter flow rates for an FRM or FEM.

Incidently, both these observations would tend to reduce the incidence of negative reading in a difference calculation of mass for either a FRM or virtual impactor based FEM system.

C. Loss of Coarse Particles During Sampling.

The assumption is made that use of a difference method is preferable to use of a virtual impactor because of the reduced loss of coarse particles in a PM_{10}, as compared to a $PM_{10-2.5}$ collected sample. I concur with the arguments associated with that assumption. Incidently, the effects which reduce particle loss are related to the effects which probably reduce semi-volatile material loss in the same comparison as discussed in the proceeding section.

D. Automated Semi-Continuous Methods.

Using a continuous sampler as an FEM sampler is fraught with many problem related to equivalency of the sampling procedures. Included in these problems are the question of semi-volatile loss in the FEM method. While the sampling time is not a factor in semi-volatile loss for the FRM, it can be for the FEM, including increased losses due to heating of the sample in the continuous sampler (e.g. a TEOM), or reduced volatile losses during sample collection and analysis (e.g. a GRIMM or FDMS TEOM monitor). The very nature of the FRM sampler will make developing a true continuous FEM sampler difficult, if not impossible. In addition, forcing a continuous method to mimic the FRM under some conditions will not guarantee it will mimic those conditions under all conditions. And finally, such forcing moves us away from the scientific advantage that can be gained from correctly measuring the semi-volatile particulate matter in the more advanced semi-continuous methods. I have additional comments on this point in my review

of the Multi-Site Evaluations manuscript. Ultimately, EPA needs to come to gripes with a balance of regulatory equivalence forced on us by the chosen FRM and the scientific advances which can be made with the more accurate sampling of fine particulate (and possible coarse particulate) material with more advanced semi-continuous methods. They have made a start in this direction in the philosophy outlined in the future national sampling network design which we have previously reviewed. EPA needs to be certain this vital scientific question is not completely sacrificed to the choice of an FRM sampler.

Delbert J. Eatough
Professor of Chemistry
Brigham Young University

Comments on Multi-Site Evaluations of Candidate Methodologies - August 2005 Undated Report

Delbert J. Eatough
15 September 2005

A. Overall.

This report includes both the original results from the 2003 - 2004 studies at Gary, Riverside and Phoenix (2 studies) plus the results obtained this year in Phoenix with Second-Generation samplers modified based on the 2003 - 2004 results. Samplers evaluated include the proposed FRM (a PM_{10} and $PM_{2.5}$ difference calculation), an R&P Dichot Sampler, a Beta Gauge, the APS, a BGI Saturation Sampler, the GRIMM Monitor and a Dichot FDMS TEOM sampler.

In addition to the above, I also requested and obtained from Dr. Vanderpool the results for $PM_{2.5}$ measured with an R&P FDMS TEOM in Gary and Riverside and the $PM_{2.5}$ data measured with a Beta Gauge in the Riverside study. My observations on the comparisons of the various samplers with the FRM results follows.

B. Specific Comments on the Results.

1. FRM Results.

The results of the study clearly indicate that all PM_{10} and $PM_{2.5}$ samplers by the three different manufacturers are equivalent and that the precision among the various instruments is all such that precise $PM_{10-2.5}$ values can be calculated from the collocated results. Based on the results obtained in the 5 different studies, the decision to use the two FRM samplers to obtain data with reliable precision for comparison of results with other potential FEM samplers seems well justified.

2. FRM Compared to Dichot Sampler Results.

The data in this report clarifying that the PM_{10} on the Dichots should be identical to that for the PM_{10} FRM are helpful, compared to the first report. Please verify that is indeed the case and I am not reading anything into the wording. This being the case, differences between the two samplers must be attributed to difference after the inlet.

It seems to be well established from the data presented that the sequential R&P Dichot has a problem with retention of coarse particles in the $PM_{10-2.5}$ channel. The fact that the problem is much more apparent in the Arizona data than in the studies at the other two sites is quite reasonable. Based on the site descriptions, the coarse particles sampled at the Gary and Riverside sites will include substantial organic material. In contrast, the Phoenix coarse particles will be typified by crustal material, dominated by dust from construction projects. It would be expected that the Phoenix coarse material will be more subject to particle bounce. This problem seems to

be well controlled in the modified single event R&P dichot sampler, based on the 2005 data, but still not solved in the sequential sampler. Of course, the improvement in results for the single event Dichot is obtained with some sacrifice in sampler operation simplicity.

However, the data also point to another difference between the Dichot and the FRM samplers. In the Phoenix data there is a consistent difference in the $PM_{2.5}$ results obtained with the two samplers, with the Dichot giving the higher results. Both for the 2003-2004 initial studies and for the 2005 studies with the single event and sequential Dichots. This effect is not seen in the data at Gary and Riverside. It seems to me that this is suggestive of a difference in particle size distribution near the 2.5 cut point. It is expected that the shape of the curve around the 2.5 cut point is not the same for the two samplers. It can also be expected that the higher concentrations of crustal material at Phoenix may well lead to a larger tail of coarse material in the aerosol sampled at the Phoenix site being present in the $PM_{2.5}$ sample. At the time of the last review of the initial study it was suggested that information on this point could be obtained using the APS data. After all, these data will give information on the particle size distribution around the 2.5 cut point for the aerosol present at the different sites. That has still not been done. This is an important potential factor in explaining the data which must be explored before the precision between the two samplers can really be established. I urge EPA to inform us on the results of that analysis.

This suggested analysis (APS exploration of reasons for the difference in the FRM and Dichot) will be important because if the difference is due to the difference in the curve for the 2.5 WINS and the virtual impactor, EPA must make a decision as to whether this difference is acceptable, or whether the virtual impactor, as presently configured, is not suitable for a FEM.

3. FRM Compared to Coarse TEOM Results:

There are three important ways in which the Coarse TEOM differs operationally from the $PM_{10-2.5}$ FRM measurement:

1. The inlets are different. The inlet on the TEOM is a modified PM_{10} inlet designed to provide a comparable cut to the 16.7 lpm with a flow of 50 lpm. Design on this inlet was changed between the 2003-2004 and the 2005 tests. Furthermore, the inlet was assumed to have 9% particle losses in the 2005 study, but not in the earlier studies. No validation tests have been performed on this assumed loss. Nor is the loss as a function of the coarse particle size distribution known.

2. The $PM_{2.5}$ cuts are different. In the FRM the 2.5 cut is made with a WINS. A virtual impactor is used with the TEOM. Furthermore, the design (and main to major flow ratios) of the virtual impactor on the TEOM is different from that used in the Dichot. The data suggest that in Phoenix, sample is preferentially directed to the fine mode in the Dichot, as compared to the FRM cut. No information is available on this point for the TEOM virtual impactor.

3. The collection filter is heated in the TEOM sampler, but not for the FRM samplers. Loss of semi-volatile material, to the extent it is present in coarse particles, would be expected. The temperature was changed for the Phoenix 2005 study from 50°C to 40°. However, losses would still be expected. For example, it has been shown that the reduction of the filter temperature to 30°C in the $PM_{2.5}$ SES TEOM reduces slightly, but does not eliminate semi-volatile loss.

The challenge in the interpretation of the Coarse TEOM data is to try and deduce the relative effect of these three operational differences.

The slope of FRM vs Coarse TEOM $PM_{10-2.5}$ data for the various studies indicates that the Coarse TEOM gave a concentration an average of about 25% lower than the FRM. The slope of the Phoenix 2003 comparison was similar to that of the other two initial studies, but the intercept was substantial, 13 $\mu g/m^3$. It would certainly be informative if the data for these studies were made available in the report. Is there any explanation for the apparent significant bias in the Phoenix 2003, but not the other data sets. Examination of the APS data indicated that the cut point of the Coarse TEOM initial inlet accounted for about 10% lower values. But no indication is given as to what the other 15% was due to. Is it the effect of the shape for the curve for the lower cut point (see for example the FRM vs Dichot data)? Is it the effect of loss of semi-volatiles? Neither of these problems will be corrected with the changes in design of the initial inlet.

APS data with the modified instrument used in the Phoenix 2005 study indicate the new cut point in the modified instrument is 10, rather than 9, µm. However, an equally important question is, what does the shape of the curve at the cut point look like compared to the FRM curve? The Phoenix data give a regression slope of 1.09 for the 2005 Coarse TEOM vs FRM $PM_{10-2.5}$ results. This slope is conveniently played down in the discussion. However, looking only at ratios does not give a solid statistical look at the data. The ratio as a function of concentrations is equally important. It is interesting that the higher slope matches the assume loss correction. Clearly data on the actual losses as a function of environment are needed if the instrument is to be designated FEM. The higher concentrations seen for the Coarse TEOM in the 2005 Phoenix study must be due to one of four things: 1. Differences in the upper cut point for the two instruments being compared, 2. Differences in the lower cut point for the samples collected, 3. The absence of semi-volatile material for the coarse TEOM in the 2005 study, 4. An incorrect assumption on losses in the system. Examination of the APS data may shed some light into the first two possibilities. Clearly, the Coarse TEOM is a promising instrument, but is far from being ready for FEM designation.

A question not yet tackled by EPA is, at the end of the day if the Coarse TEOM is shown to have semi-volatile losses different from the FRM, but other than that is OK, what will EPA do? A converse question coming up is what if the FDMS Coarse TEOM is shown to measure higher that the FRM due to semi-volatile losses, what will EPA do about FEM designation. This, of course is a box we are in for consideration of what to do for $PM_{2.5}$ FRM and FEM samplers as the

EPA considers the move to semi-continuous monitors.

4. FRM Compared to Beta Gauge Results.

The FRM and Beta Gauge $PM_{2.5}$ data do not agree well, with the Beta Gauge data being consistently higher than the FRM data for all studies. The EPA report attributes this to the presence of coarse particles in the fine particle mode as a result of the characteristics of the virtual impactor used in the Beta Gauge, similar to the problem noted above for the Dichot sampler. However, if this were the case, one would expect PM_{10} measurements with the two instruments to be comparable and the $PM_{10-2.5}$ concentrations to be correspondingly lower. This however, is not the case. Instead, the $PM_{10-2.5}$ results are in reasonable agreement (if anything, the Beta Gauge data are still biased a little high) and the PM_{10} results are biased high. This bias is evident in the slopes near unity and a significant positive intercept for the Gary, and Phoenix 2003, 2004 and 2005 data. Although statistic are not given in a Table in the report for the 2005 data, the results given in Figures 22-24, indicate the results are comparable to the 2003 and 2004 studies.

An interesting anomaly is the Riverside data, where the PM_{10} correlation shows a slope much greater that unity and a negative intercept. Similar results are also seen for the fine and coarse correlations. Unfortunately, the data themselves are not given in the report. But Dr. Vanderpool has provided me with the $PM_{2.5}$ Beta Gauge results and with the $PM_{2.5}$ FDMS TEOM results for Riverside. Both data sets are biased higher than the FRM. The results are also all highly correlated. The Beta gauge results average 65% higher than the FRM. The FDMS results average 31% higher. It is possible the Beta Gauge results are higher than both the FRM and the FDMS TEOM in part because of the shape of the 2.5 cut point for the virtual impactor. However, all of these results are entirely consistent with that which would be expected if the sampled aerosols contained measurable concentrations of semi-volatile species.

If it is assumed that the differences between the two measurements is due to aerosol semi-volatile species, then the conclusion would be reached that the semi-volatile species are highest in the summer studies in both Phoenix and Riverside, lower in the winter study in Phoenix and lowest in the early spring study (with much cooler temperatures) in Gary. This is exactly the trend expected for the formation of secondary semi-volatile fine particulate material at the various sites. In this connection, the $PM_{2.5}$ FDMS TEOM results (provided to me by Dr, Vanderpool) agree with the $PM_{2.5}$ FRM data, consistent with semi-volatile fine particulate material not being important in the Gary study.

5. FRM Compared to APS Results.

The key to comparison of the results obtained using this two different samplers is related to confidence in the uniform density and shape factors used to convert the APS particle count data to mass. The agreement given in the figures looks in general good. However, I do not think I can judge if the treatment is really reasonable and unbiased.

6. FRM Compared to BGI Saturation Sampler Results

I consider it most likely that the observed differences between these two samplers is a reflection of the lower precision of the BGI because of the lower flow rates and lower measured mass, coupled with a difference in the inlet curves. Are details on the shape of the curve for the BGI low flow unit available which could indicate if this is true? Given these two considerations, the comparison is not unreasonable.

7. FRM Compared to GRIMM Monitor.

The GRIMM data provide the second sampler comparison where the possibility exists for the retention of semi-volatile species in the comparison sampler measurement. The first such possibility was seen in the Beta Gauge results. While number are not given which will allow a direct comparison of the Beta gauge and GRIMM sampler results in the 2005 Phoenix study, comparison of the data in Figures 22 and 29 suggests that reasonable agreement was seen between these two samples, with both being biased higher than the FRM data by about 35 to 40%.

We have published two reports of measurement of fine particulate material with the GRIMM sampler, one being a summer study and Riverside and the other a winter study at Fresno. These studies included comparison with Beta gauge measurements (at Fresno) and with several independent measurements of semi-volatile fine particulate matter at both the sites. In both studies, the GRIMM was biased high than the FRM measurements and the difference was directly attributable to the presence of semi-volatile fine particulate species not measured by the FRM.

8. FRM Compared to the R&P FDMs Dichot TEOM Sampler.

The results for this comparison are not at all expected based on other data reported to the present with FDMS TEOM samplers. Conversations with EPA and with R&P have lead me to believe the R&P instruments in this comparison were rushed to the field before they were working correctly and the data in this section should be completely discounted.

C. Comment on the Possible Implications for Continuous Monitors.

A key focus of the NAAMS is a shift from the use of integrated monitors to continuous monitors at the Level II sampling sites to be included in the NCORE Program of the NAAMS. The NAAMS document correctly points out in several sections the value of such data in providing input for use by the scientific community in the understanding of peak exposures, atmospheric processes and diurnal variations in the atmosphere. These data will all be potentially valuable to the both the SLTs and the scientific community which will also use the data. Improving public access to and developing initial interpretation of this data is a key part of the Implementation Plan. A strong push for moving in this direction is the economic advantage which can be obtained using continuous monitors, coupled with the increased understanding of atmospheric processes using

the data, a potential win-win situation.

As pointed out in the NAAMS, there is also a clash between the monitoring needs of the NAAMS as outlined in the current regulations which define the fine particulate FRM as the basis of monitoring for attainment and the problems in the FRM which are addressed, but not carefully considered in the NAAMS. What will and will not be obtained using the new suite of continuous PM monitors is not considered in the NAAMS, however, the implication is there that the expected advances which might accrue from the availability of continuous data will accrue. The Continuous Monitoring Implementation Plan, on the other hand, outlines carefully the protocols which will be followed as the NAAMS Level II program is implemented. I have a great concern that the philosophy given in the Continuous Monitoring Implementation plan and the objectives for continuous monitors as outlined in the NAAMS cannot both be achieved.

The Continuous Monitoring Implementation Plan compared FRM and TEOM results across the country, and noted areas of agreement and disagreement and suggests that aerosol composition is responsible for the disagreements seen. The plan then proceeds to outline protocols intended to assure that the continuous monitors to be implemented will be consistent with data which would have been obtained with an FRM sampler. Substantial research has been reported since the Version 2 Draft of the Continuous Monitoring Implementation Plan which sheds additional light on implications of this approach.

EPA is probably correct in identifying the presence of semi-volatile material in fine particulate matter as being related to and responsible for the agreement or lack of agreement between an FRM and a TEOM. We now understand that both nitrate and organic material can contribute substantially to the SVM. Studies which have obtained FRM and TEOM data, as well as correctly measuring both nitrate and organic SVM have shown that agreement between the FRM and TEOM monitors will occur only when either there is no SVM, or when losses of SVM is comparable for the two monitors. For example, in summer there is a tendency for both samplers to loss both nitrate and semi-volatile organic material and for there to be agreement between the two in 24-h average data. However, in winter, the FRM is often higher than the TEOM because of better retention of SVM. Moreover, on a 24-h comparison basis, there is usually good correlation in the two data sets over a given season or meteorological condition. These correlations tend to exist between the FRM and TEOM on a 24-h average even when the slope of such a comparison is different from unity. These effects probably account for some of the observed seasonal variations and generally reasonable regression comparisons in the EPA Continuous Monitoring Implementation Plan report. With such results, there is a temptation to "correct" the TEOM data so that agreement between the two systems is generally seen. While this is, potentially, an acceptable solution for monitoring purposes, serious problems are introduced with respect to the uses of continuous data as proposed by EPA in the NAAMS.

Even when there is reasonable correlation in 24-h data, comparison of 1-h average TEOM and more state-of-the art instruments (such as the FDMS TEOM) show quite different diurnal patterns. This is because the diurnal changes in the chemistry of the atmosphere which lead to

SVM not well measured by the TEOM (and often by the FRM) are averaged out on a day-to-day 24-h comparison. The events which occur leading to significant atmospheric chemistry (and potential health risk and maximum exposure conditions) occur on a frequency which can be seen in 1-h data but not the 24-h data. These events are usually missed by a TEOM. Thus, using modified TEOM data will lead to the worst possible situation, believing we have valid data on diurnal patterns because of agreement with 24-h FRM data, but completely missing the diurnal features which will to aid in the understanding of atmospheric processes of importance to exposure and possible risk. However, using an instrument which will measure SVM (and hence catch these significant events) will produce data which do not meet the agreement protocols (particularly the ±10 % slope agreement) given in the Continuous Monitoring Implementation Plan, precisely because the atmosphere is better monitored.

EPA may well be constrained to not use these newer techniques because they will not agree with the FRM and thus fail monitoring legal requirements, especially in locations where SVM is important and variable. However, a consequence is that we will be producing data which give us an inaccurate picture of the atmosphere and thus lead to incorrect decisions based on continuous monitoring data. EPA needs to find a way around this problem which can be applied to the various Level II monitoring sites, allowing for the advances in understanding which form one of the major arguments for moving toward continuous monitors. At a minimum, some (if not many) of the Level II sites should have both a FRM equivalent continuous monitor and a state-of-the art continuous monitor which allows us to better understand atmospheric processes. The comparison between the two will give valuable insights on SVM in the atmosphere and aid the health community in obtaining a better understanding of exposure.

Delbert J. Eatough
Professor of Chemistry
Brigham Young University

Comments on the 21-22 CASAC AAMM Review and Comment Meeting

Delbert J. Eatough
29 September 2005

This will provide my observations on what I consider to be the more important points touched on at the Durham meeting. I have previously provided details on my thoughts on:

1. Comments on Summary and Rationale for the $PM_{10-2.5}$ FRM.
2. Comments on Multi-Site Evaluations of Candidate Methodologies - August 2005 Undated Report.

Which were provided before the meeting. I believe the points made therein are unchanged by the input obtained at the meeting and will not repeat those thoughts here. There are a few items which seem to me to be key points discussed at the meeting which EPA should consider as the FRM and associated regulations are put in place for both the $PM_{2.5}$ and anticipated $PM_{10-2.5}$ standards.

1. Consideration of the $PM_{2.5}$ Cut Point in the Proposed FRM and in Candidate FEM Samplers Which Measure Both $PM_{2.5}$ and $PM_{10-2.5}$.

As pointed out in my previous remarks, on the results of the EPA studies covered in Attachment 2, there is a consistent tendency for potential FEM samplers which use an FRM PM_{10} standard inlet and then a virtual impactor to make the cut between the coarse and fine measurements to agree with the proposed FRM on PM_{10}, but to have a bias between the too fractions with the fine being higher and the coarse lower than results obtained with the FRM. The EPA description of the data talks about the loss of coarse into the fine channel. However, looking at the total data it seems to me the difference may be due solely to the shape of the cut point curve for the two $PM_{2.5}$ separations. Thus for example, this tendency in seen in the Phoenix, but not the Gary data. EPA can evaluate this possibility using the APS and/or GRIMM size data. I have encourage that evaluation in my earlier writeup.

However, there is another point which may be equally important. The argument for using the proposed difference FRM samplers is the tie of filter pack data to the past health data. However, the very sharp cut point inlet proposed for the new $PM_{2.5}$ FRM is not the same as that used in the past health studies. In fact, the virtual impactor inlets used by potential FEM samplers which make both $PM_{2.5}$ and $PM_{10-2.5}$ measurements in a single instrument is closer to that which is used. Thus the only reason for using the proposed difference measurement with the fine FRM impactor or a very sharp cut cyclone is consistency with the current FRM, not tie to the health data. I see no reason to tie the hands of potential FEM candidates which can indepenently measure both fractions based on a rather arbitrary choice of a $PM_{2.5}$ cut, tied to what we think is good science (a sharp cut point curve), but not particularly to the health data. The health science can certainly allow for either choice. Perhaps EPA should also allow for this latitude of choice in the requirement for FEM instruments.

2. Consideration of an FRM or FEM Which Allows for Direct Chemical Speciation.

As was discussed frequently during the meeting, the proposed difference FRM method does not allow for the collection of a separate $PM_{2.5}$ and $PM_{10-2.5}$ cut for potential chemical characterization. This is a disadvantage of the prosed FRM which should be further considered by EPA. Assuming the analysis of the $PM_{2.5}$ and PM_{10} cuts colected by the two samplers of the FRM will provide solid data on the composition of the $PM_{10-2.5}$ cut is fraught with many assumptions and unknowns. These include the uncertainties in a difference measurement to obtain the coarse particle composition and the probable alteration in the coarse particle composition as it is collected with fine material of quite different composition. This concern further supports the need to make allowance for the use of virtual impactors in ideally an FRM, but certainly FEMs, recognizing the difference seen between the two samplers is not a measure of relevance to the health data set on which the standards are based. EPA should not discard the logical decision to create two FRMs, one based on a difference measurement and one based on technology such as that present in the dichot. This opens the door for reasonable chemical speciation data being obtained on the coarse particle fraction in the future.

3. Consideration of Allowing for Both a Monitoring and Scientific Use for Data Obtained with A Continuous FEM Sampler.

We talked about the importance of being sure that the requirements put in place for certification of a continuous FEM based on comparison with the FRM be loose enough that samplers which do a better job of measuring the semi-volatile material in collected particles not be excluded for monitoring use. One possible way to do this is to make the window for comparison quite large, as discussed at the meting. Another way to approach this is to let those samplers which can measure both the nonvolatile and the semi-volatile fractions separately be allowed to use the former for FRM measurement (perhaps treated as fine TEOM data is now treated) and the latter for scientific input. This will open the door for the health community to begin asking questions which are currently not possible because of the limitations of the FRM data. Instruments which have the capability of providing this input include the FDMS TEOM instruments and the GRIMM monitor.

There is ample data in the current data set to demonstrate the presence of SVM in the monitored $PM^{2.5}$ fraction based on the comparison of proposed FRM and the combination of FDMS TEOM, GRIMM and Beta Gauge data. These data emphasize the need for consideration of a dual monitoring and a scientific role for the real-time instruments which EPA would like to incorporate in the national monitoring network.

There currently is not yet data in the EPA set which answers the question of the presence of semi-volatile material in the coarse particle fraction. The GRIMM data suggest there may be. Confirmatory data from the FDMS TEOM monitor are not yet available. Hopefully this will be addressed in the ongoing study.

Comments on Potential changes to the $PM_{2.5}$ FRM

Delbert J. Eatough
15 September 2005

A. Overall.

Four changes are being considered to the fine particulate FRM method:

1. Replacement of the WINS with a Very Sharp Cut Cyclone.

2. Use of a different WINS oil.

3. Increase in the time allowed for filter recovery.

4. Changes in the transport temperature and post sampling recovery time.

The reasons for each of these proposed changes are well spelled out and the recommendations seem reasonable to me.

Mr. Dirk Felton

Responses: Dirk Felton, NYSDEC (Submitted Sept 16)
Charge to the CASAC AAMM Subcommittee
September 21-22, 2005 Meeting

Questions associated with Attachment 1:

What are the scientific and operational strengths and weaknesses of the $PM_{10-2.5}$ difference method relative to other options for a proposed FRM, especially when used as the basis for the approval of other methods?

The FRM difference method uses the same size selective inlets as the FRM for PM-2.5 and PM-10, the same sampling conditions, the same filters and the same mass determination. Many of these characteristics of the FRM(s) for PM-2.5 and for PM-10 create differences in how these methods work in various regions of the country. For instance the volatile components of PM-2.5 are only partially retained on the FRM and this retained fraction varies seasonally and geographically. It is unlikely that a proposed automated method would be able to mimic the behavior of both the FRM for PM-2.5 and PM-10 in all of the expected conditions where it would be required to work and be consistent with existing FRM measurements.

The principle disadvantage of the FRM difference method is the length of time between sample collection and data availability. This is primarily a problem if in the future air monitoring entities are required to produce public health related notices of $PM_{10-2.5}$ concentrations in near-real time. The length of time between sampling and data availability is not really an issue for an $FRM_{10-2.5}$ that is primarily intended as a benchmark for potential equivalent automated methods.

Another often stated disadvantage of the difference method is the expense of servicing a manual sampler and the associated lab and shipping costs with making gravimetric mass determinations. It is likely that these costs are minimized because the field staff required to service the PM-10 sampler would already be assigned to service an existing PM-2.5 sampler. Other costs such as shipping would be minimized by capitalizing on the existing PM-2.5 network operation. These costs savings are based on a future network of $PM_{10-2.5}$ samplers that is similar to or smaller in scope than the existing PM-2.5 network.

Operationally, a primary advantage of the difference method is that it utilizes existing technologies that air monitoring agencies are comfortable with and have been able to attain adequate precision and data availability. Many agencies will also have surplus samplers if NCORE monitoring initiatives allow a portion of the existing PM2.5 network to close.

Another, often overlooked advantage of the difference method is that it utilizes samplers that do not need an external environmental enclosure. In built up urban areas it is difficult to site equipment that needs both a temperature controlled environment and an inlet suitable for aerosols that meets monitoring siting criteria.

Based on the field study report as well as any other available data, e.g. data from other State and local agencies, how does the demonstrated data quality of the $PM_{10-2.5}$ difference method support or detract from it being proposed as a FRM?

Many of the original complaints about difference data came from early data analysis that used high volume PM-10 FRM data and PM-2.5 FRM data. This is not appropriate because the high volume samples include less of the volatile material that is included as part of the PM-2.5 FRM measurement. Subtracting PM-2.5 FRM data from high volume PM-10 data results in biased and often negative Coarse mass determinations.

The proposed $PM_{10-2.5}$ difference method utilizing collocated identical samplers one with the WINs impactor removed, does produce robust measurements suitable for use as an FRM. The New York State Department of Environmental Conservation (NYSDEC) has operated $PM_{10-2.5}$ difference method samplers in Manhattan and in Niagara Falls for more than three years. The following data summary includes data from 2002 through 2004 on a one day in three schedule. The high values apparent from the "Max" row were due to the smoke from Canadian wildfires.

Manhattan, NYC (344 Samples: 94%)

	PM-2.5	PM-10	Coarse	Ratio
Max	82.71	88.79	26.71	0.93
Min	3.96	6.17	1.71	0.22
Average	15.41	25.48	10.07	0.59
Std Dev	9.31	11.98	4.73	0.12

Niagara Falls, NY (298 Samples: 82%)

	PM-2.5	PM-10	Coarse	Ratio
Max	44.75	53.28	40.79	0.92
Min	1.88	4.25	0.79	0.15
Average	11.54	20.44	8.90	0.56
Std Dev	6.95	9.80	5.57	0.16

Excludes Smoke Event

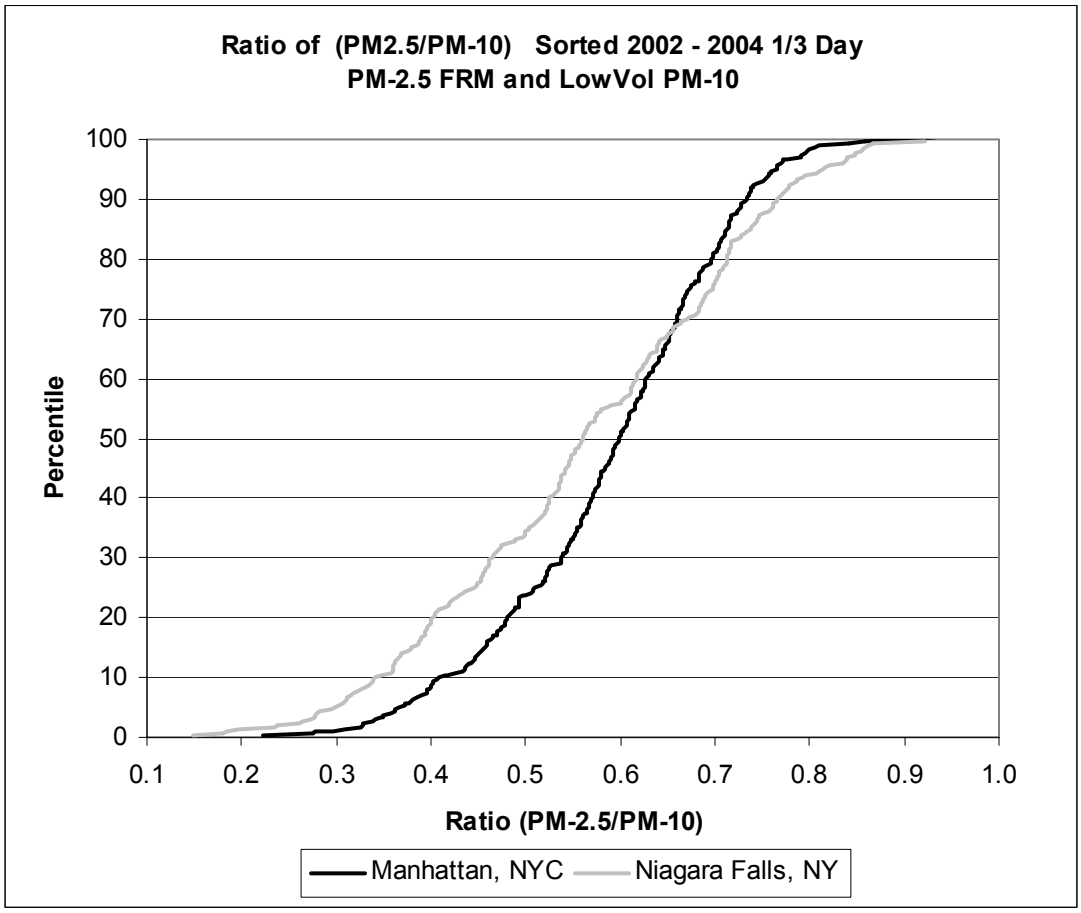
Manhattan, NYC (343 Samples: 94%)

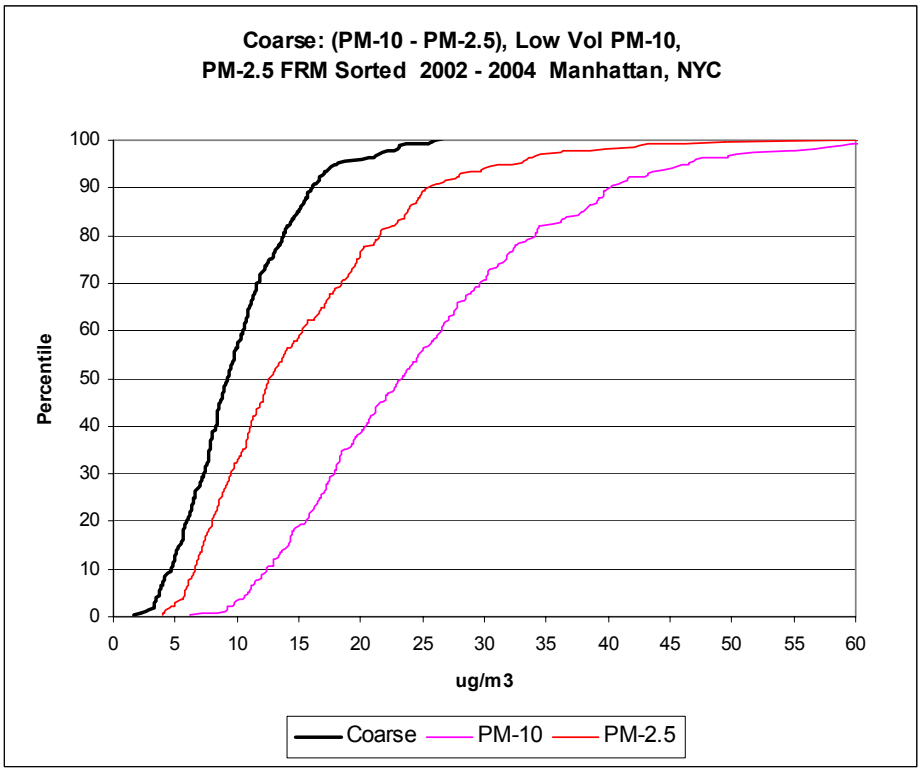
	PM-2.5	PM-10	Coarse	Ratio
Max	58.13	81.25	26.71	0.90
Min	3.96	6.17	1.71	0.22
Average	15.22	25.30	10.08	0.59
Std Dev	8.58	11.50	4.73	0.12

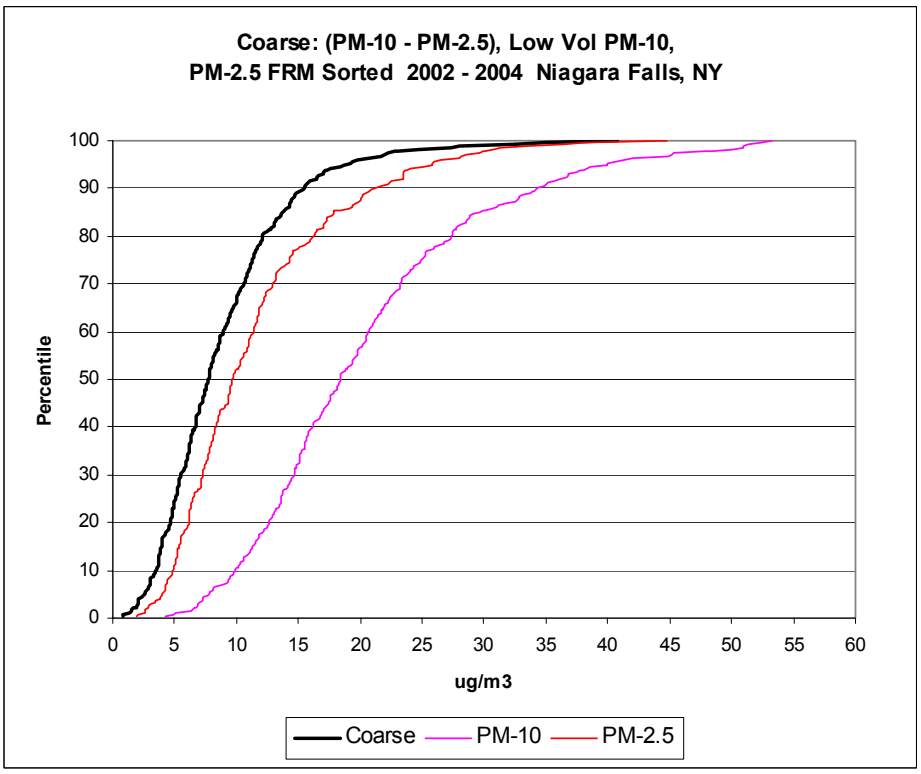
The Canadian Smoke Event was not significant in Niagara Falls.

The collocated low volume PM-10 was operated by the NYSDEC field staff who normally operate PM-2.5 FRMs. No special care or handling was directed towards the collocated sampler. State and Local Agencies comfortable with the PM-2.5 FRM program should not have difficulty producing quality $PM_{10-2.5}$ data by difference.

The NYSDEC $PM_{10-2.5}$ sampling program does not include a precision sampler but the overall results seem to be similar to what EPA ORD found in Birmingham in 2003 – 2004. This is not surprising since the Ratios of PM2.5/PM10 are similar. It is more reassuring that the results from the Multi-Site Evaluation in 2003 – 2005 demonstrated acceptable precision in each area and season where the $PM_{10-2.5}$ difference method was evaluated.


Ratio of (PM2.5/PM-10) Sorted 2002 - 2004 1/3 Day
PM-2.5 FRM and LowVol PM-10
Percentile
100
90
80
70
60
50
40
30
20
10
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Ratio (PM-2.5/PM-10)
Manhattan, NYC
Niagara Falls, NY


Coarse: (PM-10 - PM-2.5), Low Vol PM-10,
PM-2.5 FRM Sorted 2002 - 2004 Manhattan, NYC
Percentile
100
90
80
70
60
50
40
30
20
10
0
0
5
10
15
20
25
30
35
40
45
50
55
60
ug/m3
Coarse
PM-10
PM-2.5


Coarse: (PM-10 - PM-2.5), Low Vol PM-10,
PM-2.5 FRM Sorted 2002 - 2004 Niagara Falls, NY
Percentile
100
90
80
70
60
50
40
30
20
10
0
0
5
10
15
20
25
30
35
40
45
50
55
60
ug/m3
Coarse
PM-10
PM-2.5

Question associated with Attachment 2:

Based upon the latest available field study data, which $PM_{10-2.5}$ methods have both sufficient utility to meet one or more important monitoring objectives and appropriate data quality to be considered for deployment as Federal Equivalent Methods (FEMs) or speciation samplers in a potential $PM_{10-2.5}$ monitoring network?

Currently it looks like the Manual Dicot and perhaps the re-designed sequential Dicots have the potential to be an FEM with their current configurations. The fact that a portion of the Coarse mode particles are deposited on the fine filter could cause some minor geographic and seasonal discrepancies in comparisons with the FRM difference method. The Dicot methods also seem to suffer somewhat from the loss of Coarse mode particles on Teflon filters. If filters could be substituted that are more suited to retaining the Coarse mode particles without other artifact issues than the method could be further improved.

Running Dicot FEMs may not be much of an advantage to air monitoring Agencies versus running two collocated FRM samplers. The Dicot PM2.5 data may not be good enough to replace the instrument used for FRM PM-2.5 monitoring and the cost of filter and equipment maintenance for a Dicot is nearly identical to that of two stand alone instruments.

Questions associated with Attachment 3:

What are the Sub-committee's views on the Very Sharp Cut Cyclone (VSCC) being approved as an alternative second-stage impactor to the WINS for use on a PM-2.5 FRM?

The attachment did not include the data showing the comparisons between the WINS and the VSCC nor any indications that there were regional or seasonal differences between the two inlets. This data must have been quite convincing as the VSCC has already been designated an FEM. The NYSDEC evaluated a SCC collocated with a WINS in 1999 in NYC over several months. The regression results (SCC = 1.017 FRM - 0.174 and $R^2 = 0.994$) showed that at least in New York City, a well maintained SCC could justifiably be used in place of a WINS impactor.

The advantages of the VSCC over the WINS and the similarity of the resulting data support the decision to approve the use of the VSSC in an FRM. The caveat is that even though the VSCC has a longer service interval than the WINS, a neglected VSCC can cause problems with data integrity. The WINS must be serviced after five 24-Hr runs. This service includes cleaning, filter exchange, re-oiling and inspection of the inlet. If there is a problem found during the service of the WINS such as an obstruction in the flow path (spider or fluff) or a leak due to a loose fitting or worn o-ring, no more than 5 sample days have to be invalidated. If a problem is found with a VSSC during a less frequent service, a greater amount of data may have to be invalidated.

A minimum service interval should be specified for the VSCC as part of the approval designation. A service interval of Monthly or after every ten 24-Hr sample days would be a

reasonable compromise between the necessary cleaning frequency and the amount of data that would have to be invalidated if there was a problem. To reduce the burden on the agencies performing the work, this VSCC service interval also coincides with when the operator would normally perform instrument audits.

The cost of a VSCC is significant and monitoring Agencies that opt not to switch inlets should not be penalized for using a WINs. This should not be a problem as long as there is no requirement for collocating a FRM and an FEM. Allowing both the WINs and the VSCC to be an FRM would eliminate this potential problem.

To what extent are the stated advantages of relaxing existing requirements identified for the PM-2.5 FRM supported by the information cited in attachment 3, available literature, or good field and laboratory practices? Does the Subcommittee have additional recommendations for the PM2.5 FRM that would be neutral with respect to bias, but would improve the performance and minimize the burden on agencies conducting the sampling?

Recommendation number 2 (WINs oil) is fine but it needs a reference to the original requirement and description for the oil in the WINs impactor. Presumably the new oil also has a low vapor pressure and will not interfere with analyses other than gravimetric that may be performed on the filters.

Recommendation number 3 is fine and the supporting draft study performed by members of the EPA and several air monitoring Agencies justifies the change in filter recovery time. The one caveat is that there is significant negative bias at two sites that may be related to the composition of the aerosol at those locations. For that reason, there is not enough information to go ahead with recommendation number 4. A study should be designed to look for significant negative or positive bias resulting from the change in the temperature requirements prior to final weighing of the filters. This study should include several geographic areas and all seasons to insure that no areas of the country would have data that become biased high or low once the changes to the regulations were enacted.

Attachment 3 has a summary that in part states that these recommendations "would provide more consistency with other filter-based networks such as IMPROVE". This statement particularly with respect to recommendation 4 is not entirely accurate. The IMPROVE program has inconsistent filter handling at many of their monitoring locations. Some IMPROVE filters are stored for 30 days in enclosed creosote coated plywood buildings while at other sites the filters are stored in air conditioned labs. This type of inconsistency will become more apparent if IMPROVE is used in more urban monitoring locations.

Dr. Philip Hopke

Comments by
Philip K. Hopke

Question 1: What are the scientific and operational strengths and weaknesses of the $PM_{10-2.5}$ difference method relative to other options for a proposed FRM, especially when used as the basis for approval of other methods?

I am very disappointed by the effort to keep pushing the difference method as an FRM. We have had several prior discussion about the difference method and the committees involved in those reviews have consistently indicated their dissatisfaction with the difference method as the basis of an FRM. We understood the potential necessity for it when we were looking at the possibility of a standard being promulgated 3 years ago, but there were serious concerns and there remain serious concerns regarding the chemical interactions between PM_{10} and $PM_{2.5}$ and not having a sample of $PM_{(10-2.5)}$ available for chemical or other analysis. The claim is that the paired sampler provides useful samples for chemical analysis. In fact, the combination of the PM_{10} and the $PM_{2.5}$ make the chemical analysis more difficult. There is the claim that they need the $PM_{2.5}$ to keep the $PM_{(10-2.5)}$ to stay on the filter. However, there are other approaches that can be used to affix the coarse particles to the sampler such as small quantities of fluorocarbon or silicon grease. This appears to be a case of having made up your minds as to the sampling approach and doing what you can to prove it is what you wanted. It is simply not possible to support this proposal. It is not the best monitoring science and represents a major step backward.

I would suggest that it will be quite possible to provide a set of performance specification that require a continuous monitoring system that provides adequate separation of the fine and coarse particles that there is no need for a subtraction correction to be made. It is not necessary to specify any specific technology to make this separation and/or enrichment of the coarse particles nor the measurement technology. Only the precision of the measurements needs to be specified.

It is clear I am in the minority on this issue, so I would suggest that the Subcommittee at least recommend that the modern dichotomous sampler be approved as a second FRM. It is unfortunate that EPA has not even developed a simple two-stage cascade impactor to be the FRM. It would provide clean samples of coarse and fine particles at a standard flow rate of 16.7 LPM. One could have easily designed, built and tested such a unit within the funds that have been expended for testing commercial units. A major opportunity has been lost to move away from the difference method and still provide integrated samples for analysis. I would prefer the performance standards for a continuous monitor as an FRM, but at least require a sampler that actually collects coarse particles!

The logic of suggestion that more finely time resolved data is not needed escapes me. If there had been no $PM_{2.5}$ data before there was a standard, there would have been no basis for creation of the $PM_{2.5}$ NAAQS. Without complete and time resolved data, we are never going to be able to sort out the time course of the effects of PM on health. The proposed FRM again only provide data on 33% of the days and only in 24 hour increments. Yes, it permits testing attainment of the

NAAQS, but we are quite certain that the current NAAQS has significant flaws because we do not yet really understand the relationship between exposure and adverse health effects. With complete data that covers a much larger percentage of the days of the year and provides time resolved data, we will not be able to better understand the time course of the onset of the adverse effects and the useful insights into potential mechanisms that such an understanding will provide.

Question 2: Based on the field study report as well as any other available data e.g., data from State and local agencies, how does the demonstrated data quality of the $PM_{10-2.5}$ difference method support or detract from it being proposed as an FRM?

The test in Jefferson County, Alabama does not really tell us very much about the ability of the average SLT agency to operate paired samplers. All of the samplers were the relatively limited use BGI samplers rather than the more commonly employed MetOne or more automated samplers like the R&P where there is more movement of the filters into and out of the sampling position. Thus, it is not clear that this is a meaningful reflection of the ability of the likely endusers to produce non-negative data.

A key part of the past tests that have been made should be the careful examination of the chemical composition data that have been accumulated from the instrument intercomparison field studies. Many of the samples have been analyzed and the data provided to the subcommittee, but there are now filters from the latest field campaigns that need to be analyzed and more importantly, time needs to be set aside to examine and interpret these results. This should be done before going ahead with designation of the difference method so that one can really see if the samples can actually be used for chemical characterization of the trace constituents of the coarse mode particles.

Consultation Question 1: Based upon the latest available field study data, which $PM_{10-2.5}$ methods have both sufficient utility to meet one or more important monitoring objectives and appropriate data quality to be considered for deployment as Federal Equivalent Methods (FEMs) or speciation samplers in a potential $PM_{10-2.5}$ monitoring network?

The performance of the continuous monitors in the latest trials was actually quite encouraging. It is clear that several of the units would have the chance to pass a set of **rigorous** performance standards that could be set for an FRM. Thus, it suggests that a more forward thinking FRM could be designated to provide the data we need to begin to really understand the health effects of $PM_{(10-2.5)}$.

Consultation Question 2: What are the Subcommittee's views on the Very Sharp Cut Cyclone (VSCC) being approved as an alternate, second-stage impactor to the Well Impactor Ninety-Six (WINS) for use on a $PM_{2.5}$ FRM?

This is a positive step. The problems identified with the WINS in terms of maintenance issues are clear and the small loss in the sharpness of the cut relative to the ease of operation of the VSCC relative to the WINS, clearly is an improvement.

Consultation Question 3: To what extent are the stated advantages of relaxing existing requirements identified for the $PM_{2.5}$ FRM supported by the information cited in Attachment 3, available literature, or good field and laboratory practices? Does the Subcommittee have additional recommendations for the $PM_{2.5}$ FRM that would be neutral with respect to bias, but would improve the performance and minimize the burden on agencies conducting the sampling?

There is the critical problem remaining with respect to defining what is to be measured. We know that the $PM_{2.5}$ FRM fails to accurate measure the actual airborne PM mass with substantial loss of semivolatile species such as nitrate and organics. It would be useful to make more accurate measurements of the actual $PM_{2.5}$ mass as has been demonstrated to be possible using the current state-of-the-art systems such as the R&P filter dynamics measurement system (FDMS). It is time to move beyond the past and get into measurements that better represent what is actually in the air and with the completeness and time resolution to permit a better understanding of the relationships between $PM_{2.5}$ mass and health effects.

The DQO tool for examining the performance of continuous monitors as FEMs for both fine and coarse should be modified to include an asymmetric multiplicative "bias" factor. It is unfortunate to call the slope above zero to be "bias" since it implies that the current flawed FRMs are accurate which has been clearly demonstrated to be untrue. However, to permit more accurate measurements to be included that will produce larger mass concentration values, an asymmetric interval would permit better measurement tools to qualify as FEMs. Given the past performance, it would probably be wise to couple the wider range of positive multiplicative factors with a narrower range of additive factors. It has been seen in the supersite and other studies that instruments like the FDMS have an intercept that is close to zero when compared to FRM mass values and thus, we should not permit as wide a range of additive "bias" terms for the larger positive side multiplicative "bias" side of the compliance parallelogram. It would make sense to look at the sites where there are co-located FRM and FDMS systems to explore an appropriate upper bound on the multiplicative bias term. It would be useful to permit the monitoring agency to either report the best estimate of mass (including the semivolatiles) as well as the base mass values. The base mass values will more closely agree with the FRM values and may be more acceptable to the local authorities who may not want to use the higher mass values that arise from a more accurate measurement of the airborne particulate matter mass. However, although they can use the base mass measurements for attainment determinations, they should be required to report the more accurate mass measurements so that they are available for other scientific purposes.

In terms of site choices for testing equivalency, it may be good to look at a matrix of effects to which one wants to challenge the samplers.

High $PM_{(10-2.5)}/PM_{2.5}$	High semivolatiles
Low $PM_{(10-2.5)}/PM_{2.5}$	Low semivolatiles

Dr. Rudolf Husar

Comments on
Particle Methods and Data Quality Objectives
by Rudolf Husar, Washington University, Sept 18, 2005

ATTACH. 1: Summary and Rationale for the PM10-2.5 FRM

For the accurate and reliable determination of daily average coarse mass concentration the suggested PM10-PM25 difference method is quite compelling. The problem lies in the marginal utility of the PM10-PM25 difference method for estimating the potential health effects, visibility effects or for that matter learning more about the nature of coarse PM. Without speciation, it is virtually impossible to separate the small fraction of anthropogenic coarse PM from the bulk of benign, mostly natural and uncontrollable soil dust. Clearly, the difference method is not suitable for coarse PM speciation analysis.

A closer coordination with the NAAQS – setting processes would seem highly beneficial. In particular, it would seem logical to recommend the relevant FRM **after** the gross features of the standard have been set.

ATTACH. 2: Evaluations of Candidate Methodologies for Coarse PM

The field studies for the evaluation of candidate coarse PM methodologies constitute the pillar of this preparatory activity prior to the promulgation of a standard. The multi-site, multi instrument, multi season and multi-stage (repeated after feedback) comparison was commendable. The presentation of results is clear and useful.

ATTACH. 4: Equivalence Criteria for Continuous PM25

In addition to the statistical intercomparison measures given in the report, it would be helpful to incorporate and weigh qualitative differences, between the methods, e.g. unattended operation/operating cost; availability of filters for subsequent speciation analysis etc. The idea is to make the methods-intercomparison as complete as possible. That way, methods-evaluations and decisions are made using compatible metrics.

ATTACH. 5: DQO Sensitivity to Spatial Uncertainties

The spatial uncertainties of the DQO remain to be a weak part of the analysis package. In response to the CASAC Subcommittee review, the spatio-temporal model has been updated with a few cosmetic changes but the procedures are still inadquate.
A well designed and tested model of cPM pattern could be of great utility for standard setting, network design and general DQO analysis. However, the basic approach of estimating the role of different design/operational parameters on the uncertainty is weak. In fact, it leads the authors

to the unreasonable conclusion that spatial variation of coarse PM is of marginal importance compared to the other factors. This is hardly defendable, particularly in urban settings with strong spatial texture of both cPM source and transport pattern.

- The model is still based on an untested set of assumptions about the aerosol pattern. Verification of the model with historical data at a few characteristic sites was probably done by the model developers, but for reasons unknown, they were not shared in the summary.
- A particularly poor assumption is that there is a mean and a **constant Coefficient of Variation (CV).** The coarse particle concentration pattern is highly episodic. Short-term but rare high concentration events are the norm at most locations and seasons. Such pattern can not be adequately modeled by a normal or even log-normal distribution.
- The inclusion of spatial variability into the model using Design Values, is obscure at best. Why not looking at spatial variations and spatial correlations using actual observations that are available in many cities.

Finally, the conclusion that the spatial variability of cPM is much less significant than sampling frequency, data completeness, the CV is unreasonable. Looking at any urban cPM pattern (the real data from actual monitors, not a model) shows that cPM is spatially heterogeneous and has a large 'gray zone'. A possible reason that sampling frequency was found to be a major contributor to uncertainty is the input assumption of the (long) 1 day or 3 day sampling. Continuous monitors virtually eliminate the temporal uncertainty at a monitoring site, so the remaining uncertainty is dominated by the spatial texture and the other factors listed

ATTACH. 6: **PM10-2.5 Methods Equivalency Development**

The evaluation of equivalency through dedicated inter-comparison of instruments is a sound beginning. In addition to the statistical intercomparison measures given in the report, it would be helpful to incorporate and weigh qualitative differences, between the methods, e.g. unattended operation/operating cost; availability of filters for subsequent speciation analysis etc. The idea is to make the methods-intercomparison as complete as possible. That way, methods-evaluations and decisions are made using compatible metrics.

Dr. Kazuhiko Ito

Comments on Particle Methods and Data Quality Objectives.

9/30/05
Kazuhiko Ito, NYU

The implication of the letter from the CASAC PM Review Panel

Before responding to the charge question, I need to briefly discuss the implication of the letter from the CASAC PM Review Panel (distributed during the review meeting but dated September 15, 2005; EPA-SAB-CASAC-05-012) on the process of establishing $PM_{10-2.5}$ FRM methods and planning. After reading this letter, it became clearer to me that, while the toxicity of the rural coarse particles remains to be determined, the focus of the coarse particle standard appears to be the "urban" $PM_{10-2.5}$, or $UPM_{10-2.5}$, and the great emphasis will be placed on identification of the compositions of $UPM_{10-2.5}$. For example, the letter says:

> "…and there is a need for more data that relate the composition of the particulate matter to adverse health effects. We anticipate that future coarse- and fine-mode particulate standards will give greater weight to particulate composition as a critical element in defining the risk of adverse health effects. Data are needed on ambient concentrations in each size range in terms of mass concentrations and speciation. Continuous monitors for mass, as well as for key components or source-related tracers, will provide the best and most cost-effective means of collecting such data for both epidemiologic research and compliance monitoring. …" (from 1st paragraph on page 3, EPA-SAB-CASAC-05-012)

Thus, the planned FRM and FEM methods need to accommodate the need to collect speciation data (or some specific component of $PM_{10-2.5}$). With this emphasis plus some of the comments I heard during the meeting, I revised some of my initial answers to the charge questions.

I also learned, from the EPA presentation during the meeting, a $PM_{10-2.5}$ network design "similar in concept to $PM_{2.5}$ monitoring for daily standard is being considered" by the EPA staff. I think it is very important to start considering the network design for $PM_{10-2.5}$ chemical speciation data now. Though we don't have data from multiple monitors' $PM_{10-2.5}$ chemical speciation monitors yet, we may be able to at least develop a conceptual framework for the $PM_{10-2.5}$ chemical speciation monitoring based on what we learn from the $PM_{2.5}$ chemical speciation data collected from multiple monitors within cities so far. I have learned, based on the data from three $PM_{2.5}$ chemical speciation monitors in New York City, that the extent of spatial correlation varies across species (Ito K, Xue N, Thurston GD. Spatial variation of PM2.5 chemical species and source-apportioned mass concentrations in New York City. Atmospheric Environment, 2004; 38: 5269-5282). During the meeting the EPA asked for the sub-committee members' opinions on the candidate cities where more test $PM_{10-2.5}$ data would be collected. I suggest that EPA collect more $PM_{10-2.5}$ data using the candidate methods (and trial $PM_{10-2.5}$ speciation data, if possible at all) in the cities where multiple $PM_{2.5}$ speciation monitors are collecting data, so that some relationship between spatial variation of $PM_{2.5}$ vs. $PM_{10-2.5}$ could be examined and applied to future $PM_{10-2.5}$ chemical speciation monitoring network design.

Questions associated with <u>Attachment 1</u> – Selection and technical summary of $PM_{10-2.5}$ FRM:

1. What are the scientific and operational strengths and weaknesses of the $PM_{10-2.5}$ difference method relative to other options for a proposed FRM, especially when used as the basis for approval of other methods?

<u>Strength:</u>
Since the existing FRM PM_{10} and $PM_{2.5}$ are also measured using the same principle, samplers, and operating procedures, there is continuity in interpreting/comparing the past PM_{10}, $PM_{2.5}$ data with the future $PM_{10-2.5}$. Its filter samples should accommodate subsequent size-specific chemical analysis ($PM_{10-2.5}$ and $PM_{2.5}$ speciation). The multi-site evaluations of the difference methods using the samplers from different manufacturers showed very high precision, which is promising.

<u>Weakness</u>:
If the operational and maintenance cost for the difference method samplers (requiring two samplers and daily changing of filters for mass measurement alone) is much higher than those for the automated continuous and semi-continuous samplers, the cost can be a weakness. This is particularly so if spatial non-uniformity of $PM_{10-2.5}$ within a city requires more monitors for $PM_{10-2.5}$ (than for $PM_{2.5}$).

There are at least two issues with the difference method that were raised by the sub-committee members during the meeting. The first one was the fact that even the current FRM method for $PM_{2.5}$ has the potential loss of volatile compounds, and the proposed difference method would have the same problem. I think the seriousness of this issue would depend on: (1) how much of the volatile compounds we are missing from the samples, and (2) the health effects of the volatile fraction (mostly nitrate, I imagine). I am not aware of studies that identified nitrate as important component of PM that are associated with health outcomes, but then again few studies had available data on nitrate. The second issue with the difference method is, for chemical speciation purpose, collecting particles on the PM_{10} filter would be potentially chemically mixing $PM_{10-2.5}$ and $PM_{2.5}$ species on the PM_{10} filter. Thus, subtracting the speciated $PM_{2.5}$ data from the speciated PM_{10} data may not give us adequate speciated $PM_{10-2.5}$ data that would represent the $PM_{10-2.5}$ chemical components in actual ambient air. I am not familiar with concrete examples of this issue, but this may need to be examined in the future testing.

My overall impression is that the difference method was reasonable for FRM, but considering the expected "large" (compared to the precision/accuracy of the alternative monitoring instruments) spatial variation of $PM_{10-2.5}$ within a city, we may need to consider putting multiple, cheaper, FEM's in addition to one or two FRM's within a city.

2. Based on the field study report as well as any other available data, e.g., data from State and local agencies, how does the demonstrated data quality of the $PM_{10-2.5}$ difference method support or detract from it being proposed as a FRM?

From the "data user" point of view (for epidemiological studies), the difference method has the desired very high precision and continuity to the past and current $PM_{2.5}$ FRM data. Since the

variation due to spatial heterogeneity is a far bigger concern to me than the precision of the proposed samplers, I would be satisfied with the proposed difference method FRM. The study in Birmingham, Al seems to suggest that $PM_{2.5}$ is more uniformly distributed than $PM_{10-2.5}$ (37% mean level difference for PM2.5 vs. a factor of three difference for $PM_{10-2.5}$).

Consultation Questions:

Question associated with Attachment 2 – EPA's Multi-Site Evaluations of Candidate Methodologies for Determining Coarse Particulate Matter ($PM_{10-2.5}$) Concentrations: August 2005 Updated Report Regarding Second-Generation and New $PM_{10-2.5}$ Samplers:

1. Based upon the latest available field study data, which $PM_{10-2.5}$ methods have both sufficient utility to meet one or more important monitoring objectives and appropriate data quality to be considered for deployment as Federal Equivalent Methods (FEMs) or speciation samplers in a potential $PM_{10-2.5}$ monitoring network?

With so many instruments and field study results (and not being an "instrument" expert myself), I could not compare the pros and cons of these methods by just reading Attachment 2. The way the results were presented for these samplers were not always the same across the sampling campaigns. Therefore, I constructed a table below, extracting the ratios of each instrument to FRM, R^2 from regression of each method's values on those of FRM, and statements to summarize the results for myself. The R^2's of these methods were mostly very high (> 0.9), except the methods that measure size distributions, and the main issue appears to be the constant bias (over- or under-estimation). However, based on the 2003 and 2004 studies, this "constant bias" appears to also vary across locations (regions) for a given method, perhaps due to the regional difference in chemical compositions or size distributions. Thus, the alternative sampler to FRM mean ratios obtained in the 2005 Phoenix, AZ study may be "snapshots" and these may also vary, and we don't have the data on these region-specific variations for the newer instruments introduced in the 2005 study. In this situation, it may be necessary to consider site-specific (or city- or region-specific) calibration of FEM samplers to FRM samplers. Obviously, since the apparent objectives of these samplers vary (e.g., near real-time measurement, size distribution measurement, speciation), the choice of methods will need to be discussed for each objective.

For routine FEM (and possibly for speciation) purposes, I could not see major differences among the R&P dichot (sequential, sequential/manual mode, single-event) and Sierra-Andersen dichot samplers. Since the discrepancy between the FEM and FRM samplers may be region specific, it seems necessary to co-locate the FRM and FEM samplers in each region, at least initially. If the operational cost for the FEM samplers were significantly lower than that for the FRM samplers, then the FEM samplers may be used to measure the spatial variability of $PM_{10-2.5}$ in the city of interest, at least initially, and reduce the number of such monitors as appropriate.

Three methods (Kimoto dichot beta gauge, R&P Coarse TEOM, and R&P dichot TEOM) are available for near real-time mass measurements. The Kimoto dichot beta gauge showed a major over-estimation of $PM_{2.5}$ even after design modification, whereas the R&P dichot TEOM sampler showed a major under-estimation of $PM_{2.5}$, with many negative values. Thus, these monitors may require further modifications. However, practically speaking, we may not need

near real-time dichot monitors that accurately measure $PM_{2.5}$ as long as a $PM_{2.5}$ TEOM monitor is running in the area (note that $PM_{2.5}$ is expected to be more spatially uniform and therefore we do not need as many $PM_{2.5}$ monitors as we do $PM_{10-2.5}$ monitors). In that sense, the R&P Coarse TEOM monitor may be a convenient choice for monitoring short-term excursions as well as spatial variations (if the cost is low enough so that multiple coarse TEOM monitors could be operated).

There were two methods (TSI APS and Grimm EnviroCheck) that measure size distributions in real time. Both seem to suffer from R^2's that are lower than those for the other methods, and may require further modifications.

So far, we are dealing with accuracy and precision in terms of $PM_{10-2.5}$ mass concentrations. We may also have to start thinking about the issues associated with chemical speciation of $PM_{10-2.5}$ data.

Table 1. Cursory comparison of various methods from the field study results

Sampler	Main purpose	Comparison of $PM_{10-2.5}$ with FRM (the ratio to FRM and R^2)	Other comments
Collocated PM and PM FRM Samplers	FRM, speciation	NA	High precision; high data capture rate/ few functional problems
R&P Model 2025 Sequential Dichotomous Sampler	Unattended multi-day operation possible with a filter exchange system	Small but consistently under-measured $PM_{10-2.5}$ in 2003 test R&P to FRM ratio = 0.80 to 0.96; 0.89 in 2004 test; 0.93 in 2005 test. R^2 ranged from 0.968 to 0.979.	Some operational problems in the field tests; high precision; some intrusion of coarse into the fine channel; The R&P to FRM ratio for PM2.5 ranged 1.00 to 1.08.
Kimoto Inc. Model SPM-613D Dichotomous Beta Gauge	Near real-time measurement	The ratios of Kimoto to FRM for $PM_{10-2.5}$, the ratios ranged from 0.91 to 1.08 for 2003 and 2004 tests; after modification, the ratios ranged ~1.05 to 1.13 in 2005 tests; R^2 ranged from 0.957 to 0.995.	Consistent overestimation for $PM_{2.5}$. The ratios of Kimoto to FRM for $PM_{2.5}$ were consistently high (1.26 to 1.70) in 2003 and 2004 tests
R&P Continuous Coarse TEOM Monitor	Near real-time measurement of $PM_{10-2.5}$	The ratios to the TEOM to FRM mostly low but varied across sites ranging 0.69 to 1.05 in 2003 and 2004 tests; but design modification appeared to have improved the ratio (~ 1.04) in 2005 tests. R^2 ranged from 0.926 to 0.999.	Few operational problems
TSI Inc. Model 3321 Aerodynamic Particle Sizer (APS)	Size distribution of particles (larger than > 0.7 µm) in real time	The 2004 tests showed a factor of two under-prediction with TSI to FRM ratios, but using an alternative specific gravity and a shape factor, the ratio are now much better (0.76 to 1.02). The 2005 tests with design modification resulted in the ratio of ~0.86 without invalidated data. R^2 ranged from 0.53 to 0.99.	Assumption of the specific gravity and shape factor makes a big difference in results. There appear to be some functionality problems (to be tested in the next tests in 2005).
R&P Single-Event	Unlike R&P Model 2025 ,	R&P to FRM ratio = 0.99 in 2005 tests. R^2 = 0.995.	Not for routine use?

Dichotomous Sampler	the potential for post-sampling loss of large particles is minimized		
Sierra-Andersen Model 241 Dichotomous Sampler	Routine monitoring	Sierra-Andersen to FRM ratio = 0.95 in 2004 tests. $R^2 = 0.995$.	The inlet has been fully wind tunnel evaluated. Some intrusion of coarse into the fine channel; No active volumetric control
BGI frmOMNI Ambient Air Sampler (Filter Reference Method)	For short-term saturation sampling at a relatively low cost.	OMINI to FRM ratio for $PM_{10-2.5}$ ~0.85. $R^2 = 0.949$.	Not for routine use? Some functionality problems in 2005 tests, reducing the data capture rate. OMINI to FRM ratio for $PM_{2.5}$ ~1.07% (with somewhat low $R^2 = 0.808$).
Grimm EnviroCheck Model 1.107 Sampler	Size distribution in real time.	The ratios for $PM_{10-2.5}$ averaged ~1.53. $R^2 = 0.847$.	Grimm to FRM ratios for $PM_{2.5}$ averaged ~1.37.
R&P Dichotomous TEOM Sampler	Near real-time measurement of $PM_{10-2.5}$ and $PM_{2.5}$	Mean dichot to FRM ratios for $PM_{10-2.5}$ were 0.85 and 0.89 for two units used. $R^2 = 0.992$ for the average of two units vs. FRM.	No operational problems during the 2005 tests; negative $PM_{2.5}$ values; Mean dichot to FRM ratios for $PM_{2.5}$ were 0.80 and 0.63 for two units used.

Questions associated with <u>Attachment 3</u> – Memo to PM NAAQS Review Docket (OAR-2001-0017) – Potential changes being evaluated for the $PM_{2.5}$ Federal Reference Method

2. What are the Subcommittee's views on the Very Sharp Cut Cyclone (VSCC) being approved as an alternative second-stage impactor to the Well Impactor Ninety-Six (WINS) for use on a $PM_{2.5}$ FRM?

I am not familiar enough with the background information and support data for VSCC vs. WINS performance to form an opinion on this.

3. To what extent are the stated advantages of relaxing existing requirements identified for the $PM_{2.5}$ FRM supported by the information cited in Attachment 3, available literature, or good field and laboratory practices? Does the Subcommittee have additional recommendations for the $PM_{2.5}$ FRM that would be neutral with respect to bias, but would improve the performance and minimize the burden on agencies conducting the sampling?

The justifications explained in Attachment 3 all seem reasonable to me, but since I did not read the original references cited, I refrain from commenting on this.

Question associated with <u>Attachment 5</u> – Sensitivity of the PM 10-2.5 Data Quality Objectives to Spatially Related Uncertainties

7. To what extent have the assessments of spatial variability and the sensitivity of the DQO process to a variety of population distributions been appropriately addressed?

The document states "The DQO development used preliminary data collected from sites

providing coarse particulate estimates from around the country as well as data from multi-site performance evaluations…" (page 1), but the document does not give us the sense of what a typical (or any example) spatial distribution of $PM_{10-2.5}$ would be like. For example, were there cases in the database in which $PM_{10-2.5}$ variation could be depicted in as small as an 8 km x 8 km grid (like Figure 1 on page 4)? One of the documents distributed for this meeting (Network Operation of the Difference Method: An Independent Study Conducted by the Jefferson County Department of Health In Birmingham, AL, by Vanderpool and Dillard) shows spatial variations of $PM_{10-2.5}$ as measured with seven monitors in Jefferson County, which could be contained in a 50 km x 50 km grid. The data showed that there was a factor of 4 difference in the mean $PM_{10-2.5}$ concentrations across the monitors during 2004. Given a case like this, my question would be: how many monitors are needed to estimate the distribution of ambient $PM_{10-2.5}$ levels the residents in this county are exposed to? I must be misunderstanding the intended use of the DQO model, but the question being asked in Attachment 5 seems to be using somewhat unrealistic scenario, though I am not sure if using more realistic scenarios would make any difference. For example, Attachment 5 states,

> "Hence, the comparison is indicating how well a single monitor does in predicting the true mean design value across the grid area. Since the day-to-day shape of the surface is not fixed, on average, throughout the three-year period, the center should be an unbiased indicator of the mean. Consequently, there is no inherent bias at any site being simulated, unless a strong autocorrelation is used to "fix" the shape of the surface." (page 4)

This (that the center should be an unbiased indicator of the mean) seems unrealistic to me. My impression of the spatial distribution of coarse particles is that they are strongly influenced by local emissions, and such local sources (e.g., some industry complex) are not spatially uniformly distributed within a city or metropolitan area. As a result, there would be some concentration gradient (i.e., spatial autocorrelation) within the city, as in the case of the Birmingham study.

The scenarios used in the simulation seem to assume that there is no constant gradient. For a 8 km x 8 km grid, this may be a reasonable assumption, but I am wondering about the usefulness of this scale. Isn't this DQO also a part of network design? If so, different scales of grids need to be considered. Then, scenarios with a constant gradient will need to be considered. This would mean that the variance (and likely CV) would also spatially vary.

Dr. Donna Kenski

Comments on Attachment 1: Summary and Rationale for the $PM_{10-2.5}$ FRM

Donna Kenski
Lake Michigan Air Directors Consortium
Sept. 20, 2005

Questions: 1. What are the scientific and operational strengths and weaknesses of the $PM_{10-2.5}$ difference method relative to other options for a proposed FRM, especially when used as the basis for approval of other methods?
2. Based on the field study report as well as any other available data, e.g., data from state and local agencies, how does the demonstrated data quality of the $PM_{10-2.5}$ difference method support or detract from it being proposed as a FRM?

This document accurately summarized the case for the $PM_{10-2.5}$ difference method. Both the multisite evaluation (Att. 2) and the Jefferson County, Alabama, data indicate that the method provides high quality data under carefully controlled conditions. None of the other candidate methods can provide the same quality data and other significant advantages inherent in this method. And practically speaking, because the existing FRM for PM2.5 defines it by the methodology used to collect it, the FRM for PM coarse is constrained by that definition because it incorporates PM2.5 as a lower bound. State and local site operators are already familiar with the technology, and instruments now being taken out of service because of downsizing the PM25 network can be redeployed for this purpose. Thus the economic advantages alone are significant, perhaps more so in the face of looming budget cuts.

Despite the clear advantages of the difference method, it is not suitable for providing the real-time data that is critical for public health awareness. I agree with EPA's intention to emphasize deployment of continuous FEMs in the network and use the gravimetric FRMs as audit devices. EPA should clearly specify the QA requirements for collocated FRM instruments and be careful to minimize the burden on state agencies implementing the new network.

I disagree with the statement that the proposed FRM will provide aerosol samples for chemical analysis. Undoubtedly the samples can be collected and analyzed, but the nature of the difference method, and the much higher uncertainty associated with the chemical analyses of these speciation samples, makes the quality of such 'speciation by difference' data highly uncertain. Perhaps it will yield acceptable data, perhaps not; further studies to examine the practicality of the difference method for speciation are needed to demonstrate its utility. The need for speciation data, however, is inescapable, and the virtual impactor has some significant advantages for providing samples for speciation analysis, although its incomplete separation of $PM_{2.5}$ and PM_{10} is problematic. As such, it seems premature to cite the possibility of speciating the difference samples as an advantage of the proposed method. Other candidate methods may be more well suited to providing speciation samples. EPA could and should address these questions about speciation by analyzing the already-collected speciation data from the field studies.

A minor shortcoming in this summary was that it did not address losses of volatile species from the proposed $PM_{10-2.5}$ FRM, except to note that losses would be equal on both filters and thus unbiased. Because the chemical composition of PM on the two filters will be different, the volatility should not be assumed to be the same; differences in acidity or hygroscopicity of the collected particles in each size fraction could affect losses (or gains) on the filters. The data analysis suggested above could also begin to address questions about differing volatility of the two size fractions.

Both the $PM_{10-2.5}$ and $PM_{2.5}$ methods would greatly benefit from a through-the-probe audit system that generates known quantities and sizes of aerosols. Efforts to develop such equipment should be pursued, either in EPA's own facilities or by funding other researchers.

Comments on Attachment 5: Sensitivity of the PM10-2.5 Data Quality Objectives to Spatially Related Uncertainties

Donna Kenski
Lake Michigan Air Directors Consortium
Sept. 20, 2005

Question: To what extent have the assessments of spatial variability and the sensitivity of the DQO process to a variety of population distributions been appropriately addressed?

This revision of the DQO model seems to have adequately addressed the concern about multimodal distributions of PM2.5 that was raised at the July 2004 meeting of the subcommittee. By introducing and testing the effects of phase shifts and biannual concentration peaks, the authors have shown that these distributions have negligent effects on the decision zone boundaries.

The report (and the accompanying documentation on the CASAC web site) was not clear about the significance or adequacy of the grid size used to simulate spatial variability. The choice of an 8-km grid seems too small to represent spatial variability. A larger grid more representative of an urban area (50 km?) seems more appropriate, although that's just a guess due to the lack of any real data. Perhaps the Jefferson County study could be used for validation, at least of spatial variability on the urban scale. Data on spatial variability at the much smaller neighborhood scale should also be collected (either in the field or from a literature review) and data from both scales used to validate the model results. Further clarification on this part of the model would be helpful, including an explanation of the choice of grid size and whether it is influential. It wasn't possible to tell how sensitive the results were to grid size. This analysis, while statistically interesting, seemed to raise more questions than it answered (which was a useful finding in and of itself). In addition to the questions above, issues raised in the subcommittee discussion included the influence of monitor height on variability, and the expectation that spatial variability of the components of $PM_{10-2.5}$ will be quite different from mass.

Dr. Thomas Lumley

Thomas Lumley, Ph.D.
09/16/2005

Comments on Attachment 1

The proposed FRM appears reasonable and well justified.

An important feature of PM10 and PM2.5, as noted in the document, is that they are to a significant extent defined operationally, by the characteristics of inlets, impactors, &c. There are definite advantages in maintaining the same effective definition of PM2.5 and PM10 when defining the coarse subfraction. This operational definition cannot be endorsed without restrictions, of course, as it would imply that improvements in measurement are impossible by definition.

Assuming that measurement error cannot be completely eliminated, no method of measuring PM2.5, PM10 and coarse mode PM can simultaneously guarantee that both subfractions will have non-negative measurements and that their sum will be equal to the PM10 measurement. The proposed FRM guarantees equality with the PM10 measurement but not non-negativity. This is reasonable for NAAQS attainment, as noted in the document, and is also reasonable for research, where occasional negative measurements need not cause any great consternation.

The summary and rationale notes that a difference method will tend to cancel out biases that are common to the PM10 and PM25 subfractions. This is certainly true, but the price to be paid is a magnification of bias in the comparison of coarse and fine subfractions, where biases will no longer tend to cancel. The reasons for not proposing a continuous or semicontinuous monitoring technology as the FRM are cogent. It is still important that development of these monitoring technologies is not retarded by the specfications of the FRM. This is especially true because the adoption of a daily rather than hourly standard for coarse PM is not due to any evidence that daily averages are more relevant to public health than are shorter time intervals. Rather, the availability of daily data has encouraged research using daily average concentrations as a measure of exposure.

An important possibility is that a continuous or semicontinuous measurement technology would give results comparable to the FRM at some sites but not at others [perhaps because they are sensitive to temperature, humidity or aerosol characteristics]. Local validation and use of such technologies for monitoring beyond that required by the standard should be encouraged, even if they cannot qualify nationally as FEM.

Comments on Attachment 4

The proposed FEM criteria seem to be well thought out and well justified. However, it is not clear to me exactly how the criteria relate to the DQO.

The proposed criteria ensure that a new method must agree with the FRM and must have been tested over a reasonably wide range of PM concentrations in at least two seasons. The proposed bounds on error appear reasonable.

One issue that is not discussed is the choice of sites. The performance of (in particular) non-gravimetric methods may well be affected by differences in aerosol characteristics and in temperature and humidity. For example, nephelometry provides high time resolution and excellent agreement with the FRM in Seattle, but does not perform as well in many other locations. There do not seem to be any criteria other than range of PM2.5 concentrations for choosing the sites, and agreement at one set of sites may not guarantee agreement under different conditions at other sites.

It seems that the criteria should include some reason to expect the chosen sites to be representative (or at least the absence of reasons to expect the contrary).

An additional non-technical note: the use of the term "precision" for a quantity that has large values for less precise measurements is unfortunate, resulting in potentially confusing statements such as: "The precision of the FRM sampler is required to be no greater than 7 percent" (p5). In addition to conflicting with ordinary language use of the term, this conflicts with the technical use of "precision" in statistics as the reciprocal of a variance.

Further reservations on attachment 4

I had previously believed that the coefficient of variation (precision) constraint together with the regression bias constraints would necessarily provide a sufficient constraint on the error in the FEM. Warren White's example suggests that this is not the case and that a more careful analysis is needed.

The basic data for determining an FEM are $(FEM - \overline{FRM})$, the difference between a single FEM monitor and the best estimate of the construct measured by the FRM. The use of coefficients of variation gives less weight to these errors at high levels than at low levels. This does not seem sensible either for regulatory purposes (where accuracy at high concentrations is important) or for scientific purposes. The coefficient of variation is widely used as a simple descriptive summary of variation, but this does not imply that it is an appropriate summary for the purpose of qualifying an FEM.

Similarly, but less importantly, it would probably lead to simpler criteria if the closeness of data to the regression line were quantified by the residual mean squared error. Unlike the correlation, this is not sensitive to the range of the FRM.

It may be particularly useful to allow an asymmetric bound on the multiplicative bias, with positive multiplicative bias permitted if precision and additive bias are good. There are good reasons to believe that particles in the air include semivolatile components and that these may be of health importance. It currently appears unavoidable that the FRM excludes these components, but it would be very helpful if FEMs could include them. This would provide an opportunity for transition to a different indicator as scientific understanding develops.

While changing the precision summary to use variance rather than coefficient of variance would be sufficient, the EPA should consider for future development a set of summaries that is more transparently related to the differences between the FRM and candidate FEM measurments. One approach is as follows:

A simple summary of the errors is the averaged squared error

$$ASE = \frac{1}{D}\sum_{d=1}^{D}\frac{1}{T}\sum_{t=1}^{T}(FEM_{it} - \overline{FRM}_{.t})^2$$

This directly measures the difference between the FEM and the FRM, with no assumptions about regression structure or between-instrument homogeneity.

Regression structure is still important for extrapolation from levels of PM encountered during testing to the presumably higher levels relevant for NAAQS attainment. The ASE does constrain the accuracy of the regression line, since

$$\begin{aligned} ASE &= \sigma^2 + a^2 + (1-b)^2\tau^2 \\ &= (1-\rho)\tau^2 + a^2 + (1-b)^2\tau^2 \end{aligned}$$

where τ^2 is the variance of the FRM data, σ^2 is the residual variance in the regression, a is the additive bias evaluated at the mean, and $1-b$ is the multiplicative bias. It also follows from this equation that the R^2 is bounded below, by $1 - ASE/\tau^2$, but σ^2 is probably a better control target as it is less sensitive to the range of concentrations studied.

It would then either be possible to control ASE directly, and to augment this with controls on a and $1-b$ that are explicitly targeted at the extrapolation needed to estimate the NAAQS threshold accurately.

Dr. Peter McMurry

Peter H. McMurry
September 28, 2005

RE: CASAC Ambient Air Monitoring & Methods (AAMM) Subcommittee Meeting: Comments on Materials provided for review

Attachment 1: Summary and Rationale for the $PM_{10-2.5}$ FRM

Proposed Sampling methodology: It is proposed that compliance measurements of coarse particle (2.5-10 µm) concentrations be determined by the difference between mass concentrations obtained using PM_{10} and $PM_{2.5}$ samplers. Both samplers would be based on the current FRM $PM_{2.5}$ instrument with the exception that an impactor or cyclone is used to remove particles larger than 2.5 µm aerodynamic diameter prior to sample collection on the filter for $PM_{2.5}$. Both samplers operate at low volume, thereby reducing evaporative losses, and both collect integrated 24-hour samples. Finally, this standard could be implemented by adopting existing equipment and without additional training of field operators.

Data collected by EPA show that when operated with proper care, the PM2.5 and PM10 instruments provide high precision data. Furthermore, EPA showed that even when mass concentrations are low, PM_{10}-$PM_{2.5}$ is nearly always positive. EPA argues that these qualities support the validity of this methodology for compliance measurements.

I am very impressed with the high quality field work that was carried out by Robert Vanderpool and coworkers when collecting data used to assess proposed sampling techniques for $PM_{10-2.5}$. This work reflects a high level of commitment, thought and energy. It is essential that such work be continued, and that resources dedicated to it be maintained or increased. It is only through field observations that the relative merits of different measurement methods can be assessed. The contributions of this team need to be recognized and valued by the agency. They are extraordinary.

My Assessment: I feel that this document focuses primarily on the strengths of the proposed methodology while inadequately acknowledging its weaknesses. Strengths include:

- -Proven technology;
- -Direct gravimetric measurement of mass;
- -Uses existing FRM equipment (minimizes equipment and training costs);
- -Presence of sub-2.5 µm particles causes coarse particles to adhere to filter;
- -Field tests show that coarse mass measured with dichot = coarse mass measured with proposed methodology, which suggests that evaporative losses are minimal;
- -Consistency with historical database of mass measurement avoids the need for expensive field comparisons;
- -Measurements are precise, and measured coarse mass concentrations are positive even when mass concentrations are low;
- -Low face velocities keep evaporative losses low.

Weaknesses include:

-Unknown measurement *accuracy* due to positive and negative adsorption on filter;

-Likelihood that measurement artifacts will be different for the $PM_{2.5}$ and PM_{10} filters because (1) reactions between fine and coarse PM species will affect the volatility of nitrates and other compounds, (2) evaporative losses on $PM_{2.5}$ will exceed evaporative losses on PM_{10} due to the pressure drop provided by the WINS or cyclone. Is this one reason that PM_{10} was systematically higher than $PM_{2.5}$ in field trials?;

-Sampling artifacts lead to inaccuracies that cannot be quantified. Therefore, measurement accuracy is unknown. (does this lead to data that have "a high degree of fidelity and faithfulness?);

-The extent to which find particles helps coarse particles adhere to filters is not quantified. While it is plausible that such adhesion should occur, it is likely to depend on mass collected and on particle composition, and these effects are not discussed, although it became clear in our meeting that empirical evidence supporting the problem with coarse particle losses from dichot data is available;

-inherently poor time resolution (24 hour) with no possibility for use in forecasting; and

-expensive manual operation.

In summary, the recommended choice is pragmatic and precise but its *accuracy* is unknown. My views have not changed since I wrote my preliminary evaluation on July 23, 2004. I would probably not be enthusiastic about any filter-based method, although I would conceptually prefer a dichotomous sampler to the proposed PM_{10}-$PM_{2.5}$ approach. Nevertheless, it is not clear to me that benefits of using a dichotomous sampler would justify the additional funds required to equip sampling stations and to train operators. If the proposed methodology is adopted, it should be done so with the understanding that it should be replaced when more suitable methodologies become available. Furthermore, work should be continued on the development of *accurate* techniques for measurements of coarse particle mass concentrations.

Peter H. McMurry
September 28, 2005

RE: CASAC Ambient Air Monitoring & Methods (AAMM) Subcommittee Meeting: Comments on Materials provided for review

Attachment 4: Criteria for Designating $PM_{2.5}$ Equivalence

A statistical methodology for determining whether a measurement method is "equivalent" to the PM 2.5 FRM is proposed. The statistical methodology involves the analysis of data acquired with 3 to 5 "equivalent" and 3 FRM co-located samplers. Four statistical measures must be satisfied to meet the equivalency designation. These include:

(1) Precision: must not exceed 15%;
(2) Correlation with FRM data (r): must range from 0.93 to 0.95;
(3) Multiplicative Bias (b) (slope of linear least squares fit of equivalent sampler to FRM data): Must range from 0.9 to 1.1;
(4) Additive bias (a) (intercept of linear least squares fit of equivalent sampler to FRM data): must fall between 15.05-17.31*b and 15.05-13.20*b. (-3.99 to 0.53 $\mu g/m^3$ for b=1.1 and –0.53 to 3.17 $\mu g/m^3$ for b=0.9).

My comments: The focus of these criteria is the ability to duplicate data acquired by the FRM, regardless of its faults or inaccuracies. The aim is to identify samplers that can be used for compliance measurements, not necessarily for *accurate* measurements. It is entirely conceivable that an alternative sampler could measure $PM_{2.5}$ with high precision and accuracy but would be excluded because, for example, it was able to detect semivolatile compounds that are not sampled effectively with the FRM. Such a sampler might have "r" values below the designated values, and might produce "a" and "b" values that differ significantly with season and from one location to the next. It is also possible that some $PM_{2.5}$ health and other effects are associated with those semivolatile compounds.

It would appear that the law and our legal system are constraining us to invest most of our energy and financial resources into developing and certifying samplers that reproduce FRM results. The proposed criteria will probably achieve this goal for most practical situations. Furthermore, these equivalent methods may offer significant advantages, such as high time resolution, automated data collection, real-time measurements, etc. It would appear to me, however, that this proposal does not address all issues that pertain to improved measurement accuracy for $PM_{2.5}$.

I feel that more thought ought to be given to comparing the responses of alternative samplers with the FRM using laboratory generated-aerosols of known composition and size or size distribution. Such work could include sampling of known semivolatile compounds, such as ammonium sulfate and selected organic compounds. While this methodology would not be applicable to equivalency determinations as specified by law, they could take us a long way towards an understanding of measurement accuracy. Because the atmospheric aerosol is so complex, there are many processes that could lead to discrepancies when samplers are used for

atmospheric sampling, even though the samplers operate with identical inlet characteristics. The laboratory tests would, enable unambiguous testing of sampler performance to particles having known physical and chemical properties. This approach would help us improve our understanding of measurement accuracy, and would lead to the design of improved samplers in the future.

I am also concerned about EPA's narrow focus on FRM measurements. There is clearly a need for more sophisticated measurements designed to support epidemiology studies at selected sites. Such measurements would also help to refine our understanding of emissions control measures that are currently being implemented, and would help with the development and evaluation of process models for atmospheric aerosols. Substantial progress was made during the supersite program to develop such measurement methodologies and to demonstrate ways in which they can be used most effectively. It is not clear to me that EPA is working systematically to build on what was learned over the past decade, and I regard this as very unfortunate and shortsighted. Examples of measurements that can be carried out routinely and continuously include particle size distributions and size-resolved composition.

Dr. Kimberly Prather

October 1, 2005

To: Fred Butterfield, Designated Federal Officer
Clean Air Scientific Advisory Committee (CASAC)
Rich Poirot, Co-Chair – Monitoring
Barbara Zielinska, Co-Chair – Methods
CASAC Ambient Air Monitoring and Methods (AAMM) Subcommittee

From: Kim Prather

Subject: CASAC Review of the Particle Methods and Data Quality Objectives

In general, the documents provide an excellent description of the plan for establishing a $PM_{10-2.5}$ Federal Reference Method (FRM). However, based on the descriptions provided, it appears the main reason for choosing the difference method is because it is the easiest to implement given the current suite of instruments located at current sampling sites. Is this reason enough? This is an important question to address up front as it is one EPA will no doubt have to defend once the new standard is implemented. It would be helpful if support and further justification were given (i.e. health data) that led EPA to make this choice.

Peer Review Questions:

3. What are the scientific and operational strengths and weaknesses of the $PM_{10-2.5}$ difference method relative to other options for a proposed FRM, especially when used as the basis for approval of other methods?

The strengths are well laid out. The weaknesses are not. The difference method's major strength is that it will use currently existing instrumentation already deployed to support the current $PM_{2.5}$ FRM. It will allow comparison with previous FRM measurements; at the same time, this is also it's major weakness in that the main basis for choosing the new sampling methodologies is that they must agree with a technique (FRM) that has been shown to have serious flaws (i.e. RH effects, volatilization losses). Yes, it will offer consistency (precision); but is a consistently incorrect answer the one we want? Accuracy should be considered as well and in order to make this a requirement, an effort needs to be made to create lab standards which can be used for validating potential FRM methods. We really need to think towards the future (as much as possible) and envision telling the public the reasons for going this route: right now we would have to say, we chose the new standard because we have these methods for measuring PM with high precision that we have used for a long time. Yes, we know the results they provide are not really representative of PM in the atmosphere, we're not really sure how the numbers are correlated with health effects, but they are what we currently have and we have invested lots of resources in them so we need to continue to use them.

Other major drawbacks include high analysis costs, requirement of skilled personnel to operate and perform filter handling protocols which will require "experts", and limited (24 hour) time resolution. PM, particularly in the coarse size range, show concentration excursions on timescales as short as minutes and as long as a couple of hours; these will be completely missed (averaged out) by the proposed 24 hour sampling times. See figure at end of report as an example. This figure shows data taken in Rubidoux CA where PM_{10} concentrations repeatedly got as high as 158 $\mu g/m^3$ at 6-7 am for 1.5 hours every day. However, the 24 hour average PM10 concentrations for these dates of 60 $\mu g/m^3$ were well within the proposed limit for the standard. These seem like high excursions that could potentially affect health and they would be missed. However, without a better understanding of the health effects of PM, is it wise to choose an FRM protocol which completely misses such concentration variations? Also, better time resolution often offers insights into the source/s leading to high concentrations of PM. With 24 hour samples, one will get a single number which will not tell us anything about the PM causing the problem. Also, the errors associated with the difference method could be substantial, particularly for chemical analysis (this has not even begun to be explored).

4. Based on the field study report as well as any other available data, e.g., data from State and local agencies, how does the demonstrated data quality of the $PM_{10-2.5}$ difference method support or detract from it being proposed as a FRM?

Choosing a method based on equivalency to the FRM has potential problems as stated. It still isn't clear why these particular size cuts were chosen other than "convenience". Are there health studies that support separating PM into these 2 size ranges? To rule out a new technique as cited in the report because it deviates from the FRM is dangerous. The newer instrument may be providing the "right" answer. Furthermore, when the results of one technique deviate from the FRM, this could be telling one something about chemical differences. Maybe the deviations are due to problems with the FRM instead of the method it is being compared against. By forcing a technique to be "equivalent" to the FRM, you're forcing it to have the same biases. This is quite problematic.

Even though the field studies reported in Appendix 2 were carefully thought out and conducted, they only represent 3 data points. One should think about expanding these sites to include other seasons. For example, Riverside in summer has low nitrate concentrations and less "issues" with SVOC since it is so hot. Going back to Riverside in the Fall (November) and comparing the results should be highly informative.

The peripheral data obtained in the field studies (APS size distributions, chemical speciation) need to be exploited to understand discrepancies between methods (see notes at end). As discussed in the meeting, it would be helpful to complement the field studies conducted to date with lab studies using standard particles of known size and composition to test new methods that are being proposed. More effort needs to be put into choosing an FRM that will provide accurate answers on PM (as well as precise).

They go into huge detail about not choosing proprietary samplers. This seems like weak logic and one that shouldn't be used. We need to step back and ask ourselves the goals of this exercise.

Consultation Questions:

9. What are the Subcommittee's views on the Very Sharp Cut Cyclone (VSCC) being approved as an alternate second-stage impactor to the WINS for use on a $PM_{2.5}$ FRM?

This change seems justified.

2. To what extent are the stated advantages of relaxing existing requirements identified for the $PM_{2.5}$ FRM supported by the information cited in Attachment 3, available literature, or good field and laboratory practices? Does the Subcommittee have additional recommendations for the $PM_{2.5}$ FRM that would be neutral with respect to bias, but would improve the performance and minimize the burden on agencies conducting the sampling?

No evidence is supplied to show why it is necessary to maintain 4 deg. C temperatures. There will be a serious cost associated with it, so it needs to be better justified.

5. To what extent have the assessments of spatial variability and the sensitivity of the DQO process to a variety of population distributions been appropriately addressed?

Real ambient data should be used to test the DQO. There are many factors that will affect this range that have yet to be addressed. The current analysis is too simplistic and needs to be expanded. There are large amounts of data available to do this from multiple sites, locations, and extremes. One doesn't have to be "hypothetical".

6. What are the Subcommittee's views on the approach identified for the development of criteria to approve continuous $PM_{10-2.5}$ equivalent methods?

Not enough detail is given here. To judge this, it would be helpful to know how many sites and seasons will be studied.

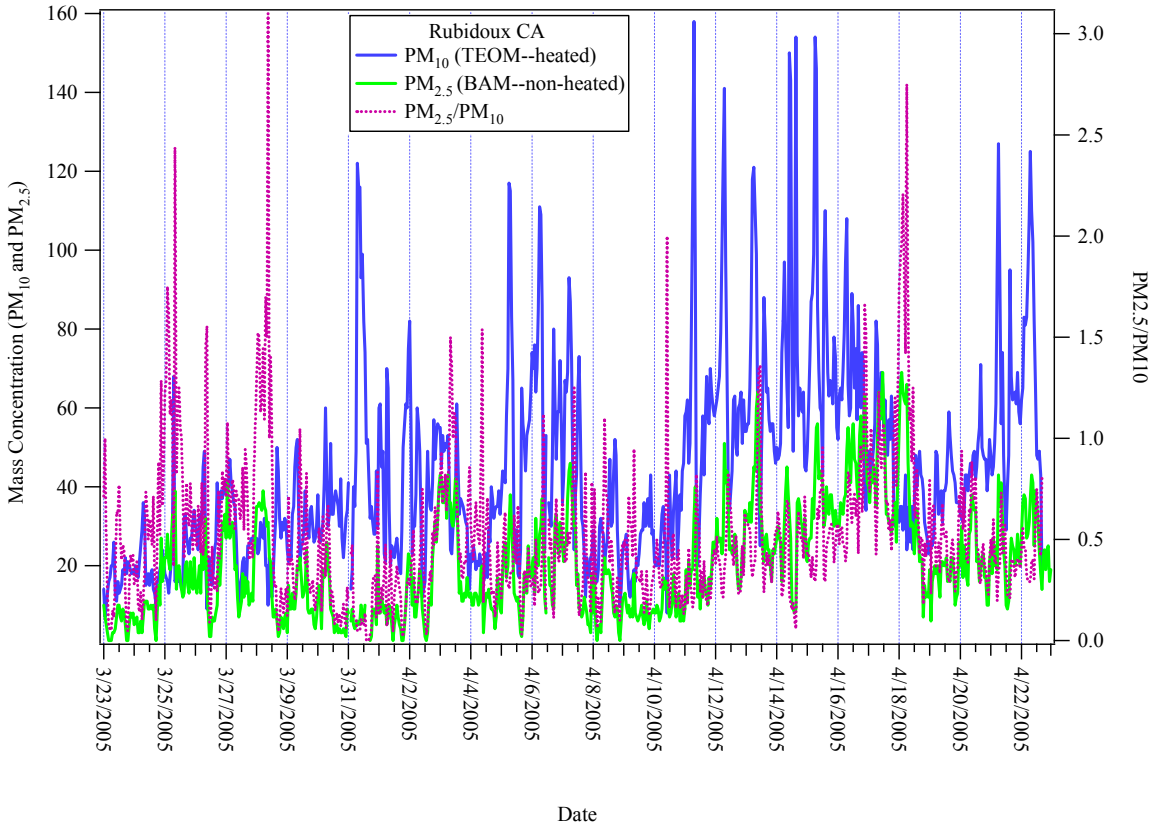
Summary:

1. As pointed out, the purpose of having the APS located at the sites during the study was not to measure FRM mass, but to understand how changes in the size distributions could play a role in measured differences in the FRM methods being tested. In the December 2004 meeting, we asked why EPA was using a PM10 cut on the front of the APS and basically throwing away information that could be used to understand discrepancies. They indicated they would take it off for future testing but this hasn't happened. It would be really straightforward to use software to add a cut-point (sharp, 50:50, etc.) and directly measure the effect on $PM_{2.5}$, PM_{10}, and $PM_{10-2.5}$. Then the issues of "bias", "error", and "zero values" could be directly addressed using real size distribution data acquired at the actual sampling locations. Also, "events" where the data from 2 techniques deviate from one another could be explored further using the measured size distributions to understand if it is playing a role.
2. Along the same line, one of the reasons for urging EPA to collect speciation data was so when 2 methods didn't agree with each other and/or the FRM, the composition at the different sites might be used to shed some light on this. I was quite surprised when

reviewing the documents for the meeting, there was not a single piece of information given on composition differences between the sites. All we were shown were time series of 24 hour mass concentrations collected by multiple instruments. Impressive data were acquired and great comparisons were done, but very little analysis and interpretation of the data was done. Now we are being asked to choose a method—but we are missing key information that would help us decide.

Suggestions:

1) In the current tests being conducted in Birmingham, EPA should ask TSI to deploy another APS without a PM10 cut point and use this alongside the other instruments for the entire study. This could provide key insight into the data and the size distribution data could be used to address "bias" concerns regarding cut points.
2) Someone should begin looking at the chemical speciation data ASAP. Without this, it would be very difficult to make a fully educated decision on the new FRM. EPA has done a series of excellent field studies to date; however, it appears most of the work has gone into collecting data and measuring differences and very little has gone into understanding these differences.
3) Choices of sites for determining "equivalency": the details on this are lacking in the report. This was purposely left as open to allow people flexibility in using new techniques. However, some amount of guidance and restrictions needs to be made or this could result in open-ended testing. More hardline guidance needs to be given: perhaps stating the instruments need to be tested in a region that has a certain range of coarse concentrations (giving a range of low to high values), a specific range of $PM_{2.5}/PM_{10}$, etc. For example, the state of CA as well as most other states with high PM levels have long historical data on PM mass concentrations for $PM_{2.5}$ and PM_{10} obtained with TEOM's and beta attenuation monitors (which don't use heating), etc. These can be studied for different seasons in previous years to better understand measurement challenges and set realistic time and locations for sampling and establishing equivalency. The figure below shows mass concentration measurements for PM10 and PM2.5 acquired in Rubidoux CA in spring 2005 generated from the AQMD web site. One can see the range of variations of coarse PM, and how the relative amount of $PM_{2.5}/PM_{10}$ varies over time. Such information can be used when choosing locations for equivalency testing.


Rubidoux CA
PM10 (TEOM--heated)
PM2.5 (BAM--non-heated)
PM2.5/PM10
Mass Concentration (PM10 and PM2.5)
PM2.5/PM10
Date
160
140
120
100
80
60
40
20
3.0
2.5
2.0
1.5
1.0
0.5
0.0
3/23/2005
3/25/2005
3/27/2005
3/29/2005
3/31/2005
4/2/2005
4/4/2005
4/6/2005
4/8/2005
4/10/2005
4/12/2005
4/14/2005
4/16/2005
4/18/2005
4/20/2005
4/22/2005

Dr. Armistead (Ted) Russell

September 19, 2005

To: Fred Butterfield, Designated Federal Officer
Clean Air Scientific Advisory Committee (CASAC)
CASAC Ambient Air Monitoring and Methods (AAMM) Subcommittee
From: Ted Russell

Subject: CASAC Review of the Coarse Particle Methods and Data Quality Objectives

This memo provides the comments by Ted Russell on two documents: Summary and Rationale for the $PM_{10-2.5}$ FRM") and ["Memo to PM NAAQS Review Docket (OAR-2001-0017): Potential Changes being Evaluated for the $PM_{2.5}$ FRM". It should be noted that I am primarily an air quality modeler and data analyst, so I am not particularly able to address the issues involved in the monitoring, *per se*, but more with the use the resulting data.

Summary and Rationale for the $PM_{10-2.5}$ FRM

Question 1: What are the scientific and operational strengths and weaknesses of the $PM_{10-2.5}$ difference method relative to other options for a proposed FRM, especially when used as the basis for approval of other methods?

In this regard, I am primarily concerned with the possible decision to use a method (i.e., the proposed approach) that will provide very little information for data analysis and model evaluation. A continuous method would be greatly preferred. At present, there appears to be an overwhelming desire to continue to use filter-based sampling for the FRM. The main reason to use a filter, as I see it, is to do speciation, for which there are no current sub-standards. A continuous method will provide more data and can also capture 24-hour exceedences that would otherwise get split between two days. The increased operating expenses and error introduction is also an issue in using a filter-based method. Looking down the road, if the real health impact is due to a more acute exposure, say over an hour, you need to have the continuous monitors to help identify the impacts and also for future regulation. We have determined that shorter term standards are appropriate for gas phase species (in part, since the method to measure them at shorter time scales were there), and I see no real reason it might not turn out that some of the effects of PM exposure (coarse, fine or whatever) are also due to shorter term events. All of the methods proposed, to date, have biases and artifacts, so use a method(s) that provide additional information.

Memo to PM NAAQS Review Docket (OAR-2001-0017): Potential Changes being Evaluated for the $PM_{2.5}$ FRM

Consultation Question 2: What are the Subcommittee's views on the Very Sharp Cut Cyclone (VSCC) being approved as an alternate, second-stage impactor to the Well Impactor Ninety-Six (WINS) for use on a $PM_{2.5}$ FRM?

Having run a couple of sampling systems, I prefer operating a cyclone vs. the WINS impactor (my system uses both). My understanding is that the differences in the results are very minor, so I would be prone to using a VSCC.

Consultation Question 3: To what extent are the stated advantages of relaxing existing requirements identified for the $PM_{2.5}$ FRM supported by the information cited in Attachment 3, available literature, or good field and laboratory practices? Does the Subcommittee have additional recommendations for the $PM_{2.5}$ FRM that would be neutral with respect to bias, but would improve the performance and minimize the burden on agencies conducting the sampling?

My interpretation of the current results from the continuous monitors suggests that they can provide reliable results. (Outside of my area.)

Dr. Jay Turner

Comments in Response to Attachment #6: *$PM_{10-2.5}$ Method Equivalency Development*
Submitted by Jay R. Turner (September 19, 2005)

Background. The overarching philosophy set forth in this attachment is "for making comparisons to the NAAQS, the decision quality [from the equivalent method] is as good as if they had used filter-based reference methods". This statement is operationalized by matching the upper- and lower-bounds for the gray zone of the decision performance curve for the candidate equivalent method to the respective bounds for the gray zone of the decision performance curve for the DQO. The equivalency criteria are assumed to have a fixed form (i.e. specified thresholds/ranges for precision, correlation, multiplicative bias, and additive bias) and will be developed based on the aforementioned matching. For equivalent methods which provide both high sampling frequency and high data completeness (e.g. continuous monitors) the equivalency criteria can actually be relaxed compared to the criteria embedded in the DQO.

Consultation Question. What are the Subcommittee's views on the approach identified for the development of criteria to approve continuous $PM_{10-2.5}$ equivalent methods?

Comments. To the extent the approach is presented at this time, it does appear reasonable. I like the explicit linkage between the DQOs and the equivalency criteria. My comments focus on what is not presented – how the results from this DQO/equivalency matching are used to actually arrive at the equivalency criteria.

1. Section 2.1 states "data collection will be replicated at multiple sites to ensure the sampling is representative of different aerosol types […]". It will be important to provide additional guidance for site selection. Not only should the aerosol types be different but rather both the aerosol types and environmental conditions should capture those conditions which are most likely to challenge the candidate method. Such conditions may be different for instruments which are based on different operating principles. There will also be issues concerning the on-site physical location of instruments – ambient conditions versus being housed in HVAC-controlled environmental shelters – which will need to be addressed in deciding what deployment characteristics actually constitute the equivalency test conditions.
2. In the proposed approach, the equivalency criteria are coupled. The example presented in Attachment 6 shows how starting with a prescribed precision and multiplicative bias one can bound the criteria for correlation and additive bias. Indeed, $PM_{2.5}$ equivalency criteria presented in Attachment 4 include an additive bias criterion which is functionally dependent on the multiplicative bias. It is not clear whether the approach presented for the $PM_{10-2.5}$ equivalency criteria includes solely fixed forms or relational forms (like the $PM_{2.5}$ equivalency criteria of Attachment 4). The text states (Section 2.0) "a fixed form for the equivalency requirements was assumed" but later states "the entire process follows that [$PM_{2.5}$ equivalency requirements] development". Relational forms provide the greatest flexibility towards meeting the equivalency requirements whereas fixed forms place additional – and potentially important – constraints. The example in Attachment 6 is quite relevant to this discussion. Figure 4 shows the acceptances ranges for additive bias and multiplicative bias for the example presented. Following

Attachment 4, a relational criterion for the additive bias would use the upper and lower lines of the parallelogram to bracket the allowable range for additive bias for a given (observed) multiplicative bias. For this particular example - which is presumably realistic (note the assumed daily standard of 60 $\mu g/m^3$ and 98th percentile falls in the range set forth in the EPA Staff Paper) the acceptable ranges for the additive bias at the extreme values for the multiplicative bias are quite large (-7 to +22 $\mu g/m^3$ at -15%, -31 to + 9 $\mu g/m^3$ at +15%). The text subsequently states that for this example a candidate monitor additive bias criterion of ± 5 $\mu g/m^3$ would be reasonable. This implies a rectangle would be drawn within the parallelogram which would define a fixed form for the additive bias, unlike the relational form set forth in Attachment 4 for $PM_{2.5}$ equivalency.[1] The decision on whether to use a fixed form or relational form could well be very important in terms of what is deemed acceptable instrument performance. If the DQO gray zone is very broad (as is the case with this example), at some point common sense dictates that a monitor exhibiting -31 $\mu g/m^3$ additive bias (lower-right corner of the parallelogram in Figure 4) should not be acceptable for a standard of 60 $\mu g/m^3$ even though it falls within the acceptance range for a relational form of the equivalency requirements.

3. As a follow-up to comment #2, *each candidate sampler in the equivalency testing should meet the defined equivalency criteria* (rather than the mean performance over multiple samplers meeting the defined equivalency criteria). The actual deployments for compliance monitoring will most certainly not have the luxury of collocated, identical samplers and thus the equivalency testing should be consistent with the real scenario. Since equivalency designations are based on performance specifications rather than design specifications, each individual instrument must be able to meet the equivalency requirements.
4. If a fixed form for the equivalency criteria is chosen, how does one get started in using the approach to generate the equivalency requirements? That is, will precision and multiplicative bias from current/emerging monitors be used to set these parameters with the correlation and additive bias criteria developed accordingly?
5. While perhaps modestly off-topic, I am curious how post-deployment performance evaluations of continuous, equivalent methods would be conducted. For filter-based PM measurements the flow rate is routinely checked in the field and the analyzer zero and span (in this case, for the microbalance) are routinely checked in the laboratory. Thus, this approach is analogous to on-line gas measurements where the analyzer is in the field but still routinely subjected to performance texts. For continuous PM monitors, we can check flow rates and analyzer zeros but we don't have a convenient, through-the-probe method for routinely checking the "span". We can monitor parameters which indirectly give us information on the performance of subcomponents of the measurement but this is different from an actual span which would require a through-the-probe aerosol. Thus, there is a need for bench testing (prior to initial deployment) and/or acceptance testing (immediately upon deployment) of each instrument as well as routine performance checks. The bench/acceptance testing could be approached by aerosol challenge tests in

[1] I enjoyed running version Beta 2 of the DQO Companion for PM_{coarse} (downloaded from http://www.epa.gov/ttnamti1/dqotool.html) and look forward to getting access to the newer version which, according to Attachment 6, can generate the acceptance range plots as well as the decision performance curves.

the laboratory; in the field, it could include collocation with a continuous monitor with documented performance characteristics and/or collocation with filter-based measurements. The routine field testing (audits) will likely require collocation with filter-based measurements (especially in the case where the continuous monitor requires an environmental shelter with HVAC controls which would complicate installation of the audit monitor). The bar is potentially much higher than conventional audits of filter-based samplers because there is no routine performance check on the continuous monitor's "span". Logistically this could be quite challenging and I look forward to learning how (methods and frequency) the field performance of deployed continuous monitors will be determined.

Again, these comments do not draw into question the overall approach of making the equivalency criteria consistent with the DQOs nor does it question the specific DQO tool being used to study the interactions between acceptance criteria. These comments are addressed at how the information would subsequently be used to actually arrive at the criteria and whether there should be some level of performance testing for each instrument deployed beyond the equivalency testing on a few instruments of that make and model.

In closing, even though it is beyond the scope of the charge question I am compelled to comment on the example provided in this attachment. To the extent the DQO example does represent a realistic worst-case scenario, the gray zone for the example 24-hour standard (which is within the range stated in the EPA Staff Paper) is very large! The difference between the lower bound and the standard is 37% of the standard and the difference between the upper bound and the standard is 59% of the standard. It will be interesting to observe what DQOs are actually established for $PM_{10-2.5}$ and how this will subsequently impact the development of the equivalency requirements.

Comments in Response to Attachment #1: Summary and Rationale for *$PM_{10-2.5}$ FRM*
Submitted by Jay R. Turner (September 19, 2005)

Background. A $PM_{10-2.5}$ Federal Reference Method is proposed which is based on the PM_{10} mass concentration minus the $PM_{2.5}$ mass concentration as determined by independent measurements.

Peer Review Questions. (1) What are the scientific and operational strengths and weaknesses of the $PM_{10-2.5}$ difference method relative to other options for a proposed FRM, especially when used as the basis for approval of other methods? (2) Based on the field study report as well as other available data, e.g., data from State and local agencies, how does the demonstrated data quality of the $PM_{10-2.5}$ difference method support or detract from it being proposed as a FRM?

Comments.

1. A compelling case is made for a filter-based FRM (be it the proposed difference method of a virtual impactor) compared to a continuous monitor being designated as an FRM.
2. The virtual impactor (compared to the difference method) offers no advantage in terms of analysis costs since the fine channel ($PM_{2.5}$) mass concentration must be measured to correct the coarse ($PM_{10-2.5}$) channel mass. That said, there may be advantages to the virtual impactor approach in terms of chemical analysis of the coarse fraction, especially for species which would have large propagated uncertainties using the difference method. The overwhelming drawback to the virtual impactor appears to be the presumed high potential for particle losses during sample handling. Are there any contemporary studies that have investigated this issue? I recall the field evaluations described in Attachment 2 used on-site gravimetric measurements? Were some of the coarse particle filters also shipped to a laboratory for analysis to mimic typical handling procedures for a routine deployment by a State or local agency? In any event, I would be very interested in seeing the evidence which demonstrates that $PM_{10-2.5}$ - laden filters are susceptible to particle losses during shipment; this is certainly the conventional wisdom but I have never seen the data.
3. The proposed advantage of the FRM providing "readily accessible aerosol samples for subsequent chemical analysis" is overstated. Following the example of the $PM_{2.5}$ FRM and $PM_{2.5}$ speciation sampler, one would need *six samplers* to characterize the aerosol (PM_{10} with quartz filter, $PM_{2.5}$ with quartz filter, PM_{10} with Teflon filter, $PM_{2.5}$ with Teflon filter, PM_{10} with Nylon filter, $PM_{2.5}$ with Nylon filter).
4. While EPA anticipated a shift to continuous monitors as such instruments obtain equivalency status, the extent to which routine on-site performance evaluations will be required is not clear. Given there are no tractable tests of the true instrument "span" for such monitors, such testing goes beyond the role of a traditional audit in that it is the sole ongoing evaluation of the instrument span. These evaluations will most likely require filter-based measurements. If the required frequency is high, the intrinsic value of deploying continuous monitors will be diminished.

Dr. Warren H. White

ATTACHMENT 1: Summary and Rationale for the $PM_{10-2.5}$ FRM

Revised comments by Warren H. White, 9/26/05

In its September 15 letter to the Administrator on EPA staff recommendations for a thoracic coarse PM standard, CASAC observed (http://www.epa.gov/sab/pdf/sab-casac-05-012.pdf, page 6) that

> ***By use of the indicator $UPM_{10-2.5}$, the Agency is taking a next step toward including composition as well as size in its regulations of ambient air PM.***

Let us not forget that the designation of "urban" $PM_{10-2.5}$ as the "indicator species" was presented – and accepted by most of the Committee – as a surrogate for an as-yet-undefined composition.

At our September 21-22 meeting, Agency staff presented an effective case for the difference method as a reference against which dichotomous samplers and continuous $PM_{10-2.5}$ monitors could subsequently be tested. Alone among candidate methods, the difference measurement is directly comparable with $PM_{2.5}$ measurement and already available in final form. Beyond these logistical advantages, Agency emphasis focused on the demonstrated precision of the difference as a repeatable measurement of $PM_{10-2.5}$. I am left, however, with three concerns that have not yet been addressed to my satisfaction:

1. **The proposed FRM has demonstrated its precision for $PM_{10-2.5}$, but there is absolutely no evidence that it provides a repeatable measurement of "those thoracic coarse particles that are generally present in urban environments."** The Staff Paper (page 5-56), for example, specifically highlights episodes of high urban concentrations of $PM_{10-2.5}$ in Spokane WA that it concludes did not represent elevated $UPM_{10-2.5}$ levels. In other words, it reports that days with comparable $UPM_{10-2.5}$ concentrations would have produced disparate FRM readings. As a measurement of the proposed NAAQS ***indicator***, the proposed FRM is wildly imprecise.

2. **The proposed FRM will not meet the planned data quality objective for $UPM_{10-2.5}$.** The DQO will require daily monitoring, which is impractical with the FRM. Isn't it a bit … well, awkward, that the Federal Reference Method itself can't deliver data of the quality needed for decision-making? Examples were given at our meeting of other FRMs that are not widely deployed, but are any of those incapable of satisfying their DQOs?

3. **The proposed FRM is incompatible with the historical rationale for filter-based PM measurements.** The Agency has always favored filter sampling as the basis of its reference methods for PM on the ground the collected deposit is then available for potential chemical analysis or toxicological testing. This consideration carries particular weight in the case of $UPM_{10-2.5}$, where we are "taking a next step toward including composition" as noted above. Clearly, however, collection of $UPM_{10-2.5}$ for chemical and toxicological analyses is best done by virtual impaction to minimize contamination of the sample by fine particles.

The third point merits elaboration. The PM_{10c} sample is an undifferentiated mix of coarse and fine PM. Agency staff note that the composition and toxicity of the $PM_{10-2.5}$ portion can be inferred indirectly, by comparisons with the associated $PM_{2.5}$ sample. Consider, however, the Staff Paper's descriptions (page 5-57) contrasting "toxic" $UPM_{10-2.5}$,

> *resuspended crustal particles may be contaminated with toxic trace elements and other components from previously deposited fine PM, e.g., metals from smelters (Phoenix) or steel mills (Steubenville, Utah Valley), PAHs from automobile exhaust, or pesticides from agricultural lands. (CD, p. 8-344)*,

with "less toxic" $PM_{10-2.5}$:

> *particles of crustal origin … are relatively non-toxic under most circumstances,*

$UPM_{10-2.5}$, in other words, may just be ordinary crustal dust that has been contaminated by fine particulate matter! How can a method that necessarily contaminates $PM_{10-2.5}$ with fine PM, in the very process of collecting it, yield useful information on ambient $UPM_{10-2.5}$?

ATTACHMENT 4: Criteria for designation of equivalence methods for continuous surveillance of $PM_{2.5}$ ambient air quality

Comments by Warren H. White, 9/15/05

A lot of careful thought and analysis have obviously gone into this paper, and I am inclined to trust the authors' and Agency's practical judgments on what constitute appropriate equivalence tests. I don't feel like I fully understand its overall rationale, however, and it seems like it should be possible to lay that out more clearly. In particular, it's not clear to me how the proposed criteria connect to the $PM_{2.5}$ DQO.

The proposed criteria involve four measures: precision, correlation, multiplicative bias, and additive bias. Precision is the only one of these that gives any information on the repeatability of the candidate measurement. The remaining three all come from a standard bivariate regression of the daily ***means*** from 2-5 candidate measurements on those from 2-3 FRM measurements.

The question I have is: How relevant is (A) the performance of means from collocated testing to (B) meeting the DQO in actual deployments of individual monitors? The table below summarizes some exploratory calculations based on the simulated data listed in Table 1. I used the exact formulae given in the attachment, which yield results identical to those obtained from the standard Excel functions CORREL, SLOPE, and INTERCEPT. I assume the differences between the two right-most columns, comparing my bias results to their reported values (e.g. 1.060 vs. 1.058 for multiplicative bias), arise from truncations in the attachment's listing of the simulated data. The full data set is included on the next page as an embedded Excel spreadsheet so the calculations can be checked (in the electronic version of these comments) by double-clicking on any cell.

		individual candidate vs mean FRM				mean vs mean	
						WHW	Battelle
r		0.966	0.938	0.963		0.981	0.981
b		1.107	1.020	1.050		1.060	1.058
a		-2.249	-1.330	-1.568		-1.728	-1.692
a bound		-4.1	-2.6	-3.1		-3.3	-3.3
a bound		0.4	1.6	1.2		1.1	1.1

The three columns under the heading "individual candidate vs. mean FRM" show the correlation r (Eq. 11), multiplicative bias b (Eq. 13), and additive bias a (Eqs. 14-16) of the three individual candidate monitors relative to the mean daily FRM measurement. Should we be concerned that two out of three of the candidate monitors fall outside the proposed criteria when tested as individuals?

Table 1

	FRM samplers			candidates			FRM		candidates	
Run	#1	#2	#3	#1	#2	#3	Mean	CV	Mean	CV
1	6.4	5.8	6.6	5.5	5.0	3.7	6.3	0.066	4.7	0.196
2	6.9	6.4	7.2	4.4	6.5	4.3	6.8	0.059	5.1	0.245
3	5.6	5.9	5.6	4.5	3.5	4.3	5.7	0.030	4.1	0.129
4	4.4	4.4	4.5	2.7	2.4	2.2	4.4	0.013	2.4	0.103
5	7.9	8.9	8.4	7.5	6.0	7.6	8.4	0.060	7.0	0.127
6	8.0	7.7	8.2	7.2	5.8	6.3	8.0	0.032	6.4	0.110
7	9.0	10.3	9.1	5.4	10.4	8.3	9.5	0.076	8.0	0.313
8	16.9	18.6	15.8	16.2	17.8	15.6	17.1	0.082	16.5	0.069
9	10.3	10.2	11.5	9.6	10.8	7.6	10.7	0.068	9.3	0.173
10	11.1	11.7	11.6	10.9	8.6	12.5	11.5	0.028	10.7	0.184
11	10.9	10.7	10.5	11.9	8.6	10.1	10.7	0.019	10.2	0.162
12	11.2	10.9	11.0	9.1	9.5	12.0	11.0	0.014	10.2	0.154
13	10.2	10.2	10.2	9.2	8.5	8.4	10.2	0.000	8.7	0.050
14	5.3	5.6	5.5	3.6	4.2	2.8	5.5	0.028	3.5	0.199
15	7.8	8.8	8.5	5.0	6.7	5.5	8.4	0.061	5.7	0.152
16	5.6	6.4	6.2	5.3	4.5	5.3	6.1	0.069	5.0	0.092
17	11.3	10.2	10.8	7.8	8.7	9.3	10.8	0.051	8.6	0.088
18	7.0	7.9	7.1	6.3	4.5	6.9	7.3	0.067	5.9	0.212
19	6.6	6.0	6.0	4.7	5.4	4.0	6.2	0.056	4.7	0.149
20	6.1	6.3	6.4	5.7	5.4	4.3	6.3	0.024	5.1	0.144
21	6.7	6.3	7.3	4.3	5.6	7.6	6.8	0.074	5.8	0.285
22	9.8	9.1	9.4	7.4	7.7	10.3	9.4	0.037	8.5	0.188
23	12.5	13.5	14.1	15.4	12.4	10.3	13.4	0.060	12.7	0.202
24	16.4	16.1	15.0	15.0	18.1	19.4	15.8	0.047	17.5	0.129
25	26.7	22.2	24.4	21.2	20.8	20.0	24.4	0.092	20.7	0.030
26	12.5	10.5	10.0	10.8	10.1	11.0	11.0	0.120	10.6	0.044
27	15.6	15.0	14.9	11.5	16.1	13.9	15.2	0.025	13.8	0.166
28	20.8	22.2	20.4	20.8	17.3	19.9	21.1	0.045	19.3	0.094
29	19.7	20.3	20.0	22.8	17.6	22.4	20.0	0.015	20.9	0.138
30	5.4	5.2	5.3	3.6	3.1	4.3	5.3	0.019	3.7	0.164
31	7.0	7.3	6.9	4.8	7.2	5.9	7.1	0.029	6.0	0.201
32	17.1	14.9	16.1	15.9	17.4	15.5	16.0	0.069	16.3	0.062
33	12.5	12.6	11.3	10.9	8.4	11.5	12.1	0.060	10.3	0.160
34	9.7	10.1	10.1	9.1	9.6	7.9	10.0	0.023	8.9	0.099
35	14.8	15.4	16.3	13.4	14.6	14.1	15.5	0.049	14.0	0.043
36	19.4	19.7	19.8	18.5	15.9	18.5	19.6	0.011	17.6	0.085
37	17.1	15.7	17.0	14.0	18.1	18.2	16.6	0.047	16.8	0.143
38	14.1	14.2	14.0	17.2	16.9	14.0	14.1	0.007	16.0	0.110
39	11.6	10.8	11.0	10.2	8.1	10.6	11.1	0.037	9.6	0.139
40	12.5	12.7	13.6	11.3	12.7	10.6	12.9	0.045	11.5	0.093
41	11.1	11.6	12.0	10.8	8.5	11.9	11.6	0.039	10.4	0.167
42	14.2	14.3	14.9	13.4	13.6	14.6	14.5	0.026	13.9	0.046
43	22.8	23.5	20.3	24.3	17.9	22.0	22.2	0.076	21.4	0.151
44	14.4	17.3	16.6	17.1	19.7	15.9	16.1	0.094	17.6	0.111
45	18.2	18.9	17.1	19.7	22.2	17.0	18.1	0.050	19.6	0.132
46	17.2	20.6	19.5	19.5	17.5	17.0	19.1	0.091	18.0	0.073
mean	11.92	12.02	11.91	10.99	10.87	10.98	11.95		10.94	
RMS	5.17	5.17	4.92	5.78	5.49	5.50	5.04		5.45	

	individual candidate vs mean FRM			mean vs mean	
				WHW	Battelle
r	0.966	**0.938**	0.963	0.981	0.981
b	**1.107**	1.020	1.050	1.060	1.058
a	-2.249	-1.330	-1.568	-1.728	-1.692
a bound	-4.1	-2.6	-3.1	-3.3	-3.3
a bound	0.4	1.6	1.2	1.1	1.1

Editorial notes (both on page A-4):

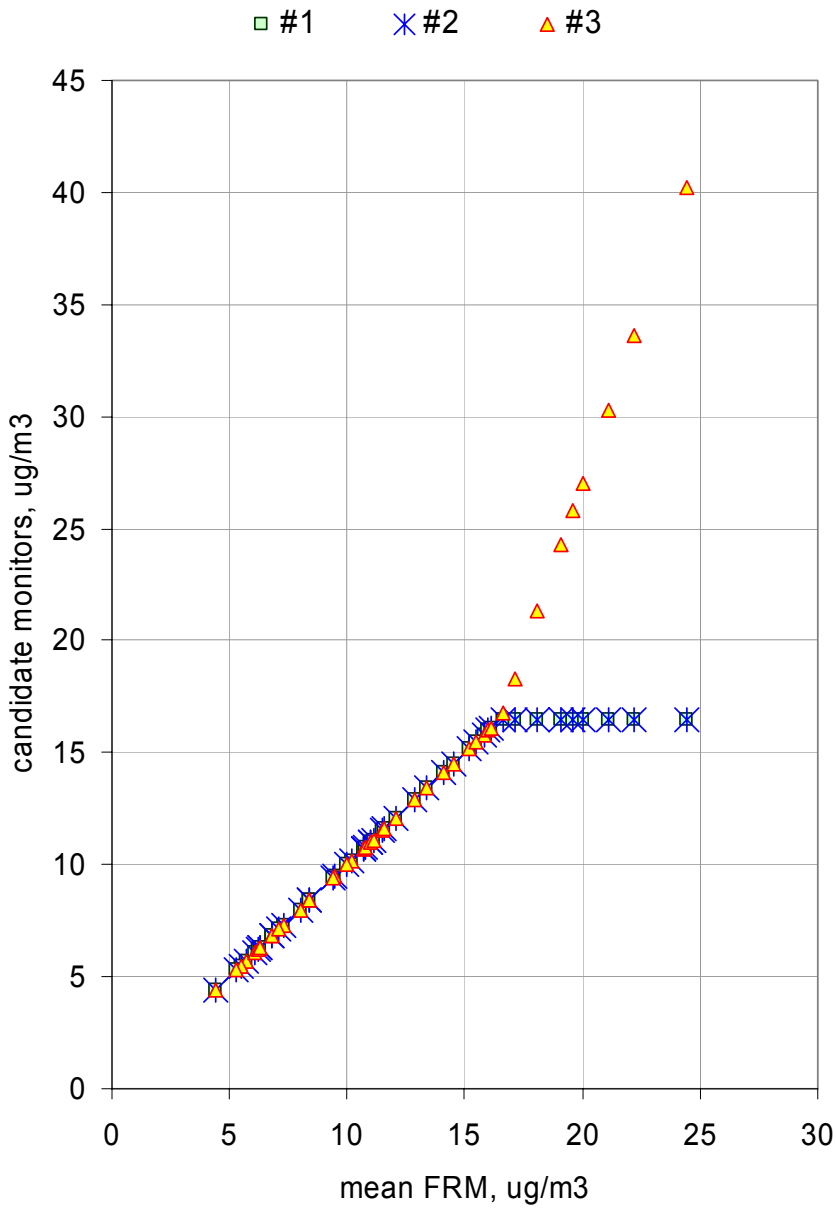
Equation A-1 is slightly garbled. A radical sign should replace the extra brackets on the right side of the denominator.

Five lines below the equation, I don't understand how Figure A-4 "shows … the relative insensitivity to the number of FRM samplers."

I would like to conclude by expressing appreciation for the care that went into this document. While it was a tough slog in places, I came to trust that the difficulty arose from the subject itself and not from any misdirection or carelessness by the authors. My compliments to them!

An example of a successful FEM demonstration. Warren H. White, 9/19/05

The accompanying plot shows (hypothetical) test data from three candidate monitors that would satisfy the FEM requirements proposed in Attachment 4.



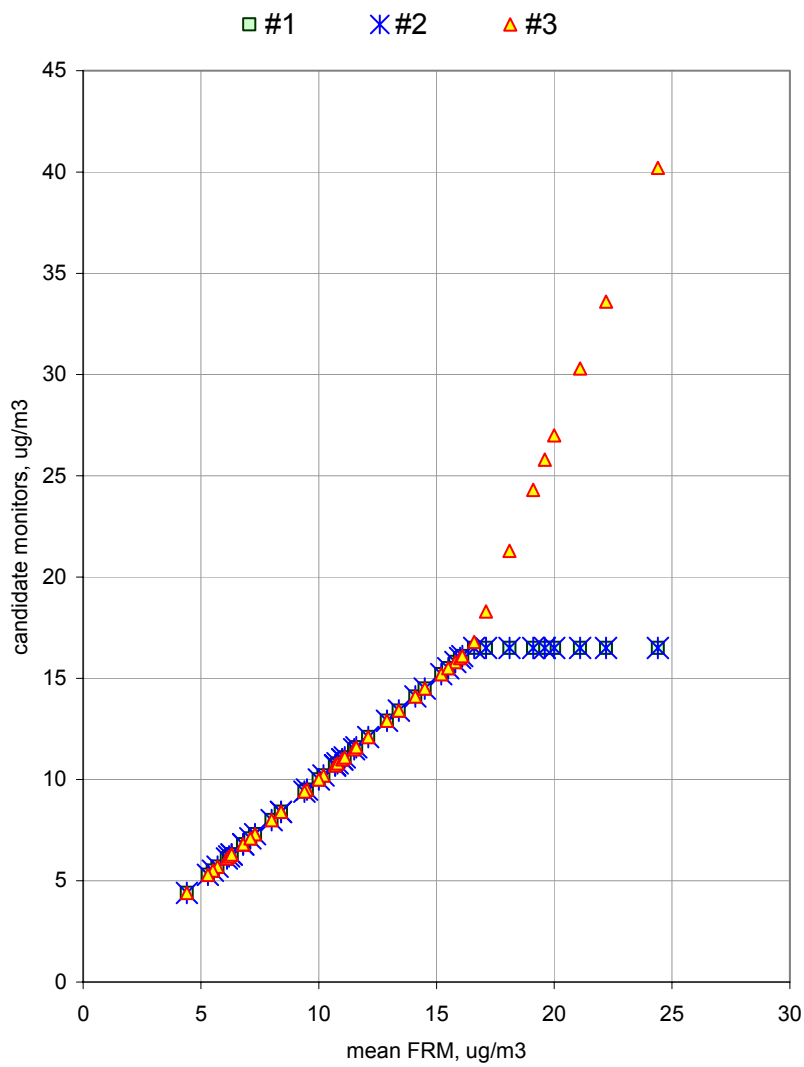
The FRM concentrations are exactly those listed in Table 1 of Attachment 4. All three candidate monitors are accurate (relative to the FRM) at concentrations below 16.5 $\mu g/m^3$, but #1 and #2 become "pegged" at 16.5 $\mu g/m^3$ when the true concentration goes higher. Fortunately for the manufacturer, candidate #3 is three times more sensitive than it should be to increments above $\mu g/m^3$, so the <u>mean</u> of the three candidates is still accurate at high concentrations. The marketing slogan is

"Buy three of our instruments and keep one in your garage!"

Here is an embedded Excel file showing the calculations. They can be verified (or played with) by double-clicking to get into the spreadsheet. Note that all four test statistics are better than those of the original example in Attachment 4: precision = 14%, correlation =1, and absolutely no multiplicative or additive bias.

maximum value for candidates 1 & 2: 16.5

	candidates			FRM		candidates	
Run	#1	#2	#3	Mean	CV	Mean	CV
1	6.3	6.3	6.3	6.3	0.066	6.3	0
2	6.8	6.8	6.8	6.8	0.059	6.8	0
3	5.7	5.7	5.7	5.7	0.030	5.7	0
4	4.4	4.4	4.4	4.4	0.013	4.4	0
5	8.4	8.4	8.4	8.4	0.060	8.4	0
6	8	8	8	8	0.032	8	0
7	9.5	9.5	9.5	9.5	0.076	9.5	0
8	16.5	16.5	18.3	17.1	0.082	17.1	0.061
9	10.7	10.7	10.7	10.7	0.068	10.7	0
10	11.5	11.5	11.5	11.5	0.028	11.5	0
11	10.7	10.7	10.7	10.7	0.019	10.7	0
12	11	11	11	11	0.014	11	0
13	10.2	10.2	10.2	10.2	0.000	10.2	0
14	5.5	5.5	5.5	5.5	0.028	5.5	0
15	8.4	8.4	8.4	8.4	0.061	8.4	0
16	6.1	6.1	6.1	6.1	0.069	6.1	0
17	10.8	10.8	10.8	10.8	0.051	10.8	0
18	7.3	7.3	7.3	7.3	0.067	7.3	0
19	6.2	6.2	6.2	6.2	0.056	6.2	0
20	6.3	6.3	6.3	6.3	0.024	6.3	0
21	6.8	6.8	6.8	6.8	0.074	6.8	0
22	9.4	9.4	9.4	9.4	0.037	9.4	0
23	13.4	13.4	13.4	13.4	0.060	13.4	0
24	15.8	15.8	15.8	15.8	0.047	15.8	0
25	16.5	16.5	40.2	24.4	0.092	24.4	0.561
26	11	11	11	11	0.120	11	0
27	15.2	15.2	15.2	15.2	0.025	15.2	0
28	16.5	16.5	30.3	21.1	0.045	21.1	0.378
29	16.5	16.5	27	20	0.015	20	0.303
30	5.3	5.3	5.3	5.3	0.019	5.3	0
31	7.1	7.1	7.1	7.1	0.029	7.1	0
32	16	16	16	16	0.069	16	0
33	12.1	12.1	12.1	12.1	0.060	12.1	0
34	10	10	10	10	0.023	10	0
35	15.5	15.5	15.5	15.5	0.049	15.5	0
36	16.5	16.5	25.8	19.6	0.011	19.6	0.274
37	16.5	16.5	16.8	16.6	0.047	16.6	0.010
38	14.1	14.1	14.1	14.1	0.007	14.1	0
39	11.1	11.1	11.1	11.1	0.037	11.1	0
40	12.9	12.9	12.9	12.9	0.045	12.9	0
41	11.6	11.6	11.6	11.6	0.039	11.6	0
42	14.5	14.5	14.5	14.5	0.026	14.5	0
43	16.5	16.5	33.6	22.2	0.076	22.2	0.445
44	16.1	16.1	16.1	16.1	0.094	16.1	0
45	16.5	16.5	21.3	18.1	0.050	18.1	0.153
46	16.5	16.5	24.3	19.1	0.091	19.1	0.236
		mean		11.95		11.95	
		RMS		5.04		5.04	



	FRM	candidates
precision	0.054	0.140
r	1.000	
b	1.000	
a	0.000	

Dr. Yousheng Zeng

Peer Review Comments on $PM_{10-2.5}$ FRM
By Yousheng Zeng
For Sept. 21-22, 2005 AAMM Subcommittee Peer Review and Consultation Meeting

Comments Revised after the Meeting

Peer Review Charge Question 1:
What are the scientific and operational strengths and weaknesses of the $PM_{10-2.5}$ difference method relative to other options for a proposed FRM, especially when used as the basis for approval of other methods?

Comment:

The $PM_{10-2.5}$ difference method has significant operational strengths as summarized in Attachment 1 in the provided review materials. I agree with EPA's analysis in Attachment 1 and consider the difference method superior to other options. Its major shortcoming is its inability to monitor continuously and automatically. Hopefully this shortcoming can be overcome by future equivalent continuous monitors that are more robust and more accurate than currently available continuous monitors. Eventually these equivalent continuous monitors, rather than the reference method, will be widely deployed in a similar fashion as continuous sulfur dioxide (SO_2) monitors in relation to the SO_2 FRM.

By definition, the difference method has to be the most accurate one among all candidate methods because all methods are compared against the PM metrics that is defined by the underlying methods of the difference method. In this sense, it has scientific strength *within the context of current PM metrics or definition*. However, from a fundamental scientific viewpoint, the difference method is fairly weak.

The current PM metrics is defined by sampling methods rather than nature of the pollutant (*i.e.*, PM). $PM_{2.5}$ is defined as the PM mass concentration measured by $PM_{2.5}$ FRM and PM_{10} by PM_{10} FRM. Therefore, it makes most sense to define $PM_{10-2.5}$ as the difference between the two. However, I would like to bring a different perspective into this discussion. Although I recognize that EPA is in the final stage of promulgating new PM standards and it is probably unrealistic to evaluate a new and fundamentally different approach in this round of the rulemaking process, I recommend that EPA evaluate it in the next cycle of the PM standard review.

We should rethink the conventional approach of defining PM by sampling methods. The PM defined in this way appears to be a very poor indicator of PM as a pollutant. This is not limited to $PM_{10-2.5}$. It applies to PM_{10} and $PM_{2.5}$ as well. The following examples demonstrate this point.

In these examples, I used simulated ambient PM samples similar to the ones specified in 40 CFR 53, Subpart F, Table F-3 and the fractionation curves of $PM_{2.5}$ FRM and PM_{10} FRM to calculate expected $PM_{2.5}$, PM_{10}, and $PM_{10-2.5}$ values as if the FRM samplers are used to collect these simulated samples. I also used the deposition curves of PM in human respiratory system in the EPA PM Criteria Document (CD), Oct. 2004 version, Figure 6-13, to see what portion of the PM will deposit onto the respiratory system if a person is inhaling the same ambient air. A more detailed description of this approach is expected to be available soon as a paper is currently under peer review.[1] An abbreviated description of the approach is provided as Appendix A to my comment.

Figure 1 includes one set of examples showing the changes of different PM metrics as the ambient PM condition changes (in this case, change of coarse mode PM concentration while fine mode holds constant, which causes change in $PM_{2.5}/PM_{10}$ ratio). This example (and many more not presented here due to space consideration) illustrate that currently used PM metrics ($PM_{2.5}$ and PM_{10}) and proposed $PM_{10-2.5}$ do not track the level of PM that can potentially deposit onto the human respiratory system. Therefore, they do not seem to be good indicators of PM as a pollutant. The scientific community is aware of the mismatch between what is measured and what can be deposited onto human respiratory system. The reason for the current PM metrics is due to lack of a system, including measurement metrics and corresponding method, that can accurately measure what can be deposited onto human respiratory system. A recently proposed system called Comprehensive Particulate Matter Measurement System (CPMMS) and corresponding dosimetry-based PM metrics have shown promising features in resolving this fundamental issue [1] (also see Appendix A). The results from CPMMS are also included in Figure 1 for comparison.

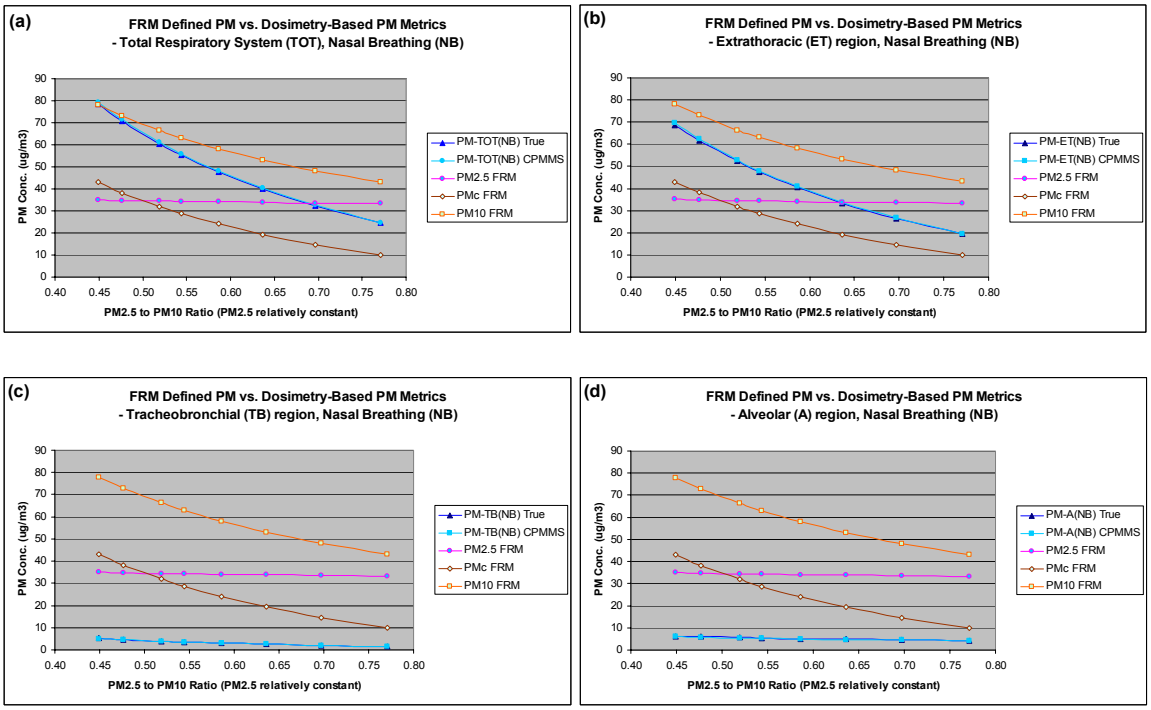


Figure 1. Changes of various PM metrics with changes in ambient $PM_{2.5}/PM_{10}$ ratio. The legend "PM-TOT(NB) True" in (a) stands for simulated true ambient PM that has potential to be deposited onto total respiratory system based on the deposition curve for nasal breathing in EPA PM CD, Figure 6-13, Oct. 2004. PM-TOT(NB) CPMMS stand for the same PM metrics obtained by a method called Comprehensive PM Measurement System.[1] Similar notations are used for different regions of the respiratory system, i.e., extrathoracic (ET) region, tracheobronchial (TB) region, or alveolar (A) region, in (b), (c), and (d), respectively.

Figure 2 is another example of how proposed $PM_{10-2.5}$ can be a misleading indicator for PM. Two ambient PM conditions, Case 1 and Case 2, are presented in Figure 2. The two cases differ in their size distribution [see Figure 2 (a) and (b)]. If the proposed $PM_{10-2.5}$ FRM is used to monitor the two ambient conditions, the $PM_{10-2.5}$ (as well as PM_{10} and $PM_{2.5}$) will be the same. However, the PM levels that correspond to the portions that have potential to be deposited onto human respiratory system [as a whole (*i.e.*, in TOT) or in extrathoracic (ET) region, tracheobronchial (TB) region, or alveolar (A) region] are significantly different between the two cases, especially in the ET region or the total respiratory system, which is the region that $PM_{10-2.5}$ intend to cover. In this example, $PM_{10-2.5}$ is about 32 $\mu g/m^3$ for both cases; but the real health effect [represented by PM-ET(NB)] is equivalent to PM level of about 52 $\mu g/m^3$ for case 1 and about 78 $\mu g/m^3$ for case 2.

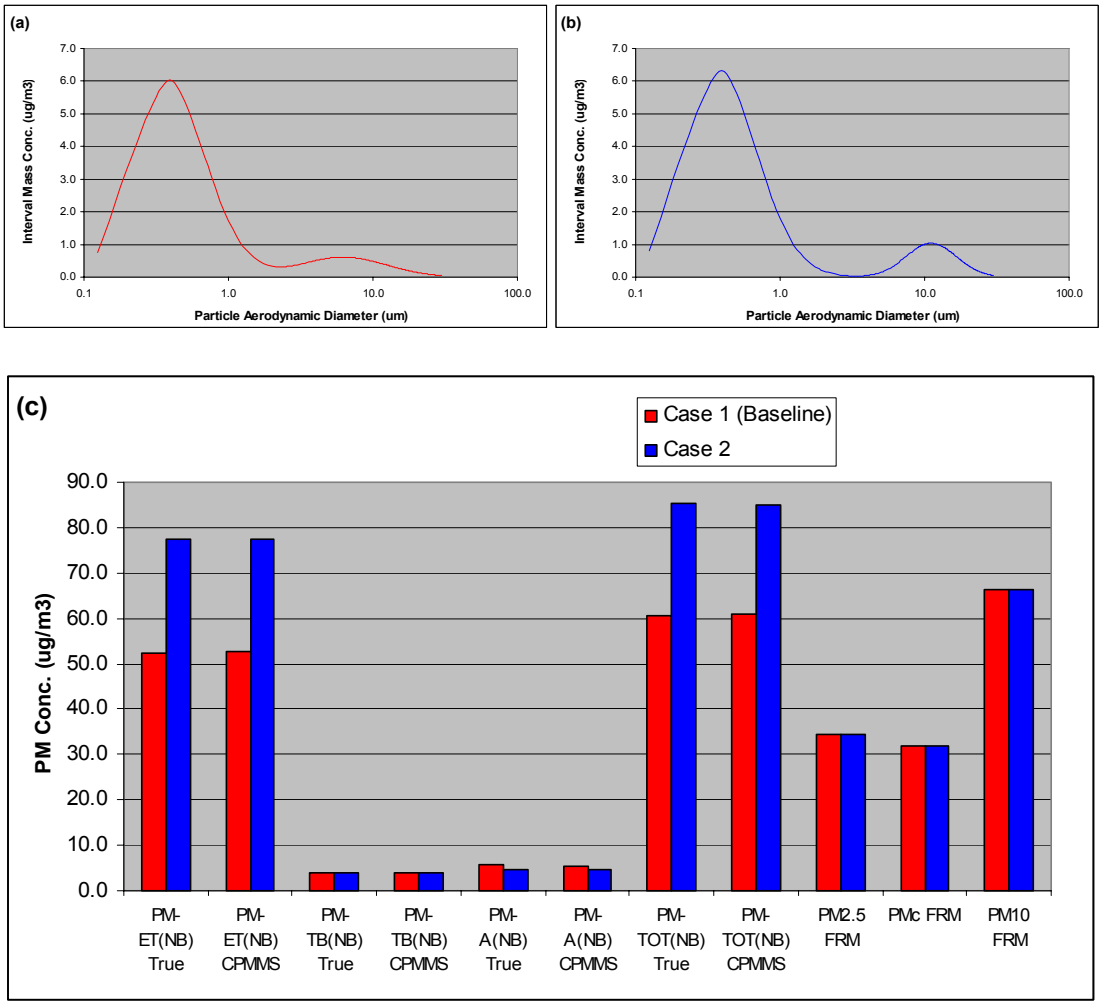


Figure 2. Illustration of two ambient PM cases with the same $PM_{2.5}$, PM_{10}, and $PM_{10-2.5}$ but different health effects; (a) is particle size distribution for Case 1 and (b) for Case 2; (c) is comparison of different PM metrics for Cases 1 and 2.

During the September 21-22 meeting, significant amount of discussions were about the evaporative loss from these candidate methods. As illustrated in Figure 2, the bias with respect to health effects due to PM definitions is as a significant (or larger) problem as evaporation loss. Most of us believe that we should measure what people in the health effect field want to see. They probably would not want to see $PM_{2.5}$, PM_{10}, or $PM_{10-2.5}$ if they were given dosimetry-based PM metrics.

We have become accustomed to think that PM metrics can only be defined by sampling methods. I am afraid that we are missing the target (*i.e.*, the PM that deposit onto the respiratory system as illustrated in the above examples). Our current rules are based on PM defined by FRM, but FRM itself seems an inadequate representative of the health related PM portion. For this reason, I think a PM metrics I refer to as "dosimetry-based PM metrics" is much more meaningful and valuable, especially this concept has be demonstrated to be feasible through CPMMS[1] (also see Appendix A). This approach

would represent a major paradigm shift away from conventional PM metrics such as $PM_{2.5}$, PM_{10}, and $PM_{10-2.5}$. More thorough evaluation is needed and it will take some time to fully develop the system. However, it makes much more scientific sense than the current PM metrics system where cut points and the shape of fractionation curves are somewhat arbitrarily set and do not represent health impact.

It is generally accepted that the steeper the sampler fractionation curve is, the better the sampler is. This means that a perfectly vertical line at 2.5 μ or 10 μ would be ideal. If we use a perfectly vertical line at 2.5 μ without any sampling errors (theoretic conditions), the current $PM_{2.5}$ FRM can produce up to 10% errors purely due to ambient PM size distribution shift alone, which is expected to happen.[1] On the other hand, there is no reason to make a perfectly vertical line at 2.5 μ or 10 μ because it does not represent anything. Nothing in the nature has such a sharp cut. Therefore, the question is what slope will be desirable. The only answer is dosimetry-based PM metrics because in such a system the "sampler fractionation curve" resembles the deposition curve of the human respiratory system.

Peer Review Charge Question 2:
Based on the field study report as well as any other available data, e.g., data from State and local agencies, how does the demonstrated data quality of the $PM_{10-2.5}$ difference method support or detract from it being proposed as a FRM?

Comment:

Based on review of the provided data, I support EPA's proposal to use the $PM_{10-2.5}$ difference method as a FRM, again *within the context of current PM metrics and regulatory framework* (see my comment on Question 1). The data quality of the difference method is higher than that of other candidate methods. The main disadvantage of the difference method is its manual and non-continuous operation. If the EPA envisions that the majority of deployed monitors will be automated continuous equivalent methods rather than the FRM (similar to the current deployment of SO_2 monitors), this disadvantage can be minimized. As far as evaporative loss is concerned, the issue is not unique to this method. If this issue needs to be addressed, the $PM_{2.5}$ FRM should also be changed.

References

1. Y. Zeng, A Comprehensive Particulate Matter Monitoring System and Dosimetry-Based Ambient Particulate Matter Standards, *J. of Air & Waste Management Association*, in review.

Appendix A to Yousheng Zeng's Comments
Dosimetry-Based PM Metrics

This Appendix contains a brief description of dosimetry-based PM metrics. A more detailed description can be found in Reference 1.

Dosimetry-Based PM Metrics

Currently ambient PM level is measured as $PM_{2.5}$ and PM_{10}. They are supposed to be surrogate parameters to indicate potential human exposure to the pollutant (*i.e.*, PM). The definitions of $PM_{2.5}$ and PM_{10} are based on sampler cut point diameter (D_{50}). How well they represent the PM that actually deposits in the human respiratory system is questionable.

A dosimetry-based PM definition is proposed to reflect the ambient PM level that may cause human health effects due to deposition of PM in the respiratory system. It is expressed as:

$$C_D = \sum_i d_{(i)} c_{(i)} \quad (1)$$

Where

C_D = Dosimetry-based ambient PM concentration, $\mu g/m^3$
$d_{(i)}$ = Respiratory tract (or a region of it) deposition fraction on a mass basis for size interval i, unitless.
$c_{(i)}$ = Ambient PM interval mass concentration for size interval i, $\mu g/m^3$

The parameter $d_{(i)}$ is selected depending on the interest of a study or the measurement objective and is fix for that purpose. For example, if the measurement is about PM deposition in the total respiratory tract using adult male nasal breathing as a benchmark, the corresponding total particle deposition fraction [TOT(NB)] as documented in the recent EPA Criteria Document for PM [2] and illustrated in Figure 1 can be used for $d_{(i)}$. The resulting dosimetry-based PM concentration (C_D) can be designated as PM-TOT(NB).

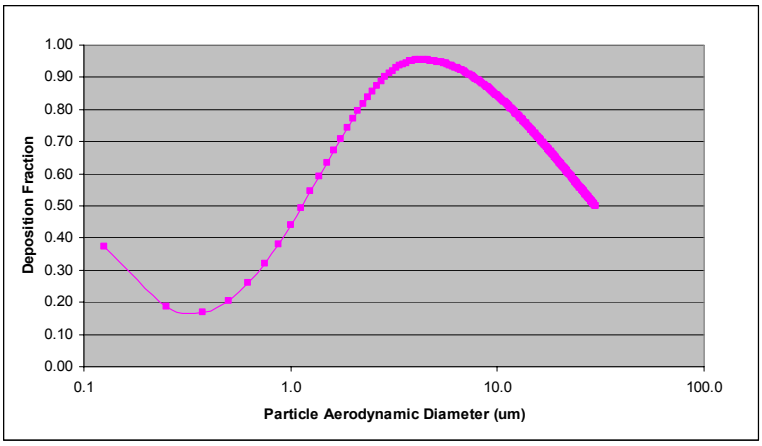


Figure 1. Total deposition in respiratory tract (TOT) for an adult male, nasal breathing (data based on Figure 6-13 in Ref. 2).

The other parameter in the right-hand side of equation (1), $c_{(i)}$, is ambient PM mass size distribution. The following PM measurement system is proposed to significantly improve the accuracy of PM mass size distribution measurement with more tolerable monitoring instruments (therefore promote a wide spread use of ambient PM size distribution).

Comprehensive Particulate Matter Measurement System (CPMMS)

A system called Comprehensive Particulate Matter Measurement System (CPMMS) is developed and is illustrated in Figure 2. It consists of the following hardware and software:

- An aerosol particle sizing device, e.g., Aerodynamic Particle Sizer (APS) Model 3321 manufactured by TSI Inc., or some other particle size distribution measurement methods;
- A mass-based PM sampler, e.g., $PM_{2.5}$ Federal Reference Method (FRM), PM_{10} FRM, or a continuous dichotomous sampler; and
- The algorithm described below.

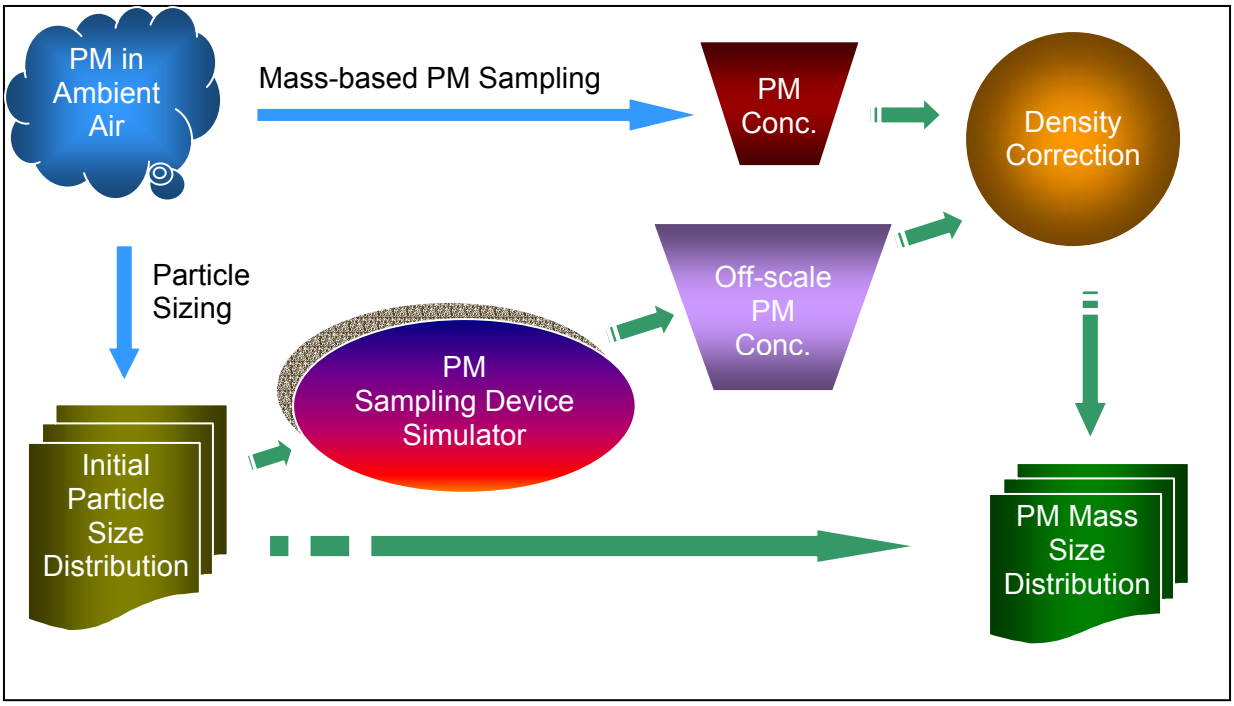


Figure 2. Schematic of CPMMS.

The particle sizing device and the PM sampler are co-located and measuring the same ambient PM sample. The PM sampler measures the mass concentration (typically in $\mu g/m^3$) of the PM fraction that can be collected by the sampler based on the sampling effectiveness curve (aspiration curve or fractionation curve) of the sampler. The result is an aggregated mass concentration over the sampler's designed size range (*e.g.*, 0-2.5 µm for FRM $PM_{2.5}$).

The particle sizing device measures initial particle size distribution in $\mu g/m^3$ for each particle size interval (e.g., 0-0.523 µm, 0.523-0.542 µm, 0.542-0.583 µm, etc.). The measurement is based on aerodynamic properties of the particles and the results in $\mu g/m^3$ are derived based on an assumed particle density. This arbitrary assumption makes the results unreliable in the absolute sense of $\mu g/m^3$ and unsuitable for estimating an aggregated mass concentration over a particle size range per regulatory definition. However, the relative $\mu g/m^3$ values between particle size intervals are much more reliable and can represent the relative particle size mass distribution. In CPMMS, the accurate aggregated mass concentration result obtained by the mass-based PM sampler is used to "calibrate" the results of the particle sizing device to produce accurate PM monitoring data in terms of mass size distribution and aggregation over any size ranges. In order to accomplish the "calibration", the following model is used to simulate the mass-based PM samplers.

A PM sampler can be characterized by its sampling effectiveness curve. The sampling effectiveness curves of $PM_{2.5}$ FRM and PM_{10} FRM based on the data points specified in 40 CFR 53 [3] are depicted in Figure 3 (a).

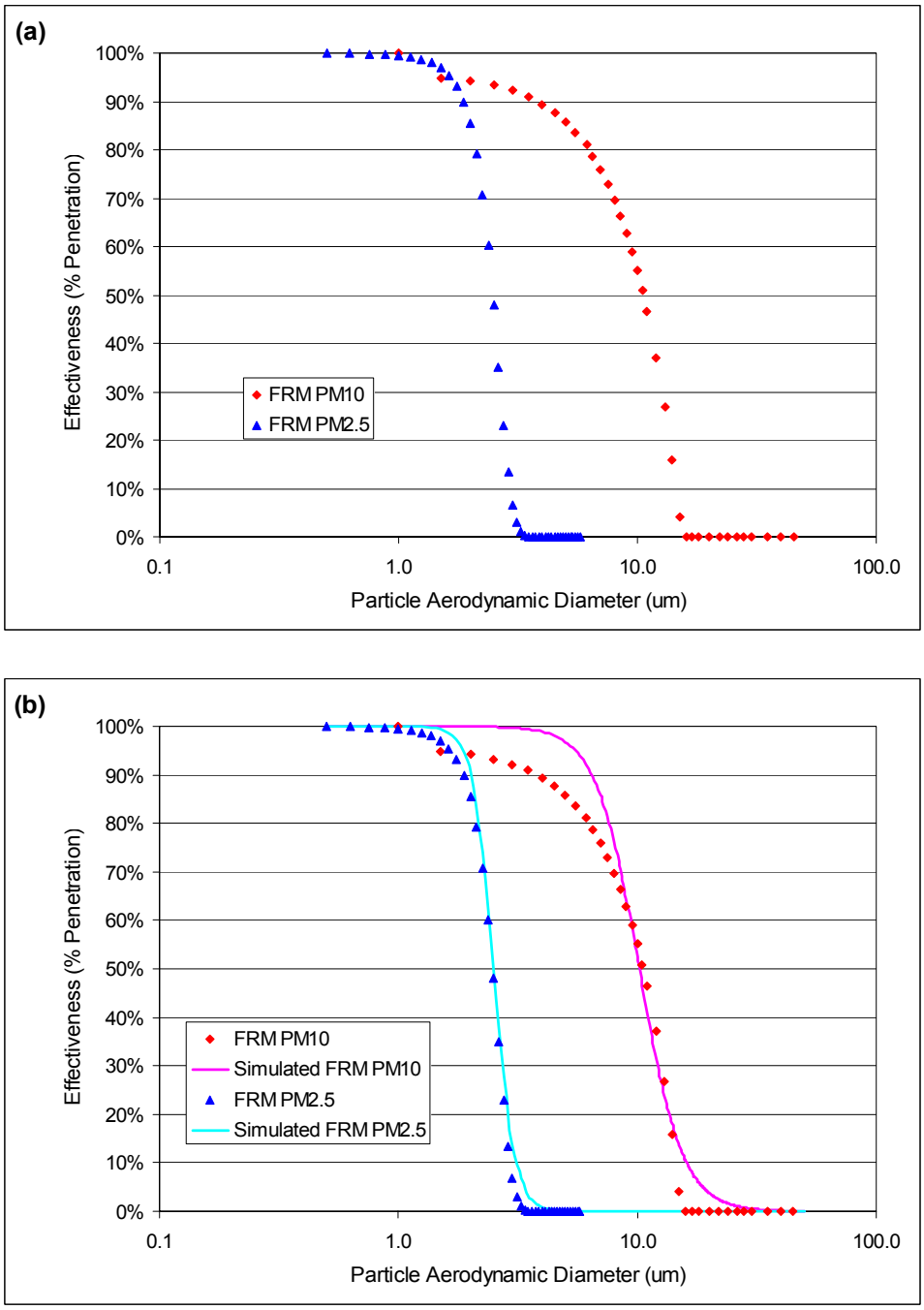


Figure 3. Sampling effectiveness curves of mass-based PM samplers: (a) PM_{10} FRM and $PM_{2.5}$ FRM based on specifications in 40 CFR 53 Subpart D Table D-3 and Subpart F Table F-4, respectively; (b) data in (a) plus sampling effectiveness curves simulated using the model presented in this Appendix.

The sampling effectiveness curves can be modeled using the following equation.

$$p_{(i)} = \frac{1}{1 + \left(D_{(i)} / D_{50} \right)^{n}} \qquad (2)$$

Where

$p_{(i)}$ = Sampling effectiveness (or penetration) for particles in size interval i, % or fraction

$D_{(i)}$ = Representative aerodynamic diameter of particles in size interval i, µm

D_{50} = Sampler cutpoint diameter (particle aerodynamic diameter corresponding to 50% penetration), µm

n = Parameter determining the steepness of the sampling effectiveness curve

PM samplers can be simulated by the above model by selecting proper values of two parameters, D_{50} and n, using the least squares (LS) approach. The accuracy of the simulation can be evaluated by virtually passing the idealized ambient particles as specified in 40 CFR 53 [3] through the model (i.e., simulated sampler) and calculating the sum of the squared deviations of the model-predicted values from the true values specified in 40 CFR 53. The LS results are normalized to be expressed as percent of the total PM mass of idealized ambient particle samples. The model parameters and simulation accuracies are summarized in Table 1. The accuracy of 0.043% in Table 1 suggests that the model represents FRM $PM_{2.5}$ very well. This excellent fit between the model and the true values can be seen graphically in Figure 3 (b). The simulated FRM PM_{10} sampler is less ideal and the accuracy is 3.864% [also see the fit curve in Figure 3 (b)]. This is consistent with the well known fact that the FRM PM_{10} is loosely defined compared to FRM $PM_{2.5}$.

Table 1. Summary of modeled FRM $PM_{2.5}$ and FRM PM_{10} samplers

		Model Parameters		Simulation Accuracy [a]
		D_{50} (µm)	n	
Modeled FRM PM_{10} Sampler		10.0	4.6	3.864%
Modeled FRM $PM_{2.5}$ Sampler	Inlet (10 µm)	10.2	4.8	0.043%
	Fractionator (2.5 µm)	2.5	10	

a. The sum of the squared deviations of the true values specified in 40 CFR 53 and the model-predicted values, expressed as percent of the total PM mass of idealized ambient particle samples.

With the PM sampling device simulator established as eq (2), the initial particle size distribution data can be processed by the simulator. The initial particle size distribution data consist of estimated mass concentrations in $\mu g/m^3$ for each particle size interval. However, the $\mu g/m^3$ value may not be correct because it is based on an assumed density for PM. At this step of the process, we can ignore the accuracy issue of the assumed density and virtually put the PM sample through the simulator. The simulator will predict the PM mass collected on the simulated sampler based on the assumed density as follows:

$$M = \sum_i m_{(i)} = \sum_i p_{(i)} c_{I(i)} \tag{3}$$

Where

M = Predicted PM mass concentration seen by the simulated PM sampler, $\mu g/m^3$

$m_{(i)}$ = Predicted PM interval mass concentration as collected on the sampler filter (i.e., passed sampler inlet cut) for particle size interval i, $\mu g/m^3$

$c_{I(i)}$ = Initial interval mass concentration measured by the particle sizing device (i.e., ambient PM size distribution before passing through the mass based PM sampler) for size interval i, $\mu g/m^3$

If the initial particle size distribution, $c_{I(i)}$, is accurate (i.e., the assumption of particle density happens to be correct), the value of M should match the PM concentration actually measured by the co-located PM sampler. Otherwise, the correct particle density can be derived using the following equations.

$$\rho = \rho_A \frac{C}{M} \tag{4}$$

Where

ρ = Correct particle density, g/cm^3

ρ_A = Assumed particle density in the initial particle mass size distribution, g/cm^3

C = Actual PM concentration measured by the PM sampler, $\mu g/m^3$

When APS is used to measure particle size distributions, it actually measures the aerodynamic diameters of individual particles using time of flight technology. The aerodynamic diameters are used to calculate the volumes of the particles to obtain the volume size distribution. The mass size distribution is finally calculated from the volume size distribution by assuming a particle density:

$$c_{I(i)} = 10^6 v_{(i)} \rho_A \tag{5}$$

Where

10^6 = Conversion factor, 10^6 $\mu g/g$

$v_{(i)}$ = Interval volume concentration measured by the particle sizing device for size interval i, cm^3/m^3 (vol. of particles/vol. of air sample)

If the actual density of particles is known (instead of assumed), an accurate mass size distribution should be:

$$c_{(i)} = 10^6 v_{(i)} \rho \tag{6}$$

Where

$c_{(i)}$ = Actual interval mass concentration for size interval i, $\mu g/m^3$

Combining eqs (4), (5), and (6) can yield the following equation that can be used to calculate actual mass size distributions:

$$c_{(i)} = c_{I(i)} \frac{C}{M} \tag{7}$$

Once an accurate mass size distribution (c_i) is obtained, any aggregated ambient PM monitoring data can be derived by summing up the c_i over the desired size range:

$$C = \sum_i c_{(i)} \tag{8}$$

For example, if the above summation is done for intervals of mass c_i representing particle diameters from 0 to 2.5 µm, the resulting C is for $PM_{2.5}$; if the diameter range is from 2.5 to 10 µm, the result is PMc or $PM_{10-2.5}$, etc.

If it is desired to have PM results that mimic a particular PM sampler, the accurate mass size distribution (c_i) produced by CPMMS should again be put through a simulator of a desired PM sampler with sampling effectiveness curve represented by p_i in the same fashion as described in eq (3), i.e.,

$$C = \sum_i p_{(i)} c_{(i)} \tag{9}$$

CPMMS relies on field measurements by a particle sizing device and a mass-based PM sampler. If the sampler is a FRM sampler, the CPMMS is referred to as CPMMS(FRM). If the sampler is a Dichot sampler, the CPMMS is called CPMMS(Dichot).

As shown above, once accurate mass size distribution is obtained [eq. (7)], PM measurement metrics can be defined (and reconstructed) by exact cut point [eq. (8)], sampler sampling effectiveness curve [eq. (9)], or PM deposition fraction of human respiratory system [eq. (1)]. The last one is dosimetry-based PM metrics and is most relevant and meaningful.

If CPMMS is implemented as a new way of managing ambient air quality for PM, the results will be high quality mass size distribution data. The monitoring data will be available to regenerate various dosimetry-based PM concentrations for different research purposes, *e.g.*, an epidemiological study on ambient PM level defined by particle deposition in the tracheobronchial region of the respiratory tract using eq (1) and the related health impact. The

study based on such PM data will be more meaningful than current PM data. Regulatory authority can also establish dosimetry-based PM standards using human respiratory PM deposition data that are representative of and protective to selected groups of population (*e.g.*, children).

Major Features of CPMMS and Dosimetry-Based PM Metrics

The major features of the above system include:

- It closely matches human health effects and there are no arbitrary cut points. This PM metrics system is expected to have much stronger correlation with epidemiological data than PM_{10}, $PM_{2.5}$, or $PM_{10-2.5}$.
- Unlike past changes from TSP to PM_{10}, to $PM_{2.5}$, etc.), the monitoring data obtained through CPMMS dose not lose its value when regulatory definition of PM changes. The sampling devices do not need to be changed even when the PM definition changes.
- It is significantly less vulnerable to imperfection in sampler design, manufacturing, and operating conditions.

Specific discussions and supporting examples can be found in Ref. 1.

References

1. Y. Zeng, A Comprehensive Particulate Matter Monitoring System and Dosimetry-Based Ambient Particulate Matter Standards, *J. of Air & Waste Management Association*, in review.
2. U.S. Environmental Protection Agency. Air Quality Criteria for Particulate Matter. EPA/600/P-99/002bF, October 2004.
3. U.S. Environmental Protection Agency. Ambient Air Monitoring Reference and Equivalent Methods. Code of Federal Regulations 40, Part 53.

NOTICE

This report has been written as part of the activities of the U.S. Environmental Protection Agency's (EPA) Clean Air Scientific Advisory Committee (CASAC), a Federal advisory committee administratively located under the EPA Science Advisory Board (SAB) Staff Office that is chartered to provide extramural scientific information and advice to the Administrator and other officials of the EPA. The CASAC is structured to provide balanced, expert assessment of scientific matters related to issue and problems facing the Agency. This report has not been reviewed for approval by the Agency and, hence, the contents of this report do not necessarily represent the views and policies of the EPA, nor of other agencies in the Executive Branch of the Federal government, nor does mention of trade names or commercial products constitute a recommendation for use. CASAC reports are posted on the SAB Web site at: http://www.epa.gov/sab.