DRAFT DRAFT DRAFT

GUIDANCE FOR DEMONSTRATING

ATTAINMENT OF AIR QUALITY GOALS FOR

$PM_{2.5}$ AND REGIONAL HAZE

draft 2.1

January 2, 2001

Disclaimer

This is a draft document which reflects comments received as part of a technical review of an earlier version. It presents a set of ideas for demonstrating attainment of air quality goals for $PM_{2.5}$ and regional haze. Our intent is to widely circulate this draft to receive comments and further suggestions. At this stage of its development, the document should not be construed as representing U.S. EPA policy.

Acknowledgments

We would like to thank members of an external review group as well as members of an internal U.S. EPA review group. Individuals on these two groups spent several months reviewing an earlier, preliminary draft of this guidance. We received many interesting and thought-provoking comments from individuals on these two groups. Many of these comments are reflected in this second draft of the guidance.

Table of Contents

Table of Contents, (continued)

Part II. How Should I Generate Modeling Results To Use In Attainment Or Reasonable Progress Tests?

Table of Contents (continued)

Table of Contents (concluded)

Table of Contents (concluded)

List Of Figures

List Of Tables

Foreword

Readers should note that the U.S. Environmental Protection Agency (EPA) is in the process of completing its periodic review of the national ambient air quality standards (NAAQS) for particulate matter. Therefore, these standards could be revised. Furthermore, the U.S. Court of Appeals for the District of Columbia Circuit issued a decision on May 14, 1999 which prohibited enforcement of the 1997 NAAQS for particulate matter and remanded them to EPA for further justification. This decision has been appealed to the U.S. Supreme Court. Oral arguments occurred on November 7, 2000. A decision by the Court is still pending. Nevertheless, concepts developed in the guidance are applicable to other standards for pollutants, such as particulate matter, which are mixtures and which have an important secondary component. Furthermore, issues related to use of models to help design effective control strategies to reduce particulate matter or regional haze are complex and controversial. This draft is being circulated to encourage a dialogue among interested parties. Our intent is to use the information gained from this dialogue, along with the outcome of the pending EPA review of the NAAQS for particulate matter and the results of ongoing analyses, to improve this draft.

This document is a guidance document rather than a regulation. Therefore, it does not impose binding, enforceable requirements on any party, and may not apply to a particular situation based upon the circumstances. U.S. EPA and State decision makers have the discretion to adopt approaches on a case-by-case basis that differ from this guidance where appropriate. Any decisions by the U.S. EPA regarding adequacy of a particular State implementation plan (SIP) to meet national ambient air quality standards (NAAQS) for particulate matter or reasonable progress goals to reduce regional haze will be based on applicable statutes and regulations. Therefore, interested parties are free to raise questions and objections about the appropriateness of the application of this guidance to a particular situation. The U.S. EPA will, and States should, consider whether or not the recommendations in the guidance are appropriate in that situation. This guidance is a living document and may be revised periodically without public notice. The U.S. EPA welcomes public comments on this guidance document at any time and will consider those comments in any future revision of this document.

(This page is left blank intentionally)

Executive Summary

An *attainment demonstration*[1] consists of (a) analyses which estimate whether selected emissions reductions will result in ambient concentrations that meet specified air quality goals, and (b) an identified set of measures which will result in the required emissions reductions. In this document we address only the first part of an attainment demonstration. That is, we describe how to estimate if a control strategy to reduce emissions of particulate matter and its precursors will lead to attainment of annual and 24-hour national ambient air quality standards (NAAQS) for particles as small or smaller than 2.5 μm in diameter (*$PM_{2.5}$*). We also describe how to use modeled and monitored data to estimate whether a control strategy will result in meeting goals to reduce regional haze in *Class I areas* (e.g., national parks, wilderness areas).

This guidance is divided into two major parts. In Part I, we explain how to use modeled and monitored data to estimate if air quality goals for $PM_{2.5}$ and regional haze will be met by a proposed control strategy. We begin by describing a *modeled attainment test* for the annual NAAQS for $PM_{2.5}$ and then for the 24-hour NAAQS. Both tests are similar. The output of each is an estimated future *design value* consistent with the respective forms of the NAAQS. If the future design value does not exceed the concentration of $PM_{2.5}$ specified in the NAAQS, the test is passed. The modeled attainment test is limited to locations with monitored data. We also describe a "hot spot analysis" which may be applied at other, unmonitored locations where there is concern that concentrations of $PM_{2.5}$ may be high.

We next recommend a *modeled test for reasonable progress* meeting goals to reduce regional haze. There are two such goals: (1) improve visibility on days when it is currently poor, and (2) ensure visibility does not deteriorate on days when it is currently good. Like the modeled tests for the $PM_{2.5}$ NAAQS**,** the test for reasonable progress uses both monitored and modeled data. Current visibility is estimated based on monitored components of $PM_{2.5}$ and *coarse particulate matter*. Models are used in a relative sense to estimate how these currently measured concentrations will respond to a control strategy. Multiplying relative (modeled) response times measured concentrations of particulate matter yields an estimate of future concentrations of particulate matter. We then derive future visibility from the estimated future component concentrations of $PM_{2.5}$ and coarse particulate matter. We conclude by comparing future estimated visibility with currently monitored visibility to see if the goals for reducing regional haze are met.

States may use other analyses in addition to the modeled attainment test, hot spot analysis and modeled test for reasonable progress to estimate whether future attainment of the NAAQS or goals for reducing regional haze is likely. Attainment is likely if a preponderance of evidence suggests so. We call this procedure a *weight of evidence determination.*

---

[1]We define terms which appear in *italics* in greater detail in a glossary which appears at the end of this guidance.

Reliability of recommended tests for estimating future attainment or reasonable progress depends on having reliable data bases. We identify and prioritize key data gathering activities and analytical capabilities which will increase credibility of analyses used to estimate if the air quality goals for $PM_{2.5}$ and regional haze will be met. We conclude Part I of the guidance by summarizing documentation and features which should be included in a sound attainment or reasonable progress demonstration.

Part II of this guidance describes how to apply air quality models to generate results needed by the modeled tests for attainment and reasonable progress. The recommended procedure for applying a model(s) has 9 steps.

1. Develop a conceptual description of the problem to be addressed.
2. Develop a modeling/analysis protocol.
3. Select an appropriate model(s) to support the demonstration.
4. Select appropriate periods to model.
5. Choose an appropriate area to model with appropriate horizontal/vertical resolution.
6. Generate meteorological and air quality inputs to the air quality model.
7. Generate emissions inputs to the air quality model.
8. Evaluate performance of the air quality model and perform diagnostic tests.
9. Evaluate prospective control strategies.

Model applications require a substantial effort. States should work closely with the appropriate U.S. EPA Regional Office(s) in executing each step. This will increase the likelihood of approval of the demonstration at the end of the process. Sections ES8.0 - ES15.0 summarize Part II of the guidance.

Part I: How Do I Use Modeled Results To See If An Air Quality Goal Is Met?

ES 1.0 What Are The Recommended Modeled Tests For Attainment Of The Two NAAQS For $PM_{2.5}$?

The two modeled attainment tests for the annual and 24-hour NAAQS use monitored data to estimate current air quality. The attainment test for a given NAAQS is applied at each monitor location within or near a designated nonattainment area for that NAAQS. Models are used in a relative sense to estimate the response of measured air quality to future changes in emissions. Future air quality is estimated by multiplying current monitored values times modeled responses to changes in emissions. Because $PM_{2.5}$ is a mixture, States should use current observations and modeled responses of major components of $PM_{2.5}$ to estimate future concentrations of each component. The predicted future concentration of $PM_{2.5}$ is the sum of the predicted component concentrations. We recommend that $PM_{2.5}$ be divided into the following major components:

- mass associated with sulfates (SO_4);
- mass associated with nitrates (NO_3);

- mass associated with organic carbon (OC);
- mass associated with elemental carbon (EC);
- mass associated with (primary) inorganic particulate matter other than primary sulfate or nitrate particles (IP);
- unattributed mass (equal to the difference between measured $PM_{2.5}$ and the sum of the other 5 components of $PM_{2.5}$) (U).

ES1.1 Modeled Attainment Test For The Annual NAAQS

Emissions and meteorological conditions vary during a year. Therefore, analyses related to the annual NAAQS should estimate future concentrations for each of the 4 calendar quarters. Predicted annual concentrations should be estimated as the average of future concentrations predicted for each of the 4 calendar quarters in a year.

Key steps in the recommended modeled attainment test for the annual NAAQS are:

1. Derive current quarterly mean concentrations (averaged over 3 years) for each of the six major components of $PM_{2.5.}$ This is done by multiplying the monitored quarterly mean concentration of $PM_{2.5}$ times the mean monitored composition of $PM_{2.5}$ for that quarter in 3 consecutive years.

2. For each quarter, apply an air quality model to estimate current and future concentrations for each of the six components of $PM_{2.5}$. Take the ratio of future to current predictions for each component. The result is a component-specific *relative reduction factor* (RRF).

3. For each quarter, multiply the current quarterly mean component concentration times the component-specific RRF. This leads to an estimated future quarterly mean concentration for each component.

4. Average the 4 quarterly mean future concentrations to get an estimated future annual mean concentration for each component. Sum the annual mean concentrations of the six components to obtain an estimated future annual concentration for $PM_{2.5}$.

5. The test is passed if the estimated future annual mean concentration of $PM_{2.5}$ is $\leq$ 15.0 $\mu g/m^3$ at each monitoring site.

The modeled attainment test for the annual NAAQS is described in Sections 3.1 and 3.3.

ES1.2 Modeled Attainment Test For The 24-Hour NAAQS

This test is applied at every monitoring site within or near a nonattainment area for the 24-hour NAAQS. It is very similar to the test recommended for the annual NAAQS. The main differences are in estimating current concentrations for the six components of $PM_{2.5}$ and in choice of days used to develop values for the component-specific RRF's. There are 4 key steps.

1. Compute the 98th percentile ambient concentration for $PM_{2.5}$ in each of three consecutive years. Average these three values to get a current *design value*. For each day with high measured $PM_{2.5}$ (i.e., $\geq$ 55 $\mu g/m^3$), estimate the fraction of measured $PM_{2.5}$ attributable to each of its six components. Develop quarter-specific average fractional profiles using all days considered. Multiply the design value times the average fractional profile to get estimated values of current concentrations of each component which correspond to the current design value for $PM_{2.5}$.

2. Apply an air quality model to simulate current and estimated future quarterly emissions. This leads to quarter-specific modeled estimates of current and future concentrations for each of the six components of $PM_{2.5}$. Compute quarterly component-specific RRF's by taking the ratio of modeled future to current concentrations.

3. Estimate future concentrations of components corresponding to a future 98th percentile concentration of $PM_{2.5}$. This is done by multiplying current concentrations of each component obtained in step 1 times the component-specific RRF derived in step 2.

4. Sum the future concentrations of each component to get an estimated future 98th percentile concentration of $PM_{2.5}$. The test is passed if the future design value calculated at each monitor is $\leq$ 65 $\mu g/m^3$.

The modeled attainment test for the 24-hour NAAQS is described in Sections 3.2 and 3.3.

ES2.0 What Is A Hot Spot Analysis And Why Do I Need It?

The modeled attainment tests are applied at monitoring sites. This is sufficient if the sites are spatially representative and major gradients in measured particulate matter are unlikely. The robust nature of the annual NAAQS (i.e., an annual mean averaged over three years), long sampling time (24 hours) inherent in the data and importance of secondary particulate matter (requiring several hours for formation) make numerous sharp gradients in ambient $PM_{2.5}$ unlikely. However, there may be substantial sources of primary particulate matter in some limited geographic locations. Further, atypically high concentrations of primary particulate matter from one or more nearby sources are possible at a location on a given day. This may be of particular concern for meeting the 24-hour NAAQS. Therefore, we suggest a procedure to assess whether a hot spot analysis is necessary and describe what such an analysis entails.

States may determine whether a hot spot analysis is needed by reviewing spatial distributions of primary $PM_{2.5}$ emissions (i.e., directly emitted inorganic and organic particulate matter) or by reviewing past model predictions within a grid superimposed over the nonattainment area. If there are cells containing primary emissions which are well in excess of any near a monitoring site and this excess is largely attributable to a limited number of sources, a hot spot analysis may be warranted near these sources. Alternatively, a hot spot analysis may be needed if past grid modeling results identify one or more grid cells with concentrations of primary particulate matter and $PM_{2.5}$ which are consistently well above concentrations predicted

near any monitoring site.

A hot spot analysis needs to be designed on a case by case basis. Since the major focus is on primary particulate matter, Lagrangian (e.g., Gaussian) or finely gridded models which do not consider chemistry are suitable for hot spot analyses. We recommend focusing the analysis on one or a limited number of sources and their proximity. The model should be applied to simulate primary particulate concentrations everyday within the 3-year period used to calculate the current design values for $PM_{2.5}$. This information is reviewed to obtain the highest future mean 98^{th} percentile prediction (24-hour NAAQS) or the highest future average mean concentration (annual NAAQS). Results are then superimposed on mean concentrations (averaged over 3 years) for the other components of $PM_{2.5}$ derived from the modeled attainment test for the annual NAAQS. If the result is ≤ 65 $\mu g/m^3$ (24-hour NAAQS) and/or ≤ 15.0 $\mu g/m^3$ (annual NAAQS), no remedial action is needed for the sources which are modeled in the hot spot analysis beyond that already proposed in the SIP revision.

Hot spot analyses are discussed in Section 3.4.

ES3.0 What Is The Modeled Test For Reasonable Progress Reducing Regional Haze?

We recommend a modeled test for reasonable progress which is conceptually similar to the attainment tests for the $PM_{2.5}$ NAAQS. The modeled test for reasonable progress needs to reproduce as closely as possible the procedure to be used at a future date (e.g., 2018) to assess whether observed trends in regional haze are consistent with reasonable progress.

The U.S. EPA's 1999 rule for reducing regional haze identifies two goals. First, visibility on the worst current days needs to be improved by an agreed upon number of *deciviews*. Second, visibility on the best current days may not deteriorate. A similar test is applied separately to address each of these two goals.

Light extinction is initially characterized in the test by an extinction coefficient (b_{ext}) which reflects scattering and absorption of light. Values for b_{ext} are estimated using an empirically derived equation which relates the extinction coefficient to relative humidity and the following components of particulate matter:

- mass associated with sulfates (SO_4);
- mass associated with nitrates (NO_3);
- mass associated with organic carbon (OC);
- mass associated with elemental carbon (EC);
- mass associated with inorganic particulate matter other than primary sulfate or nitrate particles(IP) (assumed to be all "crustal material");
- coarse mass (CM) (i.e., PM_{10} - $PM_{2.5}$).

For each Class I area, the test uses monitored concentrations of these components, together with a "relative humidity adjustment factor", to estimate typical current values for b_{ext} on days with worst and best visibility. The test uses location- and month-specific climatological mean values for the humidity adjustment factor. These have been derived beforehand by the U.S. EPA. Thus, we do not use monitored or modeled day-specific values for relative humidity to derive day-specific humidity adjustment factors. Instead, our procedure is analogous to normalizing a trend to account for meteorological differences. In assessing reasonable progress, we are interested in seeing whether a control strategy has been effective in reducing haze. This is difficult to do unless changes in visibility resulting from uncontrollable changes in relative humidity are factored out.

The test uses an air quality model to estimate relative changes in each of the preceding components of particulate matter (i.e., to derive component-specific *relative reduction factors*). Current, monitored components then get multiplied by the corresponding component-specific relative reduction factors to estimate future concentrations for each component. Future values for b_{ext} are estimated using future estimated concentrations of each component of particulate matter and climatological values for the humidity adjustment factor. Finally, differences in future vs. present values for b_{ext} are converted to differences in deciviews and compared to the relevant goal for reasonable progress. The test is passed if both goals for reasonable progress are met.

Summarizing, the recommended modeled test for reasonable progress has 6 steps.

1. For each Class I area, rank visibility on each day with observed speciated $PM_{2.5}$ data plus PM_{10} data for each of the 5 years comprising the base period.

2. For each of the 5 years comprising the base period, calculate the mean extinction coefficient for the 20% of days with worst and 20% of days with best visibility. For each Class I area, calculate the average mean extinction coefficients for worst and best days from the five year-specific values.

3. Use an air quality model to simulate current emissions and future emissions. Use the resulting information to develop relative reduction factors for each component of particulate matter identified in an empirical relationship between b_{ext} and the components.

4. Multiply the relative reduction factors times the measured PM components to estimate future values for each of these components on modeled days with poor (good) visibility.

5. Using the results in Step 4, recalculate the mean extinction coefficients for the previously selected days with best and worst visibility in each of the five base years. Compute the future average mean extinction coefficients for the worst and best days.

6. Convert the difference in average mean extinction coefficients to a difference in deciviews.

Compare the difference in deciviews to the goals for reasonable progress.

The recommended modeled test for reasonable progress is described in Section 4.0.

ES4.0 What Is A Weight Of Evidence Determination?

A State should always utilize available air quality, meteorological and emissions data to complement a modeling analysis. As we discuss in Section 9.0, these data are used to develop a conceptual description of an area's nonattainment or visibility problem. This description is useful for guiding a modeling analysis. If the modeled test is passed or nearly passed a State may choose to use weight of evidence (WOE) to estimate if the air quality goal will be met.

A weight of evidence determination is a diverse set of technical analyses performed to corroborate findings of the modeled attainment test. If a weight of evidence determination is used, a State should consider a recommended core set of analyses consisting of (1) a set of air quality model results which includes the previously described tests plus additional analyses of estimated concentrations, (2) an analysis of observed air quality and estimated emissions trends, and (3) an analysis of outcomes produced by observational models.

We identify factors which enhance credibility of evidence produced by each of the core analyses, as well as outcomes which would be consistent with the likelihood that a strategy demonstrates attainment.

A State may include other types of analyses, in addition to the core analyses, in a weight of evidence determination. For another analysis to be considered three criteria should be satisfied:

(1) a State should discuss why the proposed analysis is relevant for assessing attainment,

(2) a State should identify the procedure to be used and the data base available to support it, and

(3) a State should identify (in advance) outcomes which would be consistent with a hypothesis that a proposed strategy will lead to attainment.

Weight of evidence and its use is discussed in Section 5.0.

ES5.0 Why Do We Recommend These Modeled Tests And Offer An Option To Perform Weight Of Evidence Analyses?

The modeled attainment and reasonable progress tests we recommend for the $PM_{2.5}$ NAAQS divide $PM_{2.5}$ into major components. The impact of a control strategy on $PM_{2.5}$ or on light extinction is the sum of the impacts attributable to individual components. To estimate these impacts, models are used to predict the fractional decrease (increase) of each component.

These fractional changes are applied to currently monitored component concentrations to estimate future concentrations for each component and, ultimately, future concentrations of $PM_{2.5}$ and/or light extinction. Thus, model predictions are used in a relative rather than absolute sense. Results of other analyses (e.g., other model outputs, trend analysis, use of observational models) may be used to complement the modeled attainment or reasonable progress tests. If the weight of evidence produced by the tests and these other analyses suggest attainment is likely, we conclude that the strategy, if implemented, will lead to attainment. We believe this approach is appropriate for several reasons.

1. Particulate matter is a mixture. To accurately estimate effects of control measures on ambient $PM_{2.5}$ it is necessary to start out with the components in approximately the correct proportion. This is made more likely by using monitored values to characterize current composition of $PM_{2.5.}$ Models are best used to estimate how these correctly proportioned components respond to simulated control measures.

2. Secondary particulate matter constitutes an important fraction of $PM_{2.5}$ and source of light extinction. Modeling requirements for secondary and primary particulate matter differ in their need to consider atmospheric chemistry and in the degree of spatial resolution needed for the modeling. Treating components explicitly facilitates using a mix of models which are responsive to primary vs. secondary PM and which is more cost effective than using a single model to address all parts of a problem involving particulate matter.

3. It is easier to relate results of a modeled attainment test to the forms of the two NAAQS and goals for regional haze if we use currently monitored data to establish current conditions. The annual and 24-hour NAAQS both address a statistically determined concentration which, in addition, is averaged over 3 consecutive years. These concentrations are called "design values". It may be difficult to relate model results to such NAAQS unless numerous days are modeled from each year. Models which can address secondary particulate matter are resource intensive and thus may be incompatible with considering many alternative strategies simulating many days. However, monitoring data should be available for many days each year. These data can readily be used to estimate current design values and corresponding species concentrations. In effect, the monitoring data serve to "anchor" model results to the form of the NAAQS as well as to current mean "worst" and "best" visibility.

4. Using models to predict future attainment has associated uncertainty. Our suggested approach does not eliminate uncertainty, but reduces it in three important ways. First, monitored data (e.g., current design values) are incorporated into the test. These data are likely measured with greater accuracy than an absolute model prediction, and precision of the measurements is better known. Second, the outcome of the test is based on a composite set of calculations from several modeled days rather than a single day. This reduces the risk of choosing an inappropriate strategy on the basis of a single outcome which is subject to uncertainty. Third, if the outcome of the test is close to pass/fail, a weight of evidence determination may be used to see whether other analyses provide corroborative evidence for conclusions drawn from the test.

ES6.0 What Data Gathering And Other Priorities Should I Consider To Support This And Future SIP Revisions?

We note three major activities requiring data and/or enhanced analysis: generating modeling results; applying the results in attainment or reasonable progress tests and applying weight of evidence analyses. We identify data needs and other priorities related to these three activities. Data gathering priorities depend on the nature of the existing data base. However, in general, we suggest that highest priority be given to the following efforts.

1 (tie). Develop accurate emission factors and seasonal/diurnal activity levels for major anthropogenic sources and source categories of $PM_{2.5}$, SO_2, NH_3, VOC, NOx and CO.

1 (tie). Make speciated ambient measurements of $PM_{2.5}$ at mass monitoring sites likely to exceed the NAAQS and in Class I areas.

3. Develop locally applicable source speciation profiles for primary PM and VOC.

4. Make continuous, or at least several, upper air meteorological measurements/day and ensure these are accompanied by surface meteorological measurements below the site of the resulting vertical profiles.

5. Develop software to post process modeling and monitored data so that the modeled tests for attainment and reasonable progress can be readily applied.

6. Arrange a limited duration study in which $PM_{2.5}$ and its components are sampled continuously or over periods not exceeding 6 hours at selected locations with routine 24-hour sampling.

States should retain the ability to perform or sponsor modeling or related analyses subsequent to approval of an initial demonstration of attainment or reasonable progress. Subsequent reviews will likely be necessary.

Data needs and priorities are addressed in Section 6.0.

ES7.0 What Documentation Is Needed To Support The Modeled Demonstrations?

States should address the subject areas shown in Table ES.1. Documentation should be accompanied by an executive summary which covers each of the areas shown in the table. Documentation requirements are covered in Section 7.0.

Table ES.1. Recommended Documentation For Demonstrating Attainment Of Air Quality Goals For $PM_{2.5}$ Or Regional Haze

Subject Area	Purpose of Documentation	Issues Included
Conceptual Description of $PM_{2.5}$ or Regional Haze Problem	**To provide a short narrative of observational evidence which led to the selected modeling approach and strategies which were investigated.**	**Measurements used** **Analyses performed** **Resulting rationale used to support the modeling approach and strategies investigated**
Modeling/Analysis Protocol	**Communicate scope of the analysis and document stakeholder involvement**	**Names of stakeholders participating in preparing and implementing the protocol** **Types of analyses performed** **Steps followed in each type of analyses** **Days and domain considered**
Emissions Preparations and Results	**Assurance of valid, consistent emissions data base and that appropriate procedures are used to derive emission estimates needed for air quality modeling**	**Data base used and quality assurance methods applied** **Data processing used to convert data base to model-compatible inputs** **Deviations from existing guidance and underlying rationale** **Emissions model(s) used and justification for choice of models** **$PM_{2.5}$, SO_2, NH_3, VOC, NOx, CO emissions by State/county for major source categories.**

Table ES.1. Recommended Documentation For Demonstrating Attainment Of Air Quality Goals For $PM_{2.5}$ Or Regional Haze (continued)

Subject Area	Purpose of Documentation	Issues Included
Air Quality/Meteorology Preparations and Results	**Assurance that representative air quality and meteorological inputs are used in analyses**	**Extent of data base and procedures used to derive & quality assure inputs for analyses used in the weight of evidence determination** **Departures from guidance and their underlying rationale** **Meteorological model(s) used** **Performance of meteorological model if used to generate meteorological inputs to the air quality model**
Performance Evaluation for Air Quality Model (and Other Analyses)	**Show decision makers and the public how well the model (or other analyses) reproduced observations or otherwise performed on the days selected for analysis**	**Summary of observational data base available for comparison** **Identification of performance tests used and their results for major components of PM, mass of $PM_{2.5}$ and key gas phase species** **Ability to reproduce observed temporal and spatial patterns for major components of PM, mass of $PM_{2.5}$ and key gas phase species** **Overall assessment of what the performance evaluation implies**
Diagnostic Tests	**Ensure rationale used to adjust model inputs or to discount certain results is physically justified and the remaining results make sense.**	**Results from application prior to adjustments** **Consistency with scientific understanding and expectations** **Tests performed, changes made and accompanying justification** **Short summary of final predictions**

Table ES.1. Recommended Documentation For Demonstrating Attainment Of Air Quality Goals For $PM_{2.5}$ Or Regional Haze (continued)

Subject Area	Purpose of Documentation	Issues Included
Description of the Strategy Demonstrating Attainment (or Reasonable Progress)	**Provide the EPA and the public an overview of the plan selected in the attainment or reasonable progress demonstration.**	**Qualitative description of the selected strategy** **Reductions in $PM_{2.5}$, SO_2, NH_3, CO, VOC, and/or NOx emissions from each major source category for each State/county from current (identify) emission levels** **Clean Air Act mandated reductions and other reductions** **Show relative reduction factors estimated for each major component of $PM_{2.5}$ and (for Regional haze applications) CM** **Show predicted site-specific future $PM_{2.5}$ design values for the selected control scenario and identify any location for which a hot spot analysis was needed together with results of the hot spot analysis** **Identification of authority for implementing emission reductions in the selected strategy** **Evidence that emissions will remain at or below projected levels throughout the 3-year period used to determine future attainment for $PM_{2.5}$-related applications, and the 5-year period used to determine if reasonable progress has occurred for visibility-related applications**

Table ES.1. Recommended Documentation For Demonstrating Attainment Of Air Quality Goals For $PM_{2.5}$ Or Regional Haze (continued)

Subject Area	Purpose of Documentation	Issues Included
Data Access	**Enable the EPA or other interested parties to replicate model performance and simulation results for the selected strategy, as well as results obtained with other analyses**	**Assurance that data files are archived and that provision has been made to maintain them** **Technical procedures for accessing input and output files** **Identify computer on which files were generated and can be read, as well as software necessary to process model outputs** **Identification of contact person, means for downloading files and administrative procedures which need to be satisfied to access the files**
Weight of Evidence Determination ($PM_{2.5}$ NAAQS-related applications)	**Assure the EPA and the public that the strategy is likely to produce attainment of the NAAQS within the required time.**	**Description of the modeled attainment test and observational data base used** **Identification of air quality model(s) used** **Identification of other analyses performed** **Outcome of each analysis, including the modeled attainment test** **Assessment of the credibility associated with each type of analysis in this application** **Narrative describing process used to conclude the overall weight of available evidence supports a hypothesis that the selected strategy is adequate to attain the NAAQS**

Table ES.1. Recommended Documentation For Demonstrating Attainment Of Air Quality Goals For $PM_{2.5}$ Or Regional Haze (concluded)

Subject Area	Purpose of Documentation	Issues Included
Weight of Evidence Determinations (related to the reasonable progress goals to reduce regional haze)	**Assure the EPA and the public that the strategy is likely to meet reasonable progress goals to reduce regional haze**	**Description of the modeled test for reasonable progress and observational data base used** **Identification of air quality model used** **Identification of other analyses performed** **Outcome of each analysis, including the modeled test for reasonable progress** **Assessment of the credibility associated with each type of analysis in this application** **Narrative describing process used to conclude the overall weight of available evidence supports a hypothesis that visibility on the days with best visibility will not deteriorate and the goals for days with poor visibility will be met**
Review Procedures Used	**Provide assurance to the EPA and the public that analyses performed in the attainment or reasonable progress demonstration reflect sound practice**	**Scope of technical review performed by those implementing the protocol** **Assurance that methods used for analysis were peer reviewed by outside experts** **Conclusions reached in the reviews and the response thereto**

Part II: How Should I Generate Modeling Results To Use In Attainment Or Reasonable Progress Tests?

Generating model results for subsequent use in the attainment or reasonable progress tests requires several steps.

- develop a conceptual description of the problem at hand;
- develop and implement a modeling/analysis protocol;
- select appropriate air quality model(s) to use;
- select periods which need to be modeled;
- choose a modeling domain and select an appropriate spatial resolution;
- develop appropriate meteorological and air quality inputs to the model;
- develop emissions inputs to the air quality model;
- evaluate and diagnose model predictions;
- simulate and evaluate prospective control strategies.

ES8.0 What Is A "Conceptual Description"?

A conceptual description is a qualitative way of characterizing an area's nonattainment or light extinction problem. For example, is the problem one dominated by local emissions or are regional factors also important? Do sites violating the NAAQS reflect a spatial or temporal pattern in some way?, Do poor visibility days tend to be characterized by relatively high concentrations of the same components of particulate matter?, etc.. A conceptual description is based on use of readily available air quality, meteorological and emissions information. It may be refined later as additional analyses are performed. States should develop a conceptual description of the nonattainment or light extinction problem(s) they are seeking to solve as an initial step in developing a solution. It serves as a means for guiding later decisions which are needed in a modeling analysis. Suggestions for developing a conceptual description are found in Section 9.0.

ES9.0 What Does A Modeling/Analysis Protocol Do And What Should It Contain?

A modeling/analysis protocol is a document which identifies methods and procedures to be used in the analyses. The protocol also identifies ground rules to be followed in undertaking analyses to estimate emission reductions needed to meet the NAAQS or a reasonable progress goal. Ground rules include identifying the methods, data bases and procedures to be used to obtain results, a description of how affected stakeholders in the modeling/analysis process will be encouraged to participate, the process by which decisions will be made, and means used for communicating issues and decisions. As the name implies, the protocol should address use of other analyses as well as air quality modeling. The document is usually prepared by the State/local agency(ies), or regional planning organization having lead responsibility for the modeling/analysis, in consultation with stakeholders. The protocol should be kept up to date to reflect major subsequent decisions made after the initial version is completed. Specific topics which should be included in the protocol are identified in Section 10.0.

ES10.0 What Should I Consider In Choosing An Air Quality Model?

$PM_{2.5}$ consists of primary and secondary particles. Modeling requirements for primary

and secondary particulate matter differ. For example, spatial gradients in primary particulate matter may require more spatial detail in the modeling, perhaps over a relatively limited domain. Relationships between secondary particulate matter and its gaseous precursors may be non-linear, requiring use of a numerical grid model with an atmospheric chemistry algorithm applied over a large domain. As a result, States may find it necessary to use one air quality model to address primary components of $PM_{2.5}$, and another model to address secondary components.

Modeling used to support attainment demonstrations for the $PM_{2.5}$ NAAQS may consist of the following combinations:

- a photochemical grid model to address secondary components plus a finer scale (nested) grid model (without chemistry) to address primary components;

- a photochemical grid model for secondary components plus a Lagrangian model to address primary components;

- a fine scale grid model (without chemistry) or a Lagrangian model to address primary components, with no modeling of secondary components.

If the last option is chosen, it should include a justification why an increase in secondary components is not likely to lead to a violation of the NAAQS despite the prescribed reduction in primary particulate matter.

A regional scale photochemical grid model should be used to support demonstrations of reasonable progress reducing regional haze. Primary components may be modeled using the same spatial resolution used for the regional analysis. Exceptions may be required on a case by case basis.

For an air quality model to qualify as a candidate for use in an attainment demonstration for $PM_{2.5}$ or in a reasonable progresss demonstration, a State needs to show that it meets several general criteria.

1. The model has received a scientific peer review.

2. The model can be demonstrated applicable to the specific problem.

3. Data bases needed to perform the analysis are available and adequate.

4. Available past appropriate performance evaluations have shown the model is not biased toward underestimates.

5. A protocol on methods and procedures to be followed has been established.

6. The developer of the model must be willing to make the source code available to users

for free or for a reasonable cost, and the model cannot otherwise be proprietary.

So far as choosing individual models meeting the preceding requirements, States should determine what attributes are needed for a qualifying model to address an area's $PM_{2.5}$ or regional haze problem, and then choose among models possessing these attributes. Five factors should be considered in selecting an air quality model for a specific application. Selection of an air quality model should be concurred with by the appropriate U.S. EPA Regional Office and U.S. EPA Model Clearinghouse. The five factors are listed approximately in order of importance.

1. Nature of the air quality problem leading to nonattainment of the $PM_{2.5}$ NAAQS or light extinction should first be assessed, and the selected model(s) should have attributes and capabilities consistent with the resulting conceptual description of the problem.

2. Availability, documentation and past performance should be satisfactory.

3. Relevant experience of available staff and contractors should be consistent with choice of a model.

4. Time and resource constraints may be considered.

5. Consistency of the model with what was used in adjacent regional applications should be considered.

Prior to using model results in a specific application, those implementing the modeling protocol should show that the model performs adequately. We discuss choice of air quality models in Section 11.0.

ES11.0 How Do I Choose Appropriate Periods To Model?

Choice of days to model depends on the air quality goal being addressed (i.e., annual NAAQS, 24-hour NAAQS or reasonable progress goals for days with current best and worst visibility). However, there are some generally applicable criteria for selecting days to model.

1. Choose a variety of meteorological conditions to model.

2. Include some days with intensive data bases, to the extent feasible.

3. Model a sufficient number of days so that relative reduction factors (RRF) are likely to be stable.

In addition, choice of days may reflect several practical considerations.

1. What days have already been modeled?

2. Give preference to days occurring during the period corresponding to the current design value used in the modeled attainment test or the base period used in the modeled test for reasonable progress.

3. Include weekends among the selected days to improve the potential for evaluating model performance.

4. If applying a regional model, choose periods meeting other selection criteria in as many nonattainment or Class I areas as possible.

For the modeled attainment test addressing the annual NAAQS, it is preferable to model every day in a year which is representative of the 3-year period used to determine the current annual design value. However, we believe that modeling 15 or more days per calendar quarter (i.e., 60 or more days in a year) should yield relative reduction factors which are close to what would be obtained if every day were modeled. States may instead choose to model a set of prototypical meteorological conditions and then weight the results according to how frequently each set of conditions occurs.

For modeled attainment tests addressing the 24-hour NAAQS, States should model periods which include all observed exceedances of 65 $\mu g/m^3$ in the nonattainment area. If it is not feasible to model 10 or more days with exceedances, States should model periods including at least 10 days with observations $\geq$ 55 $\mu g/m^3$.

Hot spot analyses may sometimes be needed to supplement the modeled attainment test. Since these analyses are performed with non-resource intensive models, we recommend that the model which is used be applied for every day during the 3-year period used to compute the current design value. If this is not feasible, a similar 3-year period may be chosen.

Since it is likely that regional planning organizations/States will choose to model effects of a strategy in many Class I areas simultaneously, it may be most efficient to model a single representative year, drawn from within the base period. This will ensure that modeled estimates for each Class I area are based on simulations of a large number of "worst" and "best" visibility days. If this is not feasible, those implementing the modeling protocol should try to model at least 10 days with current "worst" visibility and 10 days with current "best" visibility in as many Class I areas as possible. States should avoid missing this minimal requirement by a large margin in any Class I area, as it increases the possiblity that the estimated RRF may not be representative of how air quality in such an area will respond to a control strategy.

Selecting periods to model is discussed in greater detail in Section 12.0.

ES12.0 How Do I Select A Modeling Domain And Its Horizontal/Vertical Resolution?

A modeling domain identifies the geographic bounds of the area which is modeled. In a grid model, horizontal resolution is related to the horizontal dimensions of individual grid cells. Vertical resolution is determined by the number of grid cells (i.e., layers) considered in the vertical direction. Appropriate domain size and resolution differ for applications related to the NAAQS for $PM_{2.5}$ vs. regional haze. Further, different recommendations are appropriate when using models to assess effects of reducing emissions of primary vs. secondary particulate matter.

If secondary particulate matter is an important cause of a nonattainment problem and/or a strategy includes altering SO_4, NO_3 or secondary OC, we recommend using a regional scale modeling domain, at least 1000 km on a side. Smaller, urban scale modeling domains (< ~ 300 km on a side) may be used to consider strategies affecting concentrations of primary components (IP, EC and primary OC). A State may consider a strategy which addresses both primary and secondary components of $PM_{2.5}$ by using a nested grid in which the regional scale domain encompasses the urban scale one.

States should consider performing diagnostic tests to asses sensitivity of relative reduction factors to horizontal grid cell size. However, as default recommendations, States may use grid cells as large as 36 km on a side outside of nested urban areas. Horizontal cell size within urban nonattainment areas and their immediate surroundings should be no larger than 12 km on a side. If a strategy depends on reducing primary emissions, maximum recommended horizontal cell size in and near the nonattainment area should be reduced to 4-5 km on a side.

A State may choose to use a Lagrangian (e.g., Gaussian) model rather than numerical grid models to consider portions of a strategy which address emissions of primary particulate matter. If this is done, the nested urban grid (used to consider secondary particulate matter in/near a nonattainment area) may use horizontal cells as large as 12 km on a side. Lagrangian models should not be applied to sources which are more than 50 km from a monitoring site where the modeled attainment test is being applied. When Lagrangian models are applied in a hot spot analysis, States may assume that the maximum impact from a source of primary particulate matter occurs within 15 km of the source.

Because secondary particulate matter is likely to be an important cause of light extinction and most Class I areas are located in rural/remote areas, a regional modeling domain at least 1000 km on a side should be used to support the modeled test for reasonable progress. Horizontal grid cell size as large as 36 km on a side may be used. States may use plume in grid modeling to characterize effects of large point sources within ~3 grid cells of a Class I area, or a monitoring site in a nonattainment area. Because regional planning organizations are likely to consider effects of a strategy on numerous Class I areas simultaneously, they should consider using one of two superdomains for their analyses: one for the western part of the contiguous U.S. (with mostly low relative humidity) and one for the eastern part of the contiguous U.S. (with

mostly high relative humidity).

Guidance for the number of grid cells in the vertical direction (i.e., layers) used in modeling is the same for $PM_{2.5}$ NAAQS-related and regional haze-related applications. We recommend using a minimum of 9-11 layers. Demarcations between layers should correspond to those used in meteorological models generating input to the air quality model. Priority should be given to having fine vertical resolution near the earth's surface where emissions occur. The surface layer should be less than 50 meters thick. Defining maximum afternoon mixing height as precisely as feasible should also be a goal.

Choice of domains, horizontal grid cell size, number of layers, use of plume in grid models and geographic limits on use of Lagrangian (Gaussian) models are discussed in Section 13.0.

ES13.0 How Do I Produce Meteorological And Air Quality Inputs Needed By An Air Quality Model?

We recommend that States use a dynamic meteorological model with four dimensional data assimilation (FDDA) as the principal means for generating meteorological inputs needed by numerical grid or Lagrangian models used for attainment or reasonable progress demonstrations. Any such meteorological model which has received a scientific peer review may be used. As with the output from emissions models, it is critical that results of meteorological models be quality assured. We identify several potentially useful means for doing so:

1. comparison with upper air measurements "held back" from use in FDDA;

2. comparison of calculated trajectories with observed air quality patterns;

3. use of computer graphics to discern spatial discrepancies;

4. simulation of inert tracers to identify discontinuities or mass balance problems;

5. comparing results obtained with different meteorological models;

6. calculating and comparing divergence and/or dimensionless parameters and comparing these with expected ranges;

7. comparing spatial air quality patterns obtained with a grid model vs. observed patterns,

8. comparing outputs from a meteorological model for key variables vs. estimates for the same variables used in the air quality model, and

9. using process analysis to flag contributions made to unexpected concentrations of

$PM_{2.5}$ components by meteorological factors.

Applying meteorological models over extensive domains with fine scales (e.g., 4-12 km) can be very resource intensive and present data base management problems. We suggest means for reducing such problems.

Air quality inputs are needed for initial conditions and for boundary values at the edges of a numerical model's domain. There is no satisfactory way to use available air quality observations to specify initial conditions. Thus, States should diminish their importance by beginning a simulation two or more days prior to the period of interest for urban applications and three or more days earlier for regional applications. Nested regional models are the usual preferred means for generating boundary conditions to a portion of the large regional domain which is the focus of an attainment or reasonable progress demonstration. If an urban scale model is used, the domain should be large enough so that emissions occurring in the center of the domain just before sunrise remain within the domain until the end of the same calendar day. This should reduce importance of boundary conditions specified for such applications.

If a State chooses to use a Lagrangian air quality model to support an attainment demonstration, meteorological input should be generated for a 3-year period, preferably the one used to calculate the nonattainment area's current design value. These data may be obtained from National Weather Service or equivalent sites or by using a meteorological model (with 12 km or smaller horizontal grid cells). If a Lagrangian model is used, estimates are also needed for "background" concentrations of primary particulate matter. Background estimates depend on the geographic scale over which the Lagrangian model is applied. For urban scale applications (used in the modeled attainment test), States should use observations or regional scale model estimates just outside the area modeled with the Lagrangian model. For geographically more limited hot spot analyses, background may be estimated by spatially averaging observations or modeled estimates made within a defined distance of the source(s) being reviewed.

Derivation and use of meteorological and air quality inputs to an air quality model are discussed in Section 14.0.

ES14.0 How Do I Produce Needed Emissions Inputs?

Whenever possible, States should derive current emissions estimates from locally available data. This is called a "bottom-up" approach. Often however, it may be necessary to rely on Statewide or national data, at least in some parts of the domain (i.e., it may be necessary to use a "top-down" approach).

Producing needed emissions inputs using a top-down approach requires several steps. Many of these steps need to be followed with bottom-up approaches as well. First, compile Statewide and then countywide estimates for inorganic and organic $PM_{2.5,}$ SO_2, NH_3, VOC, NOx and CO emissions for point, area, mobile and biogenic emissions. Estimates are needed for each

month of the year. Countywide estimates are also needed for PM_{10} emissions to estimate changes in coarse particulate matter required for reasonable progress determinations. Second, quality assure the outputs. Third, convert the resulting estimates into speciated, gridded hourly emissions using emissions models. Fourth, once again, quality assure the results. Finally, project gridded, speciated hourly emission estimates to a future year which corresponds to two years prior to the deadline for meeting the NAAQS or for realizing the first increment of reasonable progress toward reaching natural background visibility. We also recommend projecting emissions to an intermediate year to facilitate a subsequent "mid-course review" to see if a strategy is working or to diagnose why not.

The U.S. EPA has prepared a series of guidelines relating to these steps as a part of the Emission Inventory Improvement Program (EIIP), as well as a guideline for developing emissions inventories. States should be familiar with these guidelines. States should use the most recent emission estimates commonly available when applying the modeled attainment or reasonable progress tests. As of late 2000, the National Emissions Trends (NET) Inventory compiled for 1999 is the preferred source of information for State and countywide estimates in portions of the modeling domain for which States who are stakeholders have no better information. If NET inventory estimates for 2002 become available in a timely manner, they will be the preferred choice for use in modeled demonstrations for attainment or reasonable progress.

Different means are used to obtain emissions information in the form needed by air quality models. Ideally, location and daily/weekly emission patterns should be directly available for point sources. Spatial distribution of surrogates for activity factors needs to be estimated for major area source categories and for mobile sources in order to estimate gridded emissions. Diurnal and weekly activity patterns are also useful. Current, locally applicable $PM_{2.5}$ and VOC speciation profiles are desirable for point sources, mobile and major area source categories. Emission models are then used to characterize emissions from point sources, stationary area, mobile and biogenic sources. We identify several commonly used emissions models.

Quality assurance of emissions estimates is necessary for results of a modeled test to be credible. We recommend that quality assurance checks occur during several stages of the process needed to derive required emission inputs to an air quality model. Comparing inventory estimates made for different studies and in different States, computer graphics and comparisons with available, speciated air quality data are useful means for quality assuring emission estimates.

In the modeled tests for the annual NAAQS and regional haze, projected emissions should ordinarily reflect best estimates (i.e., projected "actual" emissions). For applications related to the 24-hour NAAQS, States should ordinarily assume projected emissions reflecting maximum feasible or permitted (i.e., "allowable") emissions, whichever is less, for a small number of sources. The number and types of sources for which this assumption is made needs to be determined case by case. Projections for all other sources should reflect best estimates of seasonally typical (i.e., "actual") emissions.

Hot spot modeling analyses may be needed in some nonattainment areas to provide assurance that the 24-hour and annual NAAQS will be met in suspect locations where there are no monitors. For analyses related to the 24-hour NAAQS, States should assume maximum feasible or permitted emissions (whichever is less) for sources subject to this review, as well as for a limited number of large, nearby sources. Number of sources for which projected "allowable" or maximum feasible emissions is used needs to be determined case by case. Typical emissions may be assumed for all other modeled sources contributing to ambient "background" for these analyses. For analyses related to the annual NAAQS, typical (i.e., "actual") emissions should ordinarily be assumed for all sources, including those which are the subject(s) of the hot spot analysis.

Generating emission inputs is addressed more fully in Section 15.0.

E15.0 How Do I Evaluate Model Performance And Make Use Of Diagnostic Analyses?

In Section 16.0, we identify the following approaches as useful for evaluating performance of an air quality model:

1. get a "big picture" assessment of discrepancies between observations and predictions using computer graphics;

2. compute various metrics reflecting comparisons between predicted and observed concentrations of PM components and gaseous precursor species;

3. compute metrics which compare predictions and observations of $PM_{2.5}$ and extinction coefficients derived from predicted vs. observed components of particulate matter;

4. compare source attribution estimates obtained with air quality models vs. those derived from monitored data using receptor models;

5. compare observations vs. predictions of PM components and $PM_{2.5}$ on weekends vs. weekdays;

6. compare predicted vs. observed ratios of indicator species, and

7. perform retrospective analyses, in which observed trends are compared with past modeled projections.

All of the preceding approaches have shortcomings. Thus, States should perform a variety of tests and view all of the results in a "mini-weight of evidence" assessment of the model's performance. Greatest weight should be given to those tests which are most closely

related to how the model is used in an attainment or reasonable progress demonstration (i.e., to compute relative reduction factors for individual components of particulate matter). Therefore, we suggest spending the most effort evaluating the model's ability to reproduce observed concentrations of components of PM and some related gaseous species (e.g., ozone, nitric acid). We identify a metric (comparison of predicted, spatially averaged component concentrations near a monitor, averaged over several days with corresponding monitored observations, averaged over the same days). This metric corresponds closely to how we compute the denominator of the relative reduction factor. We also identify several additional, potentially useful metrics.

We recommend that States gather the data and expertise to apply approaches which evaluate the model's ability to accurately predict *changes* in components of particulate matter. Thus, we identify weekend/week day and seasonal comparisons, comparisons between predicted and observed ratios of indicator species and retrospective analyses as having the potential for evaluating model performance in a most relevant way.

Diagnostic tests should be applied throughout a modeling analysis. We identify several key stages for use of these tests: (1) during model setup, (2) during model performance evaluation, (3) to study predictions at specific times and locations in greater depth, (4) during the process of choosing/comparing prospective control strategies, and (5) to help estimate uncertainty in the resulting air quality projections after tentatively selecting a strategy.

Two types of diagnostic tests are identified: sensitivity tests and process analysis. In designing and evaluating results of diagnostic tests, States should consider how models are to be used to support an attainment demonstration. That is, models should be used in a *relative* sense. Thus, diagnostic tests should consider how relative reduction factors (RRF) or other predicted changes in the components of particulate matter are affected by various factors. Tests which are useful for exploring, and perhaps improving, model performance are described in Section 16.0. Tests that are potentially useful for helping to choose an effective control strategy are identified in Section 17.0.

1.0 Introduction

1.1 What Is The Purpose Of This Document?

This document has two purposes. The first is to recommend how to use modeling/analysis to demonstrate that a control strategy will lead to air quality which meets certain national goals. Goals include attaining national ambient air quality standards (NAAQS) for particles with aerodynamic diameters $\leq$ 2.5 micrometers (μm) ($PM_{2.5}$) and making reasonable progress reducing regional haze. Final determination of whether an air quality goal for $PM_{2.5}$ or regional haze is met will be based on a review of monitored data at the appropriate future date. Our intent in this guidance is to provide as reliable an indicator as possible of what these future monitored attainment or reasonable progress tests will show. Because this is a guidance document rather than a regulation, our recommendations are not requirements. A State may use alternative procedures if it presents compelling arguments for doing so in a particular case to the appropriate U.S. Environmental Protection Agency (EPA) regional office(s) and that office agrees with the State's assessment.

The likelihood of choosing effective strategies to meet goals for $PM_{2.5}$ and regional haze is greater if the strategies are derived using high quality emissions, meteorological and air quality data. There are many desirable measurements or estimates which could be made to enhance these data bases. However, time and resources to develop data for regulatory applications are often limited. Therefore, the second purpose of this guidance is to identify how air quality, emissions and meteorological data will be used in models or other analyses to demonstrate future attainment of the NAAQS for $PM_{2.5}$ or reasonable progress reducing regional haze. By identifying how data will be used, our intent is to help States and others with an interest in emission control strategies to prioritize their data needs.

1.2 Does This Guidance Apply To Me?

This guidance applies for all States which need to submit a State implementation plan (SIP) revision to meet one or more NAAQS for $PM_{2.5}$ (U.S. EPA, 1997). It also applies for States needing to submit a SIP revision to help meet goals to reduce regional haze in "*Class I areas*" (i.e., generally, national parks or wilderness areas) subject to the regional haze rule (U.S. EPA, 1999).

Two NAAQS have been promulgated for $PM_{2.5}$: an annual standard and a 24-hour standard. The annual NAAQS is met at a monitoring site if the annual arithmetic mean $PM_{2.5}$ concentration, averaged over 3 consecutive years, is $\leq$ 15.0 $\mu g/m^3$. The 24-hour NAAQS is met at a monitoring site if the 98th percentile 24-hour concentration, averaged over 3 consecutive years, is $\leq$ 65 $\mu g/m^3$. The "$PM_{2.5}$" measurements referred to in the NAAQS must be measured using a Federal Reference Method (FRM) or an equivalent measurement technique. FRM and equivalent techniques are described in 40 CFR Part 50, Appendix L and in 40 CFR Part 53. Generally, a designated "nonattainment area" for $PM_{2.5}$ is a geographic location containing one or

more monitoring sites where observations do not meet one or both of the NAAQS for $PM_{2.5}$.

The regional haze rule requires visibility to approach "natural background" levels by 2064. It requires reasonable progress to be made toward this goal over a series of consecutive periods. The first of these periods covers 2004-2018. The guidance herein focuses exclusively on this initial period. We expect that guidance for the following periods will be developed as they become more imminent. The rule identifies two criteria which must be met in order for reasonable progress reducing regional haze to be demonstrated for a Class I area during the initial period.

1. Mean regional haze on the 20% of days with the worst visibility averaged over a consecutive 5-year period (i.e., 2000-2004) should be reduced by an agreed upon difference in *deciviews*[2] between 2000/2004 and 2018, and

2. Mean regional haze on the 20% of days with the best visibility averaged over a consecutive 5-year period (i.e., 2000-2004) should not increase.

1.3 How Does The Perceived Nature Of Particulate Matter And Tools For Describing It Affect My Use/Interpretation Of Model Results?

Guidance for demonstrating attainment of NAAQS for $PM_{2.5}$ or reasonable progress reducing regional haze needs to be consistent with the perceived nature of $PM_{2.5}$ and regional haze. In this section, we identify probable attributes of $PM_{2.5}$ in most locations of interest in attainment or reasonable progress demonstrations. There may well be exceptions. As we discuss in Section 9.0, States need to develop a conceptual description of the $PM_{2.5}$ or regional haze problem in each of their areas subject to a demonstration. If a substantially different picture emerges from the general one presented in this section, modeling/analysis procedures which differ from some of those we believe are generally applicable may be warranted.

This guidance is based on a conceptual description of $PM_{2.5}$, regional haze, the air quality goals we are trying to meet and the tools that are available to estimate whether a prospective strategy will meet the goals. This description is summarized by the following 9 premises.

1. Particulate matter is a mixture.
2. "Secondary" PM is a more important part of $PM_{2.5}$ than it is of PM_{10}.

---

[2]A deciview is a variable depicting maximum range of visibility. It is related logarithmically to a physical measurement of attenuation, the *extinction coefficient*. Deciviews increase as visibility impairment increases. They are analogous to the decibel system for sound. A deciview is calcuated by taking the logarithm of the ratio of a measured extinction coefficient to a background value for the coefficient (see Section 4.0 for further details). The scale of deciviews is set up so that changes are approximately linear to changes perceived by a human eye.

3. Regional haze is closely related to high concentrations of fine particulate matter.
4. The Federal Reference Method used to determine attainment of the NAAQS for $PM_{2.5}$ is subject to potential interferences which are difficult to measure and model.
5. Many factors affecting formation and transport of secondary particles are the same as those affecting formation and transport of ozone.
6. Spatial gradients for primary particulate matter may be more pronounced than those for secondary particulate matter or ozone.
7. Seasonal differences exist in emissions of $PM_{2.5}$ and its precursors as well as in meteorological conditions affecting source/receptor relationships.
8. Causes for violations of $PM_{2.5}$ NAAQS and poor visibility may be diverse.
9. Ability of models to predict future concentrations of $PM_{2.5}$ and its components is limited by a variety of factors.

These premises imply several things about how modeling should most appropriately be used to demonstrate whether a prospective control strategy will meet goals for $PM_{2.5}$ and/or regional haze. We discuss these in the following paragraphs.

Premise 1. Particulate matter is a mixture. Unlike a compound (e.g., ozone) or an element (e.g., lead), a mixture has components which (a) can behave independently of one another (e.g., primary vs. secondary components) or (b) are related to one another in a complex way (e.g., different secondary components). Thus, if one only considers $PM_{2.5}$ as a single entity, rather than as the sum of its major *components*, there is a greater risk of choosing an ineffective control strategy. This follows, because models may not perform equally well in predicting major components of $PM_{2.5}$. Nevertheless, balancing errors could (erroneously) indicate good model performance predicting $PM_{2.5.}$ If a control strategy focused on reducing a component of $PM_{2.5}$ which was overestimated by the model, the subsequently observed impact on $PM_{2.5}$ could be less than expected.

Characteristics of $PM_{2.5}$ as a mixture and the possibility that models perform unevenly in predicting the major components of the mixture have two important implications for our guidance. First, the modeling should divide $PM_{2.5}$ into a half a dozen or so major components and note the effects of a strategy on each. The effect on $PM_{2.5}$ should be estimated as a sum of the effects on individual components. Second, to reduce the effects of uneven performance and possible major bias in predicting absolute concentrations of one or more components, models are best used in a "relative" sense in concert with measured $PM_{2.5}$ and estimated composition of the measured $PM_{2.5}$ derived from speciated measurements. That is, responses predicted by models should be applied to observed component concentrations derived from $PM_{2.5}$ measurements and composition of $PM_{2.5}$ estimated from measurements of ambient species. Third, failure to model components explicitly robs one of much diagnostic insight. For example, secondary components of $PM_{2.5}$ (e.g., sulfates and nitrates) are not independent of one another. Under some circumstances reducing sulfates can increase nitrates. Thus, to realize the full potential of reducing sulfates, it might be necessary to reduce precursors to nitrates in this example. To consider these interrelationships, it is best to use a single air quality model to consider secondary

components of PM.

Premise 2. "Secondary" PM is a more important part of $PM_{2.5}$ than it is of PM_{10}. Size-differentiated ambient particulate data suggest that mass of particulate matter follows a bimodal distribution, with one peak (*fine mode*) reflecting particles with aerodynamic diameters ~ 0.1-1.0 micrometers arising from nucleation and accumulation phenomena, and a second (*coarse mode*) occurring with aerodynamic diameters in the range of 1.0-20 micrometers. As shown in Figure 1.1, derived from Wilson and Suh, 1997, mass of fine particulate matter (i.e., $PM_{2.5}$) attributable to coarse mode particulate matter $\leq$ 2.5 micrometers is relatively small. Mass attributable to fine mode particulate matter with aerodynamic diameters $\geq$ 2.5 micrometers (i.e., coarse particulate matter) is also relatively small.

Origins of coarse and fine mode particulate matter are quite different. The former results mostly from physical types of activities (e.g., crushing, grinding, resuspension due to motions, etc.). Nearly all these activities result in particulate matter being emitted as particulate matter, with little subsequent chemical change. We call such emissions of particulate matter "primary emissions", because they are measured in more or less the same form in which they are emitted. In contrast, origins of fine mode particulate matter are more diverse. For example, some fine mode particulate matter is directly emitted to the atmosphere as a result of combustion. Such emissions occur either directly as particles or as a result of condensation which occurs very shortly after the emissions occur. These are primary emissions, because what is measured in the ambient air is essentially unchanged (chemically) from what is released. However, many fine mode particles are the result of physicochemical reactions which occur in the atmosphere among gaseous precursors or through absorption or adsorption onto previously existing aerosols. Such particles constitute "secondary" particulate matter, because they undergo transformations in the atmosphere causing the chemical and/or physical nature of what is measured to be different from what is emitted.

Because of the size distribution of ambient particulate matter, most measured $PM_{2.5}$ is likely to be fine mode particulate matter. As a result, it is dominated to a much larger extent than PM_{10} by "secondary" particulate matter and primary particulate emissions arising from combustion. Some of the physicochemical processes leading to secondary particulate matter may take several days, as do some of the removal processes. Most require good mixing between emissions from a source and its environment. Thus, many of the sources of measured secondary particulate matter may not be in the immediate vicinity of a measured concentration of $PM_{2.5}$. This implies that modeling to support *attainment demonstrations* for $PM_{2.5}$ (and, as we will discuss later, regional haze-related applications) will need to cover a very large domain, and will need to include chemical/physical mechanisms important in formation/removal of secondary particulate matter. Because several of the processes are slow and first require thorough mixing with the environment, spatially detailed treatment near area or mobile sources of precursors may not be necessary. Individual treatment of precursor emissions from relatively nearby large sources of combustion may be needed on a case by case basis.

Premise 3. Regional haze is closely related to presence of high concentrations of fine particulate matter. Light extinction results from scattering and absorption of light. Some scattering occurs by gas molecules in pristine air (i.e., Rayleigh scattering). Nearly all remaining light extinction is caused by presence of aerosols. For any given mass, fine particles (i.e., $\leq$ 2.5 μm) are more efficient at scattering light than are particles > 2.5 μm aerodynamic diameter. Further, certain components of $PM_{2.5}$ are more efficient at scattering or absorbing light than others. Many of the most efficient are secondary particulate species. For example, sulfates (secondary), nitrates (secondary) and organic (secondary and primary) components scatter light more efficiently than do small ($\leq$ 2.5 μm) coarse mode (i.e., primary) particles (e.g., composed of crustal material). Light extinction is exacerbated by high relative humidity. Water vapor combines with hygroscopic particulate matter (e.g., sulfates and nitrates) to greatly increase the light scattering efficiency of these species. Previously, we noted that secondary particulate matter is likely to comprise an important fraction of measured $PM_{2.5}$. Secondary particulate matter will be even more important as a cause of regional haze. This follows from the greater efficiency with which these already important components of $PM_{2.5}$ scatter light. This importance is enhanced still further by high relative humidity, which is especially relevant in the Eastern U.S.

The discussion in the preceding paragraph suggests that modeling to assess reasonable progress reducing regional haze will need to address secondary particulate matter. This, in turn, means that large modeling domains will be necessary. Secondary particles arise from source plumes being thoroughly mixed with their environment over a period of hours or days. Further, Class I areas are generally likely to be far removed from most sources of precursors for secondary particulate matter. Finally, the measure of visibility (deciviews) with which we are most concerned addresses maximum range of visibility. This measure reflects an effect which is an integrated one over a relatively large distance. Thus, the need for a large domain, prerequisite mixing for secondary particles, the relative remoteness of Class I areas from most sources of precursors and the visibility measure of greatest interest suggest that modeling related to reasonable progress may be done without a fine degree of spatial resolution.

Premise 4. Sampling anomalies associated with the Federal Reference Method for $PM_{2.5}$ are difficult to consider with an air quality model. A NAAQS needs to be related to observed health effects. In order to establish this link most clearly, the U.S. EPA has adopted a Federal Reference Method (FRM) for measuring $PM_{2.5}$ similar to procedures used in epidemiological studies in which adverse health effects are associated with exposure to $PM_{2.5}$. However, the sampling protocol inherent with the FRM leaves some uncertainty about the water content of material measured on the filter, as well as measured semi-volatile organic particulate matter and nitrates. At present, it is not possible to adjust a model prediction to take account of positive or negative sampling artifacts inherent in the FRM. Thus, one might *expect* some disagreement between unadjusted model estimates and monitored concentrations of $PM_{2.5}$. This has several implications regarding how models might best be used to support attainment demonstrations for the $PM_{2.5}$ NAAQS. First, when evaluating model performance, it is important to assess how accurately the model is able to reproduce observed components of particulate matter (i.e.,

measured using different, often less uncertain, methodologies). Second, there may well be a difference between the sum of measured components and the gravimetric mass of $PM_{2.5}$ measured with the FRM.. This difference needs to be accounted for in some way in a modeled test for attainment of the NAAQS. Since the FRM is included in the definition of the NAAQS, the modeled attainment test we will recommend starts with measured concentrations of $PM_{2.5}$ as the "ground truth". Models are used to assess how these measurements are likely to respond to growth plus control measures applicable at some future date. Thus, we believe models are most suitable for predicting relative changes in particulate matter and its components rather than using them by themselves to make statements about future absolute concentrations of $PM_{2.5}$.

Premise 5. Many of the factors affecting formation and transport of secondary particles are the same as those affecting formation and transport of ozone. Although there is not necessarily a positive correlation between measured ozone and secondary particulate matter, many of the same factors affecting concentrations of ozone also affect concentrations of secondary particulate matter. For example, similarities exist in sources of precursors for ozone and secondary particulate matter. Sources of NOx may lead to formation of nitrates as well as ozone. Sources of VOC may also be sources or precursors for organic particles. Presence of ozone itself may be an important factor affecting secondary particulate formation. For example, as ozone builds up, hydroxyl (OH) radicals do also as a result of equilibrium reactions between ozone, water and OH in the presence of sunlight. Hydroxyl (OH) radicals are instrumental in oxidizing gas phase SO_2 to sulfuric acid, which eventually gets absorbed by liquid aerosol and converted to particulate sulfate in the presence of ammonia. Hydroxyl radical and NO are also precursors for gas phase nitric acid, which gets absorbed by liquid aerosol and, in the presence of ammonia, leads to particulate nitrate. Chemistry of secondary particulate matter and its relationship to that for ozone are described in the U.S. EPA Criteria Document for Particulate Matter (U.S. EPA, 2000).

Strategies to reduce ozone can also affect formation of secondary particulate matter. A few of the more obvious examples are reducing VOC emissions could reduce ozone and OH. If sulfate or nitrate production is limited by lack of availability of oxidizing agents, the ozone reduction strategy could also reduce secondary particulate matter. Reducing NOx emissions diminishes one of the precursors for nitric acid (i.e., NO_2 which results from NO). Therefore, in the presence of sufficient ammonia, reducing NOx emissions could reduce particulate nitrate concentrations. There are also more subtle interfaces between strategies to reduce ozone and secondary particulate matter. For example, reducing NOx in the presence of substantial particulate sulfates and lack of sufficient ammonia could exacerbate the particulate sulfate problem, or reducing SO_2 in the presence of substantial NOx and ammonia could exacerbate the particulate nitrate problem.

The preceding discussion implies that models intended to address secondary particulate matter problems need also to be capable of simulating ozone formation and transport and related factors. It also suggests that ability to predict ozone is a first step which States should consider in evaluating performance of models for secondary particulate matter. Finally, it is clear from the

discussion that States should include previously implemented or contemplated measures to reduce ozone in their air quality model applications to predict secondary particulate matter.

Premise 6. Spatial gradients for primary particulate matter may be more pronounced than those for secondary particulate matter or those for ozone. As previously noted, secondary particulate matter and ozone result from an interaction of meteorology and chemistry taking several hours to days. Because of the time scales and mixing required, many sharp spatial gradients in concentrations of these pollutants are unlikely. In contrast, primary particulate matter is emitted in the form it appears at monitoring sites. It is likely that concentrations of primary particulate matter are greatest near major source areas of primary particulate matter.

The preceding implies that it is necessary to estimate concentrations of primary particulate matter using models with finer spatial resolution than is necessary for secondary particles. Further, there may be several sources or concentrations of sources of primary particulate matter within an area designated as "nonattainment" for $PM_{2.5}$. It is unlikely that there will be monitored $PM_{2.5}$ data near each such location. The guidance will need to address how to evaluate model performance and how to estimate whether attainment of the NAAQS is likely in such locations.

Premise 7. Seasonal differences are likely in emissions of $PM_{2.5}$ and its precursors, as well as in meteorological conditions affecting source/receptor relationships. Emissions from several potentially important sources of $PM_{2.5}$, such as residential wood burning, agricultural burning, prescribed forestry burns and biogenic sources have distinctive seasonal patterns. Further, meteorological factors which may affect $PM_{2.5}$ or regional haze, such as relative humidity, sunlight intensity, mixing heights, precipitation and temperature, have marked seasonal differences in many parts of the United States. The annual NAAQS for $PM_{2.5}$ and the goals for regional haze address a composite of conditions measured over many days. To understand how such composites respond to changes in emissions, it will be necessary to model a variety of days with varying emissions and meteorological conditions. This implies that States will need to develop base emissions estimates for each of four seasons and that a representative portion of days may need to be modeled from each season.

Premise 8. Causes of $PM_{2.5}$ concentrations which violate NAAQS may be diverse. Modeling approaches needed to address primary vs. secondary particulate matter differ in their requirements. Earlier, we noted differing requirements for size and resolution of the modeling domain. Another difference is the need to consider atmospheric chemistry in the modeling. It is essential to have some understanding of the nature of an area's $PM_{2.5}$ or visibility problem *before* modeling begins. Otherwise, a State runs the risk of selecting inappropriate analysis tools as well as selecting a strategy which will prove to be ineffective at reducing its problem. Thus, a State needs to perform data analysis before using air quality models. This analysis should be used to develop a *conceptual description* of the problem at hand. The conceptual description may then be used to select a broad strategy (e.g., do I focus on reducing primary or secondary particulate matter or both?) as well as to help implement a modeling protocol to best address the nature of

the problem and the qualitative strategy which has been tentatively selected to address it. The guidance needs to provide States with flexibility in choosing model(s) to address specific problems.

Premise 9. Ability of models to predict future concentrations of the components of $PM_{2.5}$ is limited by a variety of factors. Our ability to characterize emissions on a day to day basis or on a source-specific basis is limited. Fully characterizing meteorological conditions on any given day is also problematic. Further, for regulatory models to be tractable, they must characterize chemical and physical processes by simplifying them in some reasonable manner. In some cases, most notably for secondary organic particulate matter, the extent to which current simplifications are reasonable is uncertain. These limitations (and others) make the ability of a model to accurately predict concentrations of $PM_{2.5}$ and its components at a given time and location doubtful.

The preceding paragraph has several implications for using models to demonstrate future attainment of a NAAQS for $PM_{2.5}$ or reasonable progress goals for regional haze. It suggests that we should focus on composite responses of the model averaged over several days to help circumvent the problem of not knowing all of the details on an individual day. This composite response then needs to be related to the form of the air quality goal in some manner. Limitations in available models and their underlying data bases also suggest that the guidance should recognize a need for performing other, corroboratory analyses to confirm conclusions reached with a model. Such corroboratory analyses should include periodic reviews of air quality data, implementation schedules, meteorological measurements, etc. (e.g., one or more "*mid-course reviews*") using updated data bases and modeling tools (if available) so that a strategy can be adjusted, if warranted.

Premises 1-9 represent a kind of "conceptual description" of likely $PM_{2.5}$ and regional haze problems viewed from a national perspective, as well as the ability of models to deal with these problems. Guidance in subsequent chapters reflects needs and problems raised in this conceptual description.

1.4 What Topics Are Covered In This Guidance?

Part I (Sections 2.0-7.0) of this document describes how results of air quality models and other analyses should be used to see if a proposed emissions control strategy will lead to attainment of the NAAQS for $PM_{2.5}$ and/or reasonable progress reducing regional haze. Section 2.0 provides an overview of a modeled *attainment demonstration* for the annual and 24-hour NAAQS for $PM_{2.5}$. This demonstration consists of a *modeled attainment test* and, if a State so chooses, a *weight of evidence determination.* Section 2.0 also summarizes the modeled test for reasonable progress reducing regional haze. Section 3.0 describes attainment tests for each of the two NAAQS for $PM_{2.5}$. Section 4.0 outlines a recommended test to assess whether a proposed control strategy will lead to air quality changes which meet reasonable progress goals to reduce regional haze. Section 5.0 explains how weight of evidence may be used to supplement the

attainment test in an attainment demonstration or the modeled test for reasonable progress reducing regional haze. Section 6.0 identifies key data gathering priorities and related activities which could lead to better estimates of air quality for a SIP revision and in subsequent reviews. Section 7.0 identifies documentation which should accompany a SIP revision in order to adequately describe the basis for assuming that the selected strategy demonstrates attainment or reasonable progress.

Part II (Sections 8.0 - 18.0) of this document identifies key data needed in modeling which supports the analyses described in Part I. Part II also describes how the data should be used in models and other analyses to generate results used in attainment or reasonable progress demonstrations. Section 8.0 provides an overview of the steps needed to produce credible air quality modeling results. Each of the major steps is outlined in the following sections. Section 9.0 discusses why a conceptual description is needed of each area's $PM_{2.5}$ or regional haze problem, and explains how such a description should be developed. Section 10.0 addresses the need to develop a modeling protocol. The protocol translates implications of a conceptual description into a modeling/analysis plan to evaluate effects of proposed control strategies. Section 11.0 describes factors which a State should consider in selecting one or several models to address goals for $PM_{2.5}$ and regional haze. Section 12.0 provides guidance on choosing suitable periods to model when addressing the annual and 24-hour NAAQS for $PM_{2.5}$ and goals to reduce regional haze. Section 13.0 covers choice of modeling domains and the degree of horizontal and vertical spatial detail needed for the types of model applications we discuss. Section 14.0 indicates procedures which are appropriate for generating meteorological and air quality inputs to models used to support demonstrations for $PM_{2.5}$ and regional haze. Section 15.0 identifies needed emissions data and describes how available data should be converted to a form needed by air quality models. A discussion on estimating future emissions is also included. Section 16.0 identifies a series of tests which are potentially useful for assessing whether a model's performance is adequate for it to be used to support an attainment or reasonable progress demonstration. Section 17.0 identifies several procedures which may be useful for identifying control strategies which are effective in meeting air quality goals for $PM_{2.5}$ or regional haze. Section 18.0 lists references cited in our guidance document.

We have tried to make this guidance comprehensive. However, situations which we have failed to anticipate will undoubtedly occur. Further, we expect that our knowledge about fine particulate matter will evolve as data bases improve and the user community gains more experience with the use of models for particulate matter. Thus, we recognize that States may find it appropriate to deviate from the default recommendations in this guidance. Indeed, the guidance itself is likely to be updated from time to time as our knowledge and experience regarding fine particulate matter, regional haze and related analysis methods improves. States should coordinate their approaches for demonstrating future attainment of air quality goals for $PM_{2.5}$ and regional haze with the appropriate U.S. EPA regional office(s) to reach consensus on the most appropriate procedure(s) for the case at hand.

Part I

How Do I Use Model Results To See If An Air Quality Goal Is Met?

(This page is left blank intentionally)

2.0 What Is A Modeled Demonstration?–An Overview For The $PM_{2.5}$ NAAQS And Regional Haze Goals

Modeled attainment demonstrations meeting goals for $PM_{2.5}$ and regional haze consist of (a) analyses which estimate whether selected emissions reductions will result in ambient concentrations that meet the NAAQS or reasonable progress goals, and (b) an identified set of measures which will result in the required emissions reductions. This guidance focuses on the first component of an attainment demonstration--interpretation and conduct of analyses to estimate the amount of emission reduction needed to reduce concentrations of $PM_{2.5}$ or its components to levels consistent with meeting the NAAQS or reasonable progress goals. Emission reduction strategies should be simulated by reducing emissions from specific source categories rather than through broad "across-the-board" reductions from all sources.

States should estimate the amount of emission reduction needed to demonstrate attainment the NAAQS by using a modeled attainment test. In some cases, where there are sources which are major emitters of primary particulate matter and no nearby air quality monitoring, States should also perform a "hot spot analysis" for the source(s) in question. We recommend a similar modeled test for demonstrating reasonable progress toward reducing regional haze to natural background levels. In addition to the modeled tests and hot spot analysis (for NAAQS-related applications), a State may consider a broader set of corroboratory analyses to determine whether the "weight of evidence" indicates that a proposed emission reduction will lead to attainment of the NAAQS or reasonable progress goals.

2.1 What Is The Recommended Modeled Attainment Test?–An Overview

The *modeled attainment test* for the annual or 24-hour NAAQS needs to be applied if there is a violation of the corresponding NAAQS within or nearby a designated nonattainment area. The test we recommend reflects $PM_{2.5}$'s nature as a mixture. This is done by dividing monitored $PM_{2.5}$ into major components. In the test, these are

- mass associated with sulfates
- mass associated with nitrates
- mass associated with organic carbon
- mass associated with elemental carbon
- mass associated with primary inorganic particulate matter other than primary sulfate and nitrate particles
- mass in a catchall category reflecting the difference in measured mass of $PM_{2.5}$ and the sum of the mass from the other components of $PM_{2.5}$.

In the recommended modeled attainment test for the annual or 24-hour NAAQS, an air quality model is used to simulate current and future air quality. Our recommended test uses model estimates in a "relative" rather than "absolute" sense. That is, we take the ratio of the model's future to current predictions of the components of $PM_{2.5}$ at FRM or equivalent $PM_{2.5}$

monitoring sites. We call each of these site-specific ratios, component-specific *relative reduction factors*. Future $PM_{2.5}$ design values are estimated at existing monitoring sites by multiplying modeled relative reduction factors "near" each monitor times the observed "component specific design value". This latter quantity is estimated using measured site-specific design values for $PM_{2.5}$ in concert with available measured composition data. Future site-specific $PM_{2.5}$ design values at a site are estimated by adding the six future "component specific design values" computed at the site. If the predicted future design value for $PM_{2.5}$ is ≤ 15.0 $\mu g/m^3$ (applications related to the annual NAAQS) or ≤ 65 $\mu g/m^3$ (applications related to the 24-hour NAAQS), the test is passed. If all such future site-specific $PM_{2.5}$ design values are # the concentration specified in the NAAQS, the test is passed.

2.2 What Is A Hot Spot Analysis And Why Is It Needed?

The modeled attainment test we recommend predicts whether or not all observed future $PM_{2.5}$ design values will be less than or equal to the concentration level specified in the NAAQS. By itself, the test makes no statement about future $PM_{2.5}$ at locations where there is no nearby monitor. Secondary particulate matter requires mixing and hours or days to materialize. Further, the long sampling time (i.e., 24 hours) and averaging time inherent in the NAAQS is also likely to reduce spatial gradients in $PM_{2.5}$. Nevertheless, we recognize that concentrations of $PM_{2.5}$ which are substantially higher than measured design values could exist near major sources of primary $PM_{2.5}$ emissions. Thus, we recommend States use a supplementary "hot spot" analysis to identify other locations where passing the test might be a problem if monitoring data were available.

Like the modeled attainment test itself, the hot spot analysis is described more fully in Section 3.0. Briefly however, it entails the following two steps.

- The first step is to identify locations where a hot spot analysis may be needed. This may be done in one of two ways. The first approach is to review the spatial distribution of estimated primary $PM_{2.5}$ emissions for grid cells within a designated nonattainment area for $PM_{2.5}$. Note the location of grid cells in which primary emissions (i.e., from point, area and mobile sources) are well above (e.g., > 20% higher) those "near" any monitoring site for $PM_{2.5}$ within the designated nonattainment area. Major sources within such cells are candidates for a hot spot analysis. The second approach for flagging major sources subject to a hot spot analysis makes use of available results from air quality modeling. In this approach, a State should flag major sources in any grid cell for which predicted concentrations of primary particulate matter are substantially (e.g., > 20% greater) and consistently (e.g., $\geq$ 50% of the days modeled) greater than predicted concentrations of primary particulate matter near any site monitoring $PM_{2.5}$.

- Apply a point/area/line source model to flagged sources for 3 consecutive years to estimate a projected future highest 3-year average annual mean concentration and/or highest 3-year average 98^{th} percentile 24-hour concentration of primary particulate matter

near each flagged source. Add each of these estimates to concentrations estimated from interpolating future $PM_{2.5}$ annual mean concentrations obtained with the modeled attainment test for the annual NAAQS. If the resulting estimate exceeds the level specified in the NAAQS, remedial measures are needed.

2.3 What Is The Recommended Modeled Test For Reasonable Progress?—An Overview

The recommended modeled test for reasonable progress reducing regional haze is similar conceptually to the recommended tests for the two NAAQS for $PM_{2.5}$. Models are used to develop relative reduction factors for each of 6 components of particulate matter between a base period (2000-2004) and a future 5-year period which will be reviewed in 2018. Components used for regional haze-related applications differ slightly from those used for NAAQS-related applications. They are:

- mass associated with sulfates;
- mass associated with nitrates;
- mass associated with organic carbon;
- mass associated with elemental carbon;
- mass associated with primary inorganic particulate matter other than primary sulfate and nitrate particles (i.e., crustal material);
- mass associated with coarse particulate matter (i.e., PM_{10} - $PM_{2.5}$).

Current speciated measurements in a *Class I area* are used in an empirically derived equation to estimate light extinction for each day with measurements. Days are ranked according to their resulting light extinction coefficients. This ranking is used to identify the 20% of days with worst and 20% days with best visibility during each year in the base period. The 20% worst and best days are examined to estimate appropriate observed concentrations for the components of PM on "best" and "worst" days.

Observed component concentrations are multiplied by the corresponding relative reduction factors to estimate future concentrations for each component on "best" and "worst" days. Future component concentrations are then inserted in the equation relating light extinction to concentrations of particulate matter. The resulting estimates for future light extinction on "best" and "worst" days are compared with observations made during the base period to see whether the 2 goals for reasonable progress are realized.

2.4 What Does A Recommended Weight Of Evidence Determination Consist Of?--An Overview

As we note later in Section 9.0, States should perform complementary analyses of air quality, emissions and meteorological data, and consider available modeling outputs other than the results of the $PM_{2.5}$ attainment test, related hot spot analyses or reasonable progress tests.

Such additional analyses are used to develop a conceptual description of an area's problem. A conceptual description is instrumental for guiding an air quality modeling application. Sometimes, States may choose to use results of corroboratory analyses in a *weight of evidence determination* to conclude that attainment of the air quality goal is likely despite modeled results which do not quite pass the attainment or reasonable progress test or for which a hot spot analysis suggests remedial measures are needed to meet a NAAQS. If a conclusion differs from the outcome of the modeled tests, the need for subsequent review (several years hence) with more complete data bases is increased.

In a weight of evidence (WOE) determination, States should review results from several diverse types of analyses, including results from the modeled attainment or reasonable progress test and, if applicable, one or more hot spot determinations. States should next note whether or not results from each of these analyses support a conclusion that the proposed strategy will meet the air quality goal. States should then weigh each type of analysis according to its credibility as well as its ability to address the question being posed (e.g., is the strategy adequate for meeting the $PM_{2.5}$ NAAQS by a defined deadline?). Next, conclusions derived in the two preceding steps are combined to make an overall assessment of whether meeting the air quality goal is likely. This last step is a qualitative one involving some subjectivity. If it is concluded that a strategy is inadequate to demonstrate attainment or reasonable progress, a new strategy is selected for review, and the process is repeated. States should provide a written rationale documenting how and why the conclusion is reached regarding adequacy of the final selected strategy.

Results obtained with air quality models are an essential part of a weight of evidence determination and should ordinarily be very influential in deciding whether the NAAQS or reasonable progress goals will be met. This follows from including *ability to address the question being posed* as one of two criteria for weighing results from different analyses and from a model's ability to integrate information from scientific theory and observed data. For example, in NAAQS-related applications, if the modeled attainment test is passed and the hot spot analyses indicate no potential problem, this supports a hypothesis that the strategy is adequate to meet the NAAQS. This information is included as one of several elements in a weight of evidence determination to assess the strategy's adequacy. The further model results are from passing the modeled attainment test, the more compelling contrary results from other analyses have to be for a control strategy to demonstrate attainment. If either the modeled attainment test or a hot spot analysis produces results far in excess of the concentration specified in the NAAQS, it is doubtful that other evidence will be sufficiently convincing to conclude that the NAAQS will be attained. States should ordinarily consider a revised control strategy.

2.5 Why Should A Model Be Used In A "Relative" Sense And Why May Corroboratory Analyses Be Used In A Weight Of Evidence Determination?

The procedures we recommend for using model results to address problems related to $PM_{2.5}$ and regional haze differ from those used in the past to address attainment demonstrations

for PM_{10} and for total suspended particulate matter (TSP). Major differences include using models in a relative sense in concert with monitored data, estimating changes in mass of particulate matter as the sum of estimated changes in its major components, and recognizing use of supplementary analyses in a weight of evidence determination in a more formal way. There are several reasons why we believe these differences are appropriate.

1. The definitions of the air quality goals and the importance of secondary particulate matter supports using such a test. Both NAAQS focus on a concentration which is averaged over 3 consecutive years. The goals for regional haze focus on mean worst and best visibility averaged over 5 consecutive years. Such goals present difficulties using models which address secondary particulate matter. Non-linearities which may exist between secondary particulate matter and its precursors mean that models addressing this important part of $PM_{2.5}$ and light extinction are likely to be resource intensive. This, in turn, means that it will probably not be feasible to model 3 or 5 consecutive years to estimate the effects of a strategy on meeting the goals. The problem of relating model results to the NAAQS is reduced by using monitored design values as an inherent part of the modeled attainment test. The monitored design values are calculated consistently with the form of the NAAQS, and serve to "anchor" a model response to the form of the NAAQS. Problems relating model results to the goals for regional haze can be addressed in a similar fashion.

2. $PM_{2.5}$ consists of a diverse mix of primary and secondary components. This raises a concern about a model's potential inability to correctly predict values for each component which are proportional to the observed mix of components. Failure to predict major measured components of $PM_{2.5}$ in the correct proportion increases the possibility of choosing ineffective control strategies on the basis of incorrect model predictions. This possiblity is reduced if the model responses are instead applied to components of $PM_{2.5}$ which are derived from measurements.

3. Starting with an observed rather than modeled concentration as the base value subject to improvement reduces problems in interpreting model results. This is especially true for air quality goals, like the NAAQS, in which an absolute value is established as the goal. If one relies on absolute model predictions, interpretive difficulties result if the model under (or over) predicts an observed daily or annual concentration. For example, if a 24-hour $PM_{2.5}$ concentration of 70 $\mu g/m^3$ were observed and a model predicted 80 $\mu g/m^3$ on that day, should the target for the day, nevertheless, be 65 $\mu g/m^3$? Although good model performance remains a prerequisite for using a model in an attainment demonstration, problems posed by disagreements which are within acceptable bounds are reduced by the new procedure.

4. Model results and projections will continue to have associated uncertainty. The procedure we recommend recognizes this by allowing use of modeling plus other analyses to determine whether weight of available evidence supports a conclusion that a proposed emission reduction will suffice to meet the NAAQS or goals for regional haze.

5. Focusing the modeled attainment test on monitoring sites could result in control targets which are too low if the monitoring network is limited or poorly designed. We recommend supplementing the recommended attainment test with a hot spot analysis performed for major point sources of primary particulate matter not having a nearby monitoring site. This exercise provides a means for remedial actions despite passing the modeled attainment test. Because models for individual sources of primary particulate matter are significantly less resource intensive than models needed for the modeled attainment test, it should be possible to apply these models in a way which is consistent with the form of the NAAQS for $PM_{2.5}$.

<u>Recommendations</u>. States should estimate emission reductions needed to demonstrate attainment using a modeled attainment test which focuses on sites measuring $PM_{2.5}$ with Federal Reference or equivalent methods . The modeled attainment test specifically addresses 6 major components of measured $PM_{2.5}$. States should use model predictions in a relative sense to compute a relative reduction factor for each component associated with a strategy. These factors should be multiplied by "component specific design values" derived from speciated measurements and $PM_{2.5}$ measurements at different monitors. A site specific future $PM_{2.5}$ design value is the sum of effects of a strategy estimated for each of the 6 major components of measured $PM_{2.5}$. If each estimated future site-specific design value for $PM_{2.5}$ is $\leq$ the concentration specified in the NAAQS, the test is passed.

States should review emissions and available modeled data to see if the modeled attainment test needs to be supplemented by additional modeling performed near major sources of primary particulate matter. If a hot spot analysis is performed, the impact of an individual source, when superimposed on a background value applicable to the designated nonattainment area should be $\leq$ the concentration specified in the NAAQS.

We recommend that States use a modeled test for reasonable progress reducing regional haze which is similar to the recommended NAAQS attainment test. States should use a model to estimate relative changes in 6 types of particulate matter which have been previously related to visibility extinction. These relative reduction factors should be used to estimate future concentrations for each type of particulate matter. The resulting future concentrations are used to derive future values for the mean 20% worst and mean 20% best extinction coefficients. Future extinction coefficients should be compared to current ones to determine if the goals for reasonable progress have been met.

Prior to applying the modeled tests, States should undertake complementary analyses of air quality, meteorological and emissions data. These additional analyses are needed to design and focus modeling which underlies the modeled attainment or reasonable progress tests. Provided results of the modeled tests and (if applicable) hot spot analyses are not failed by a wide margin, States may also use

evidence produced by corroborative analyses together with results of the tests in a weight of evidence determination. A weight of evidence analysis may be used either to increase or decrease emission reductions identified by the modeled tests as necessary to meet the NAAQS or goals for regional haze.

This page is left blank intentionally

3.0 What Are The Modeled Attainment Tests For The Two $PM_{2.5}$ NAAQS?

Monitored data will ultimately provide the basis for determining whether the two NAAQS for $PM_{2.5}$ have been attained. The intent of this guidance is to identify a modeled attainment test which is a good predictor of the monitored attainment test's future outcome. Accordingly, our recommended modeled attainment tests for the annual and 24-hour NAAQS attempt to closely mimic the corresponding monitored attainment tests. In this section, we begin by illustrating the monitored tests for the annual and 24-hour NAAQS. We then recommend a modeled attainment test for the annual NAAQS. This is followed by recommending a modeled attainment test for the 24-hour NAAQS. We next identify how to derive needed inputs from available modeled data for the two modeled attainment tests and highlight some implications that result from the recommended tests. We conclude the section by describing a hot spot analysis which may be needed for unmonitored locations near one or more major sources of primary particulate matter.

By definition, the annual NAAQS is met if, over a consecutive three year period, the average arithmetic mean concentration of $PM_{2.5}$ is less than or equal to 15.0 $\mu g/m^3$. 40CFR Part 58, Appendix N stipulates that the annual arithmetic mean at a monitoring site is to be calculated by averaging the four quarterly arithmetic mean concentrations observed during a calendar year at the site. Generally, the annual mean concentration averaged over three years must be $\leq$ 15.0 $\mu g/m^3$ at all monitoring sites. An exception occurs however at sites meeting criteria for spatial averaging[3]. For such sites, the NAAQS is met if the 3-year average of the average annual arithmetic means of all sites so lumped together is $\leq$ 15.0 $\mu g/m^3$. Tables 3.1 and 3.2 illustrate how to use monitored data to determine if the annual NAAQS for $PM_{2.5}$ is met. A more complete description of data handling conventions for the $PM_{2.5}$ NAAQS can be found in U.S. EPA (1999a).

Table 3.1. Monitoring Data In Area "A" (with no sites qualifying for spatial averaging)

(1)	(2) Arithmetic Mean, Year 1, $\mu g/m^3$	(3) Arithmetic Mean, Year 2, $\mu g/m^3$	(4) Arithmetic Mean, Year 3, $\mu g/m^3$	(5) Average Arithmetic Mean, $\mu g/m^3$
Monitor 1	**18.0**	**14.0**	**15.0**	***15.7***
Monitor 2	**16.0**	**14.0**	**14.0**	***14.7***
Monitor 3	**12.0**	**13.0**	**14.0**	***13.0***

[3] See 40 CFR Part 58, Appendix D, Section 2.8 for a description of criteria needed for sites to qualify for spatial averaging to determine compliance with the annual NAAQS for $PM_{2.5}$.

Table 3.2. Monitoring Data In Area "B" (having two clusters of spatially averaged monitors plus two additional monitoring sites which are not spatially averaged)

(1)	(2) Arith.Mean Year 1, $\mu g/m^3$	(3) Arith.Mean Year 2, $\mu g/m^3$	(4) Arith.Mean Year 3, $\mu g/m^3$	(5) Average Arith.Mean, $\mu g/m^3$	(6) Spatially Averaged Average Arith.Mean, $\mu g/m^3$
Monitor 1	**10.0**	**12.0**	**11.0**	***11.0***	**NA**
Monitor 2	**14.0**	**12.0**	**10.0**	***12.0***	**NA**
Monitor 3 (in cluster 1)	**17.0**	**14.0**	**17.0**	**16.0**	
Monitor 4 (in cluster 1)	**13.0**	**12.0**	**12.0**	**12.3**	
Cluster 1 Spatial Average	**15.0**	**13.0**	**14.5**	**NA**	***14.2*** (= (15.0+13.0+14.5)/3)
Monitor 5 (in cluster 2)	**11.0**	**12.0**	**10.0**	**11.0**	
Monitor 6 (in cluster 2)	**15.0**	**14.0**	**14.0**	**14.3**	
Monitor 7 (in cluster 2)	**15.5**	**14.0**	**14.5**	**14.7**	
Cluster 2 Spatial Average	**13.8**	**13.3**	**12.8**	**NA**	***13.3***

To determine whether the NAAQS is met, we compare each of the italicized numbers in column (5) with 15.0 micrograms/m^3. Note that the NAAQS is not met in Area A, because the average arithmetic mean at one of the sites, monitor 1, is greater than 15.0 $\mu g/m^3$.

The NAAQS is met in Area "B". This follows since the average arithmetic means at sites 1 and 2 are not greater than 15.0 $\mu g/m^3$, and the spatially averaged average means for sites in cluster 1 and for sites in cluster 2 are also no greater than 15.0 $\mu g/m^3$ (i.e., all the italicized numbers in the table are $\leq$ 15.0 $\mu g/m^3$). Note that the NAAQS is met in Area "B", even though the average annual arithmetic mean observed at site 3 is greater than 15.0 $\mu g/m^3$. This follows,

because site 3 is one which qualifies to be treated as part of a spatial average (i.e., with site 4).

The 24-hr NAAQS for $PM_{2.5}$ is met if the 98th percentile 24-hour average concentration of particulate matter with aerodynamic diameter $\leq$ 2.5 micrometers, averaged over three consecutive years, is $\leq$ 65 $\mu g/m^3$.[4] The test applies at *all* monitoring sites--spatial averaging is not allowed for the 24-hour NAAQS. The "98th percentile" concentration (and, thus, the design value concentration) depends on the number of days on which $PM_{2.5}$ is monitored during a year. Table 3.3 is derived from information presented in 40CFR Part 50, Appendix N. The table shows how to determine the 98th percentile concentration for a given year.

Table 3.3. Determining the 98th Percentile from Number of Samples Collected in a Year

(1) Number of Samples Collected in a Year	(2) Value Corresponding to 98th Percentile for the Year
1 - 100	Highest Observed Concentration
101 - 150	2nd Highest Concentration
151 - 200	3rd Highest Concentration
201 - 250	4th Highest Concentration
251 - 300	5th Highest Concentration
301 - 350	6th Highest Concentration
351 - 366	7th Highest Concentration

Thus, if valid samples were collected at a monitoring site once every 3 days (i.e., ~ 122 valid samples/year), our recommended modeled attainment test would examine whether the 2nd highest *observation* (averaged over 3 years) at the site becomes $\leq$ 65 $\mu g/m^3$ as a result of a simulated control strategy.

3.1 What Is The Recommended Modeled Attainment Test For The Annual NAAQS?

In order to perform the recommended modeled attainment test, States should divide

[4]See 40CFR Part 50, Appendix N for a discussion of data completeness in applying the monitored attainment test.

observed mass concentrations of $PM_{2.5}$ into 6 components[5]:

- mass associated with sulfates (SO_4);
- mass associated with nitrates (NO_3);
- mass associated with organic carbon (OC);
- mass associated with elemental carbon (EC);
- mass associated with inorganic particulate emissions, excluding primary sulfate and nitrate particles (IP);
- unattributed mass (i.e., the difference between measured $PM_{2.5}$ and the sum of the other 5 components) (U).

To apply the test, States must first have run an air quality model at least twice--to simulate current emissions and to simulate the net effects of a proposed control strategy and growth projected to two years prior to the required attainment date.[6] We recommend a modeled attainment test which has 6 steps.

Step 1. Compute *observed* quarterly mean $PM_{2.5}$ and quarterly mean composition for each monitor.

For each monitoring site, compute quarterly arithmetic mean values for $PM_{2.5}$, averaged over 3 years, from monitored observations. For each day having corresponding speciated $PM_{2.5}$ data, determine the fraction of $PM_{2.5}$ comprised of mass associated with SO_4, NO_3, OC, EC, IP and U. Using the resulting day-specific fractions, estimate observed quarterly average fractions for each of the 6 components. For each monitoring site, multiply the averaged quarterly mean observed $PM_{2.5}$ concentration times the quarterly average fractions for each of the 6 components. This yields the "observed" average quarterly concentration for each of the 6 major components of measured $PM_{2.5}$.

Note that we recommend a procedure which differs slightly from that in U.S. EPA, (1999a). That is, for modeling purposes, States should average observations over a 3-year period on a quarter by quarter basis rather than on an annual basis. Our recommendation recognizes that not every day with measured $PM_{2.5}$ is likely to have speciated data needed to derive day-specific component fractions. However, quarterly mean compositional data (even if drawn from a more

---

[5]Some of these components are not independent of one another. The model(s) whose results are used in our recommended tests needs to take account of these interdependencies.

[6]"Current emissions" are those corresponding to the period reflected by the monitored design value used in the test. Regarding the recommended projection year, air quality used to assess attainment in the future will be based on 3 years of observations. Emissions need to be consistent with a projected attainment concentration throughout the 3-year period to increase the likelihood that attainment will occur. Thus, if a nonattainment area were given an attainment date of 2012, the projections in the modeled attainment test should be to 2010.

limited sample) should ordinarily be close to that which would be observed if speciated data were available for every day with a $PM_{2.5}$ measurement, provided the more limited sample is large enough. Our recommendation is made to increase the sample size of speciated measurements used to compute the fraction of observed $PM_{2.5}$ which is comprised of each of the 6 components in each quarter. Quarters should correspond to calendar quarters for two reasons. First, the NAAQS is calculated for a calendar year, so the quarters need to fit evenly within a year. Second, the monitored comparisons with the NAAQS use calendar quarters, so the modeling exercise should do so as well.

Step 2. Derive relevant quarterly *observed* spatially averaged air quality values, where appropriate.

For monitoring sites subject to spatial averaging, compute spatially averaged quarterly mean $PM_{2.5}$ values and spatially averaged component fractions for all sites which are associated with one another.

Step 3. Using model results, derive component-specific relative reduction factors (RRF) at each monitor and apply these to observed air quality to obtain a projected quarterly $PM_{2.5}$ estimate.

For each monitoring site, obtain modeled estimates for 5 of the 6 major components of observed $PM_{2.5}$ (i.e., SO_4, NO_3, OC, EC and IP). For each 3-month quarter, calculate site (i) and component (j) specific relative reduction factors $(RRF)_{ij}$. The relative reduction factor for component j at a site i is given by the following expression:

$$(RRF)_{ij} = ([C_{j,\ projected}]/[C_{j,\ current}])_i$$

where $C_{j,\ current}$ is the quarterly mean concentration predicted at or near the monitoring site with emissions characteristic of the period used to calculate the current design value for annual $PM_{2.5}$ (e.g., 2000-2002);

$C_{j,\ projected}$ is the quarterly mean concentration predicted at or near the monitoring site two years prior to the required attainment date (e.g., 2010 in an area having a 2012 attainment date).

Assume that the relative reduction factor for the sixth previously identified component of observed $PM_{2.5}$ (i.e., "U") is "1.00". Apply each component-specific relative reduction factor to the observed average quarterly mean concentration of the corresponding component, derived in step 1. Add the 6 components to obtain a projected average quarterly mean $PM_{2.5}$ concentration.

Step 4. Calculate site-specific projected average annual mean $PM_{2.5}$ for all sites not subject to spatial averaging.

For each monitoring site calculate the predicted projected average annual arithmetic mean $PM_{2.5}$ concentration by averaging the 4 projected site-specific average quarterly mean concentrations obtained in step 3. Note those sites which are not subject to spatial averaging.

Step 5. Calculate spatially averaged projected average annual mean $PM_{2.5}$ for sites subject to spatial averaging.

For sites subject to spatial averaging, calculate the arithmetic mean of the projected average annual mean $PM_{2.5}$ concentrations for each site in a cluster of sites which have been previously identified (by others) as subject to spatial averaging.

Step 6. Compare all projected average annual mean $PM_{2.5}$ concentrations obtained in steps 4 and 5 with 15.0 $\mu g/m^3$.

For each site not subject to spatial averaging, compare the projected average annual arithmetic mean $PM_{2.5}$ concentration obtained in Step 4 with 15.0 $\mu g/m^3$. Also compare each spatially averaged projected average annual arithmetic mean $PM_{2.5}$ concentration obtained in Step 5 with 15.0 $\mu g/m^3$. If all values are $\leq$ 15.0 $\mu g/m^3$, the test is passed.

We illustrate application of the recommended test in example 3.1.

Example 3.1

Given: (1) Area "C" has 3 monitoring sites. Sites 2 and 3 meet the criteria for spatial averaging.

(2) Monitored air quality data show the following average quarterly mean $PM_{2.5}$ concentrations based on 3 years' observations at each site (values are in $\mu g/m^3$).

Site	Quarter 1	Quarter 2	Quarter 3	Quarter 4
1	17	16	16	14
2	15	17	18	13
3	17	19	15	15

(3) Monitored data also show the following regarding composition of each observed quarterly mean $PM_{2.5}$ concentration. For clarity of presentation, we assume there is no quarter to quarter difference in measured composition at sites 2 and 3.

Site	Quarter	EC	OC	SO_4	NO_3	IP	U
1	1	10%	40%	30%	5%	5%	10%
1	2	10%	20%	50%	5%	5%	10%
1	3	10%	20%	50%	10%	5%	5%
1	4	5%	30%	40%	10%	10%	5%
2	All	10%	40%	40%	0%	5%	5%
3	All	5%	20%	45%	5%	5%	20%

(4) Modeled results show the following relative reduction factors $(RRF)_{ij}$ in predicted mass of the 6 components of $PM_{2.5}$:

Site	Quarter	RRF_{EC}	RRF_{OC}	RRF_{SO4}	RRF_{NO3}	RRF_{IP}	RRF_{U}[7]
1	1	.90	.80	.60	.80	1.00	1.00
1	2	1.00	1.00	.50	.60	.90	1.00
1	3	1.00	1.00	.50	.60	.90	1.00
1	4	.90	.80	.60	.80	.90	1.00
2	1	.90	.80	.60	.80	1.00	1.00
2	2	.90	.80	.60	.80	1.00	1.00
2	3	1.00	.90	.60	.60	.90	1.00
2	4	1.00	.90	.60	.60	.90	1.00
3	1	.90	.80	.70	.80	.90	1.00
3	2	.90	.80	.70	.80	.90	1.00
3	3	.90	.80	.70	.80	.90	1.00
3	4	.90	.80	.70	.80	.90	1.00

Find: Will Area "C" will meet the annual NAAQS for $PM_{2.5}$ by the required attainment date?

Solution:

Step 1. Derive relevant quarterly observed site specific quarterly air quality values

[7]Note: the RRF for the unattributed component (U) is always "1.00".

This information is provided in "given" items (2) and (3).

Step 2. Derive relevant quarterly observed spatially averaged air quality values for sites 2 and 3.

Using the information in "given" item (2), the spatially averaged average quarterly mean $PM_{2.5}$ concentrations for sites 2 and 3 are:

Quarter 1: 16.0 $\mu g/m^3$ (i.e., (15 + 17)/2)
Quarter 2: 18.0 $\mu g/m^3$
Quarter 3: 16.5 $\mu g/m^3$
Quarter 4: 14.0 $\mu g/m^3$

Using the information in "given" item (3), the spatially averaged composition data for sites 2 and 3 by quarter are:

Quarters 1: EC(7.5%), OC(30%), SO_4 (42.5%), NO_3 (2.5%), IP (5%), U (12.5%)
Quarters 2-4 are the same as Quarter 1 in this example.

Step 3. Apply quarterly component-specific relative reduction factors to quarterly observations to obtain quarterly projected air quality estimates for each monitoring site.

Use information in "given" items (2), (3) and (4) for sites 1, 2 and 3.

Site 1:

Quarter 1:

$$[PM_{2.5}]_{1\ projected} = ([PM_{2.5}]_{1\ current})\ 3\ [(RRF)_j\ (\text{Fraction of observed } PM_{2.5} \text{ comprised of component j})]_1$$

$$= [17.0]\ [(.10)(.90) + (.40)(.80) + (.30)(.60) + (.05)(.80) + (.05)(1.00) + (.10)(1.00)]$$

$$= 13.3\ \mu g/m^3$$

Quarter 2:

$$[PM_{2.5}]_{1\ projected} = [16.0]\ [(.10)(1.00) + (.20)(1.00) + (.50)(.50) + (.05)(.60) + (.05)(.90) + (.10)(1.00)]$$

$$= 11.6\ \mu g/m^3$$

Quarter 3:

$$[PM_{2.5}]_{1\ projected} = [16.0]\ [(.10)(1.00) + (.20)(1.00) + (.50)(.50) + (.10)(.60) + (.05)(.90) + (.05)(1.00)]$$

$$= 11.3\ \mu g/m^3$$

Quarter 4:

$[PM_{2.5}]_{1\ \text{projected}} = [14.0]\ [(.05)(.90) + (.30)(.80) + (.40)(.60) + (.10)(.80) + (.10)(1.00) + (.05)(1.00)]$

$= 10.6\ \mu g/m^3$

Using the relevant information in "given" items (2), (3) and (4) and identical procedures followed for site 1, average quarterly projected mean $PM_{2.5}$ concentrations are computed for sites 2 and 3. The following estimates result.

Site 2:

Quarter 1: $[PM_{2.5}]_{2\ \text{projected}} = 11.3\ \mu g/m^3$
Quarter 2: $[PM_{2.5}]_{2\ \text{projected}} = 12.8\ \mu g/m^3$
Quarter 3: $[PM_{2.5}]_{2\ \text{projected}} = 14.3\ \mu g/m^3$
Quarter 4: $[PM_{2.5}]_{2\ \text{projected}} = 10.3\ \mu g/m^3$

Site 3:

Quarter 1: $[PM_{2.5}]_{3\ \text{projected}} = 12.8\ \mu g/m^3$
Quarter 2: $[PM_{2.5}]_{3\ \text{projceted}} = 14.3\ \mu g/m^3$
Quarter 3: $[PM_{2.5}]_{3\ \text{projected}} = 11.3\ \mu g/m^3$
Quarter 4: $[PM_{2.5}]_{3\ \text{projected}} = 11.3\ \mu g/m^3$

Step 4. Calculate site-specific projected average annual mean $PM_{2.5}$ for site 1.

Using the quarterly mean projections for site 1 obtained in step 3, we get

Annual $[PM_{2.5}]_{1\ \text{projection}} = (13.3 + 11.6 + 11.3 + 10.6)/4 = 11.7\ \mu g/m^3$

Step 5. Calculate spatially averaged projected average annual mean $PM_{2.5}$ for sites 2 and 3.

First, use the information generated in step 3, to compute spatially averaged quarterly means for the two sites

Quarter 1 $[PM_{2.5}]_{\text{projection}} = (11.3 + 12.8)/2 = 12.1\ \mu g/m^3$
Quarter 2 $[PM_{2.5}]_{\text{projection}} = (12.8 + 14.3)/2 = 13.6\ \mu g/m^3$
Quarter 3 $[PM_{2.5}]_{\text{projection}} = (14.3 + 11.3)/2 = 12.8\ \mu g/m^3$
Quarter 4 $[PM_{2.5}]_{\text{projection}} = (10.3 + 11.3)/2 = 10.8\ \mu g/m^3$

Then, we average these 4 quarterly values to obtain the projected average annual mean $PM_{2.5}$ for the two sites.

Annual Spatially Averaged $[PM_{2.5}]_{\text{projection}} = (12.1 + 13.6 + 12.8 + 10.8)/4 = 12.3\ \mu g/m^3$

Step 6. Compare the site specific annual projection at site 1 and the spatially averaged annual projection for sites 2 and 3 to 15.0 μg/m^3.

We note from Step 4, that the site-specific annual projection for site 1 is 11.7 μg/m^3. From Step 5, we see that the spatially averaged annual projection for sites 2 and 3 is 12.3 μg/m^3. Neither of these projections is greater than 15.0 μg/m^3. Therefore, the modeled attainment test is passed in Area "C".

<u>Recommendations</u>. Calculations underlying the modeled attainment test for the annual NAAQS should be performed on a quarter by quarter basis. We recommend measured $PM_{2.5}$ be split into 6 components: mass associated with sulfates, nitrates, organic carbon, elemental carbon, inorganic particulate matter and unattributed mass. The calculations should be performed for each of 6 components of $PM_{2.5}$ rather than on $PM_{2.5}$ as a single entity. The test multiplies quarterly component-specific relative reduction factors times concentrations of each component, derived from observed data. Relative reduction factors are estimated by noting the modeled response of each component to a proposed strategy. Resulting projected quarterly component concentrations are then aggregated to estimate projected average annual mean $PM_{2.5}$ concentrations. The test we recommend has 6 steps:

1. Derive quarterly mean estimates for $PM_{2.5}$ and for elemental carbon, organic carbon, sulfates, nitrates, inorganic particulate matter and unattributed/unmodeled particulate matter from current *observed* data at each $PM_{2.5}$ monitoring site.

2. Using results of step 1, calculate observed spatially averaged air quality data for sites which qualify for spatial averaging.

3. Using model results, derive component-specific quarterly relative reduction factors for each monitoring site. Apply the relative reduction factors and results of step 1 to estimate projected quarterly mean $PM_{2.5}$ concentrations at each site.

4. Calculate site-specific average annual mean $PM_{2.5}$ for all sites not subject to spatial averaging using the information generated in step 3.

5. Calculate spatially averaged projected average annual mean $PM_{2.5}$ for sites subject to spatial averaging.

6. Compare all projected average annual mean $PM_{2.5}$ values obtained in steps 4 and 5 with the level specified in the annual NAAQS (15.0 μg/m^3).

If all values compared in step 6 are ≤ 15.0 μg/m^3, the test is passed.

3.2 What Is The Recommended Modeled Attainment Test For The 24-Hour NAAQS?

Our recommended modeled attainment test for the 24-hour NAAQS for $PM_{2.5}$ is similar to the previously described test for the annual NAAQS in that it uses model predictions in a relative sense to reduce site-specific *observations* (averaged over 3 years). In the test, we are interested in seeing what it takes to reduce current design values (i.e., 98th percentile concentrations averaged over three consecutive years) at each site to $\leq$ 65 $\mu g/m^3$.

Ideally, the modeled attainment test should reflect model results obtained for all days in each season having observed $PM_{2.5}$ concentrations below (but > 65 $\mu g/m^3$) as well as above the design value. This may seem strange at first. The underlying reasons are that $PM_{2.5}$ consists of a mixture of pollutants. Composition of the mixture could vary substantially from season to season. Second, there could be a substantial amount of uncertainty associated with predictions on any single day. Thus, our test is most likely to be reliable when relative reduction factors reflect composite responses from many days. Therefore, we recommend modeling as many days as feasible where observed $PM_{2.5}$ is greater than 65 $\mu g/m^3$. As with the annual NAAQS (and for the same reasons), the preferred approach is also to develop relative reduction factors which are season (i.e., quarter) specific.

We have noted that it is desirable to base our conclusions on a composite response of the model(s) over many days. However, from limited monitoring data available as of 2000, there is not likely to be many days with observed concentrations greater than 65 $\mu g/m^3$. If this results in a sample size of smaller than about 10 days per quarter, days with $PM_{2.5}$ observations > ~55 $\mu g/m^3$ and speciated measurements should also be considered. If, after doing this, there is still a small number of days in a single quarter, it may be combined with the most appropriate other quarter. If the sample of high days is limited in 2 or more quarters, a State may develop relative reduction factors which reflect a combined sample of days from the entire year.

We recommend a modeled attainment test for the 24-hour $PM_{2.5}$ NAAQS with 5 steps.

Step 1. For each $PM_{2.5}$ monitoring site, determine the 98th percentile concentration for each of three years. Examine all days having observations > 55 $\mu g/m^3$ to determine which have speciated $PM_{2.5}$ data.

Use the information in Table 3.3 plus knowledge of the number of $PM_{2.5}$ mass measurements to determine concentrations corresponding to the 98th percentile observation for each year at each site. Determine all occasions having observations greater than 65 $\mu g/m^3$ with available speciated measurements. The goal of the modeled attainment test is to provide assurance that measured $PM_{2.5}$ with representative species profiles is reduced to $\leq$ 65 $\mu g/m^3$. If there is a limited number of days with observed concentrations $\geq$ 65 $\mu g/m^3$, we suggest States consider days with observed $PM_{2.5} \geq$ 55 $\mu g/m^3$ to increase the available sample size for modeling.

Identify all such days which also have speciated data. These are the candidate days for modeling.

Step 2. Compute the observed 3-year average design value at each monitoring site.

This step is performed to establish a site specific concentration to be reduced to 65 $\mu g/m^3$. It is computed by taking the arithmetic mean of the 98th percentile concentration at a site for each of 3 consecutive years. Corresponding "observed" values for each of the 6 components of $PM_{2.5}$ are derived by computing an arithmetic mean composition profile[8] from speciated measurements on the modeled days and multiplying this mean site-specific composition profile times the observed site-specific $PM_{2.5}$ design value.

Step 3. For each site, develop component-specific relative reduction factors to be applied to the current site-specific "observed design values" derived for each component in step 2.

At each monitoring site, calculate the arithmetic mean concentration (using all modeled days in a quarter[9]) of each component corresponding with current emissions. Then calculate the arithmetic mean concentration of each component corresponding to the projected future emissions. For component j at site i:

$$(RRF)_{ij} = ([C_{j\ projected}]/ [C_{j\ Current}])_i$$

where $(RRF)_{ij}$ is the relative reduction factor developed at site i for component j;

$C_{j\ current}$ is the predicted arithmetic mean 24-hr concentration of component j, computed from modeled days, using emissions which correspond to the period in which the design value is measured (e.g., 2000-2002);

$C_{j\ projected}$ is the predicted arithmetic mean 24-hr concentration of component j with projected emissions corresponding to a time two years prior to the required attainment date (e.g., 2010 for an area with a required attainment date of 2012).

---

[8]Ideally, quarter-specific composition profiles should be derived. If there is an insufficient number of speciated observations on occasions with high observed $PM_{2.5}$, it will be necessary to use average annual profiles for days with high observed $PM_{2.5}$. See Section 12.3 for further discusion.

[9]Excluding "ramp-up" days

Step 4. At each monitoring site, project future $PM_{2.5}$ design values by multiplying each component-specific relative reduction factor obtained in step 3 times the corresponding component-specific "observed design value" derived in step 2. Add the results to obtain the estimated future site-specific design value for $PM_{2.5}$.[10]

This is done using the following expression

$$[PM_{2.5}]_{i\ projected} = [[PM_{2.5}]_{i\ current}]\ 3[(RRF)_j\ (\text{Fraction of observed } PM_{2.5} \text{ comprised of component j})]_i$$

where $[PM_{2.5}]_{i\ projected}$ is the projected $PM_{2.5}$ design value at monitor site i,

$[PM_{2.5}]_{i\ current}$ is the currently observed design value at site i, and

$(RRF)_j$ is the relative reduction factor calculated for component j at site i.

Step 5. Compare each projected $PM_{2.5}$ design value obtained in step 4 with 65 $\mu g/m^3$.

If all of the projected $PM_{2.5}$ design values are $\leq$ 65 $\mu g/m^3$, the test is passed.

The recommended test is illustrated in example 3.2.

Example 3.2

Given: (1) Area "D" has 2 monitoring sites. The average 98th percentile $PM_{2.5}$ concentration (i.e., the site specific design value) at site 1 is 80 $\mu g/m^3$. The site specific design value at site 2 is 70 $\mu g/m^3$.

(2) During the 3-year period used to compute the observed design values, there are 2 days with observations $\geq$ 55 $\mu g/m^3$ at site 1 having speciated data. $PM_{2.5}$ observations at site 1 on these days are 80 $\mu g/m^3$ and 70 $\mu g/m^3$. Site 2 has one candidate day with an observed value $\geq$ 55 $\mu g/m^3$. $PM_{2.5}$ concentration on this day is 66 $\mu g/m^3$.[11]

---

[10]Note that if the sample of days with observed concentrations > 55 $\mu g/m^3$ is sufficiently large to use quarter-specific monitored data, this step is altered somewhat. In this case, one should multiply the quarter-specific RRF values times an appropriate observed value. If the design value occurs in a particular quarter, use the observed design value in that quarter. For other quarters, use the highest observed concentration which is less than or equal to the design value.

[11]We anticipate that there would be more candidate days for modeling in actual applications. Indeed, it is likely that results obtained for a sample size as small as the one in this example would be highly uncertain. We are using an example with few candidate days and few monitoring sites for ease of presentation.

(3) The following day specific species category fractions are observed:

Site	Day	EC	OC	SO_4	NO_3	IP	U
1	1	5%	30%	50%	5%	5%	5%
1	2	5%	35%	45%	0%	5%	10%
2	3	10%	50%	20%	10%	10%	0%

(4) The following species-specific relative reduction factors have been calculated for the 2 sites shown in (3):

Site	RRF_{EC}	RRF_{OC}	RRF_{SO4}	RRF_{NO3}	RRF_{IP}	RRF_{U}
1	.80	1.00	.30	.50	1.00	1.00
2	.70	.70	.90	.90	1.00	1.00

Find: Is the attainment test for the 24-hour NAAQS passed in Area "D"?

Solution:

Step 1. Determine currently observed site-specific design values and candidate days for modeling.

Candidate days for modeling (to compute the RRF values) are identified in "given" item (2). Since there are so few of these days, we need to consider the year as a whole rather than performing quarter-specific analyses.

Step 2. Compute the observed 3-year average design value at each monitoring site.

Information provided in "given" item (1) reflects application of this step. The design values for sites 1 and 2 are 80 $\mu g/m^3$ and 70 $\mu g/m^3$, respectively. Note that if one of the site specific design values had been much less than 65 $\mu g/m^3$ (say, <55 $\mu g/m^3$), it would not have been necessary to apply the test at this site. It is generally a good idea to consider sites having design values somewhat below 65 $\mu g/m^3$ in the test, since relative "reduction" factors could be greater than "1.00" as a result of growth with few control measures being applied.

Step 3. Choose days for modeling and develop site-specific relative reduction factors to be applied to the design value obtained in step 2.

Note that we have chosen to include day 2 as well as day 1 in applying the test for site 1,

despite measured $PM_{2.5}$ having similar composition on the two days, and the measured $PM_{2.5}$ on day 2 being substantially less than on day 1. It is generally good practice to include all candidate days. The resulting site- and day-specific relative reduction factors are shown in "given" item (4).

Step 4. Project future $PM_{2.5}$ concentrations applying relative reduction factors developed in step 3 to design values developed in step 2.

From "given" item (4), we see that the relative reduction factors at site 1 for EC, OC, SO_4, NO_3, IP and U are, respectively, 0.80, 1.00, 0.30, 0.50, 1.00 and 1.00.

Relative reduction factors at site 2 are EC (0.70), OC (0.70), SO_4 (0.90), NO_3 (0.90), IP (1.00) and U (1.00).

Therefore, for site 1,

$$[PM_{2.5}]_{1\ projected} = [80]\ [(.80)(.05) + (1.00)(.325) + (.30)(.475) + (.50)(.025) + (1.00)(.05) + (1.00)(.075)]$$

$$= 52\ \mu g/m^3$$

and for site 2,

$$[PM_{2.5}]_{2\ projected} = [70]\ [(.70)(.10) + (.70)(.50) + (.90)(.20) + (.90)(.10) + (1.00)(.10) + (1.00)(0)]$$

$$= 55\ \mu g/m^3$$

Step 5. Compare the results in step 4 with 65 $\mu g/m^3$.

Note that all projected site-specific $PM_{2.5}$ design values are $\leq$ 65 $\mu g/m^3$. Therefore, the modeled attainment test is passed in Area "D".

<u>Recommendations</u>. The modeled attainment test for the 24-hour NAAQS for $PM_{2.5}$ should consider days with observations $\geq$ 55 $\mu g/m^3$ and with available speciated data. Observed $PM_{2.5}$ should be divided into 6 components on each day at each site. The six components are: mass associated with elemental carbon, organic carbon, sulfates, nitrates, inorganic particulate matter and unattributed mass. Modeling is used to develop site-specific relative reduction factors for five of the six components. The relative reduction factor for unattributed mass is always "1.00". Relative reduction factors, along with compositional information, should be multiplied by site specific observed design values to project a future $PM_{2.5}$ design value at each site. If the projected design value is $\leq$ 65 $\mu g/m^3$ for all sites, the modeled attainment test is passed. The recommended test has 5 steps.

1. For each monitoring site identify the 98th percentile $PM_{2.5}$ concentration for each

year. Identify individual days having observed $PM_{2.5}$ concentrations $\geq$ 55 µg/m^3 which also have speciated observations.

2. Compute the observed 3-year average $PM_{2.5}$ design value and use speciated observations to derive "observed" component-specific design values for each monitoring site.

3. For each site, use model results to develop component-specific relative reduction factors (RRF).

4. At each site, project future $PM_{2.5}$ design values by multiplying each component-specific RRF times the corresponding component-specific "observed" design value derived in step 2. Add the results to obtain the estimated future site-specific design value for $PM_{2.5}$. See the previously identified footnote for this step if there is a large number of days with speciated measurements which also have observations exceeding 55 µg/m^3.

5. Compare each estimated future design value to 65 µg/m^3. If all projected concentrations are $\leq$65 µg/m^3, the test is passed.

3.3 How Do I Select Appropriate Inputs For The Recommended Tests?

The recommended attainment tests are very similar for the annual and 24-hour NAAQS for $PM_{2.5}$. Basically, a State should (a) estimate "current" measured site specific design values for $PM_{2.5}$, (b) derive corresponding "observed" site specific design values for each of 6 major components of $PM_{2.5}$, (c) take the ratio of modeled future to current conditions to derive site-specific relative reduction factors for each of the 6 components, (d) for each site, multiply each component-specific "observed" design value times the component's relative reduction factor to estimate a future concentration, (e) add the estimated future component concentrations to estimate a corresponding future design value for $PM_{2.5}$. All this sounds relatively straightforward, but is it? It turns out there are several ways of deriving the information needed to execute the preceding steps. In this section, we identify several preferences to promote consistency among different applications. We do this by addressing several questions:

- how do I decide what "current period" and corresponding "current design value" to use in the test?

- in the test for the 24-hour NAAQS, what do I do if each of the three years used to estimate the 98th percentile concentration has a different sampling frequency?

- how do I estimate *mass* associated with each of the six components of $PM_{2.5}$?

- how do I decide what portion of the measured mass of $PM_{2.5}$ should be categorized as

"unattributed" (U)?

- how do I derive "observed" component specific design values?

- in computing a site specific RRF value, how do I decide which predictions occur "near" a monitor?

- which of several predictions "near" a monitor do I use to compute the RRF?

Choosing the current period and corresponding current design value. For the recommended tests to be most sound, "current emissions" (used to calculate the denominator in the relative reduction factor) should be consistent with emissions likely during the period reflected by the "current (observed) design value". Failure to meet this need could result in projected future design values which are inappropriate. For example, if the "current inventory" reflects an earlier period where emissions are substantially higher than those during the period corresponding to the design value, a State could underestimate the future design value. This follows, since the contrast between "current" and future emissions (and therefore predicted air quality differences) could be greater than warranted. As a result, the RRF values would be too low.

Ordinarily, if the year reflected by the most recent available inventory is included within the 3-year period used to compute a design value, the need for the two to coincide is met. For example, if the most recent available inventory is for 2002, design values reflecting 2000-2002, 2001-2003 and 2002-2004 are among the candidates for "current (observed) design value". Although this should be confirmed for each attainment demonstration, major differences in emissions are unlikely from year to year. Thus, differences in design values for three consecutive overlapping 3-year periods most likely reflect meteorological differences. Calculating a design value by averaging observations over three years will dampen fluctuations in the design values attributable to meteorological differences. Nevertheless, looking only at three 3-year periods is unlikely to adequately reflect potential variability in the design value due to meteorology. States should ordinarily choose the highest site-specific design value of the three candidate periods for use in the tests. This provides a safeguard against basing the test on design values observed during an unusually mild 3-year period. Using this approach, it is conceivable that different 3-year periods (all close to the period represented by the current inventory) could be used as the basis for the current observed design value at different monitoring sites in a nonattainment area. This is acceptable. States may demonstrate that choice of a set of the most severe of three candidate design values leads to one or more current design values which is inappropriately high for use in the test. This latter argument could be considered as part of a weight of evidence determination (see Section 5.0).

We need to mention one final issue before moving on to the next topic. What if the period used to designate an area as "nonattainment" and the period reflected by the most recent inventory do not coincide? For example, what if data from 2000-2002 were used as the basis for

a designation, but the most recent inventory reflected 1999? If a site-specific design value for the period used for designation is higher than any of the three periods "straddling" the period reflected by the current inventory, it should be used. Otherwise, States should use the highest of the three site-specific design values straddling the period reflected by the inventory.

Estimating 98th percentile concentrations and observed design values when sample sizes differ from year to year. States should use the guidance in Table 3.3 to compute the 98th percentile $PM_{2.5}$ concentration at each monitoring site for each of three consecutive years. The observed site-specific design value is the arithmetic mean of these three 98th percentile values.

Estimating measured mass associated with components of $PM_{2.5}$ and dealing with differences between measured $PM_{2.5}$ and the sum of identified components. We have been very careful to note that it is important to estimate mass *associated* with each of the major components of $PM_{2.5}$, not just presence of the component itself. Estimating associated mass requires us to make assumptions about which of several measured (or unmeasured) chemical species are bound together. For example, SO_4 aerosols could, in theory, be present in any one of four forms: ammonium sulfate [$(NH_4)_2SO_4$], ammonium bisulfate [$(NH_4)HSO_4$], sulfuric acid [H_2SO_4] or letovicite [$(NH_4)_3(HSO_4)_2$]. Mass associated with a measured µg of sulfur depends on the form of sulfate collected on a filter. That is, stoichiometry of the preceding four compounds implies that the mass associated with 1 µg of measured sulfur could range from about 4.1 µg (pure ammonium sulfate) to about 3.1 µg (pure sulfuric acid).

For many locations in the western U.S., using a default assumption that measured sulfate is in the form of [$(NH_4)_2(SO_4)$] is probably valid (i.e., muliplying measured sulfur by "4.125" or measured sulfate by "1.375" provides an appropriate estimate of the mass associated with sulfate). However, in general, we recommend that States estimate the appropriate conversion from measured SO_4 (or sulfur) to mass *associated* with SO_4 by noting the amount of measured ammonium ion [NH_4][12] and nitrate [NO_3] in the collected sample. The first step in estimating mass associated with sulfate is to subtract the amount of measured ammonium associated with nitrate. For the purpose of estimating mass associated with sulfates, a State may assume that all measured nitrate is attributable to ammonium nitrate [NH_4NO_3]. Ammonium ion associated with nitrate is subtracted from the measured concentration of ammonium ion. The remaining portion of ammonium ion is considered in Equation 3.1a or 3.1b to estimate mass associated with sulfates (Malm, et al., 2000).

If sulfate ion is measured,

---

[12]For ease of presentation in this document, we consistently omit valence notation for ions. The valence associated with ammonium ion is +1, that for sulfate is -2, nitrate -1, etc. Occasionally, as with the radical nitrogen trioxide, this could lead to some ambiguity. Unless otherwise noted, the reader should assume that our notation refers to sulfate, bisulfate, nitrate and ammonium ions along with their appropriate valences.

Mass associated with sulfate = 0.932 $[NH_4]$ + 1.02 $[SO_4]$ (3.1a)

If sulfur is measured,

Mass associated with sulfate = 0.932 $[NH_4]$ + 3.06 [S] (3.1b)

Example 3.3 illustrates how to estimate mass associated with measured sulfate ion.

Example 3.3

Given: (1) measured sulfate ion = 10 $\mu g/m^3$

(2) measured nitrate ion = 2 $\mu g/m^3$

(3) measured ammonium ion = 4 $\mu g/m^3$

Find: measured mass associated with SO_4

Solution:

(**1) Subtract mass of ammonium ion associated with nitrate**.

Assuming all nitrate is in the form of ammonium nitrate $[NH_4NO_3]$, we see that 1 mole of NH_4 is equivalent to 1 mole of NO_3.

We are given that 2 $\mu g/m^3$ of nitrate has been measured. Dividing this by the molecular weight of nitrate (62 g/mole), we note that this measurement is equivalent to 2/62 = .032 µmoles of nitrate. Thus, there are .032 µmoles of ammonium associated with the measured nitrate.

We determine the concentration of ammonium ion associated with nitrates by multiplying the molecular weight of NH_4 ion times .032 µmoles.

$[NH_4]_{\text{associated with NO3}}$ = (.032 µmoles) (18 µg/µmole) = 0.58 $\mu g/m^3$

We obtain the mass of ammonium associated with sulfate by subtracting 0.58 $\mu g/m^3$ from the measured amount of ammonium (given as 4 $\mu g/m^3$). Thus,

$[NH_4]_{\text{associated with SO4}}$ = 4 - 0.58 = 3.42 $\mu g/m^3$

(2) Estimate mass associated with sulfate considering the amount of ammonium ion associated with sulfate.

Since sulfate ion, rather than sulfur, has been measured, use Equation 3.1a to estimate mass associated with sulfate. We are given that the measured concentration of sulfate ion is 10 $\mu g/m^3$.

Mass associated with sulfate = 0.932 [3.42] + 1.02 [10] = 13.4 $\mu g/m^3$. (3.1a)

As previously noted, we believe States may generally assume that measured particulate nitrate is largely in the form of ammonium nitrate. Thus, we recommend that mass associated with nitrate be estimated by multiplying the measured nitrate ion concentration times "1.29". Other default recommendations for translating measurements into mass associated with each component of particulate matter appear in Table 3.4.

Default recommendations for translating EC, OC and the soil portion of IP are consistent with recommendations found in Sisler (1996). There is uncertainty associated with the "1.4" factor used to translate measured OC into mass associated with organic carbon. One recent study suggests higher factors may be appropriate (Turpin and Lim, 2000). However, the value of "1.4" is the one which has been most widely used as of late 2000. States may use an alternate factor in a weight of evidence assessment, if justified. Composition of soil in a local area may also differ from the recommended default assumptions in some parts of the country. Chow, et al., (1996) and Solomon, et al., (1989) describe procedures for characterizing composition of soil dust used for studies in central and southern California. States may use alternative procedures for estimating measured mass associated with crustal material in a weight of evidence analysis if they can present sufficient justification for doing so. We should note however, that the portion of a nonattainment area's measured $PM_{2.5}$ attributable to soil dust may often be low. In such cases, a major effort to refine these calculations may result in little change in the results of a modeled attainment test.

Unattributed mass (U) is the difference between mass of $PM_{2.5}$ measured with the Federal Reference or equivalent method and the sum of the mass associated with the other five components.[13] If a certain portion of inert $PM_{2.5}$ is unmodeled because it originates outside of the United States, its mass may be included within component "U". If this practice is followed, a corresponding portion of the mass allocated to one or more other components should be reduced. For example, if monitored data are used to attribute a certain portion of measured $PM_{2.5}$ to Saharan dust, this may be subtracted from component "IP", included as part of component "U" and need not be modeled, either explicitly or as assumed background. On the other hand, if Saharan dust were included as background in a model application, it should continue to be included as part of component "IP".

---

[13]In the unlikely event that the sum of the other 5 components exceeds the mass of $PM_{2.5}$ measured with the FRM or equivalent method, assume U is zero, and adjust the component of measured particulate matter with the greatest uncertainty (generally OC) downward so that the sum of the 5 components and the measured mass of $PM_{2.5}$ agree.

Table 3.4. Recommended Default Assumptions To Derive Mass Associated With Major Components of $PM_{2.5}$

(1) Component	(2) Formula To Derive Mass Associated With Component	(3) Assumptions
SO_4	See accompanying text	All elemental sulfur is from sulfate
NO_3	1.29 (measured nitrate)	Denuder efficiency is ~100% & all nitrate is from ammonium nitrate
EC	1.00 (high + low temperature EC)	All high temperature carbon is elemental
OC	1.4 (organic carbon)	Average organic molecule is 70% carbon (other elements include hydrogen, nitrogen and oxygen)
Inorganic Particulate Matter (IP) = mass associated with crustal material (soil) plus combustion	Mass associated with soil = 2.2(Al) + 2.49(Si) + 1.63(Ca) + 2.42(Fe) + 1.94(Ti) Mass associated with combustion = oxides of remaining measured metallic elements	(Soil K)=0.6(Fe), FeO & Fe_2O_3 are equally abundant, a factor of 1.16 is used for MgO, Na_2O, H_2O & CO_3 Multiplication factors for remaining elements (associated with combustion) depends on the stoichiometry of the compound when combined with oxygen
Unattributed mass (U)	$U = PM_{2.5} - (SO_4 + NO_3 + OC + EC + IP)$	Six components account for all measured $PM_{2.5}$

Deriving component-specific "observed" design values. For applications addressing the annual NAAQS, States should compute the mean mass associated with each of the 6 components of $PM_{2.5}$ at each monitoring site for each quarter over a continuous 3-year period. Each of these 6 values should then be divided by the mean $PM_{2.5}$ measured on the corresponding days at that site. The resulting 6 quotients reflect the composition of $PM_{2.5}$ at the monitor. Each of the 6 quotients should then be multiplied by the site-specific quarterly $PM_{2.5}$ mean value, obtained by averaging the observed quarterly mean concentration over 3 continuous years. The product of this multiplication is the "observed" component specific mean value for each quarter.

For applications related to the 24-hour NAAQS, States should compute the mean mass associated with each of the 6 components for all days chosen for modeling which have speciated air quality data. These values should be divided by the corresponding mean $PM_{2.5}$ measurement on these days at the monitoring site in question. The resulting 6 quotients reflect the composition of $PM_{2.5}$ at the monitor. Each of these should be multiplied by the observed site-specific $PM_{2.5}$ design value (i.e., the 98^{th} percentile concentration averaged over 3 consecutive years) to derive the "observed" component-specific design values for the 24-hour NAAQS.

The preceding approaches are straightforward. However, what should a State do if the speciated and $PM_{2.5}$ mass measurements are not collocated? To minimize this problem, we suggest States review preliminary FRM data and give priority to making speciated measurements at sites showing a potential for violating one or both NAAQS. In the event an FRM monitoring site is unaccompanied by collocated species measurements, we recommend that those responsible for implementing a State's modeling protocol review available speciated data, monitor placement, meteorological conditions and local geographic features to choose among available composition observations. For example, if available speciated data suggest little gradient in components dominated by secondary particulate matter (e.g., SO_4, NO_3 and, perhaps, OC), concentrations of these components could be assumed as constant from site to site and further interpolation need only address components consisting largely of primary particulate matter. Any remaining problem may be addressed by interpolating from existing speciated data using objective analysis (e.g., such as weighing available observations according to the inverse square of a site's distance from a measurement).

One additional question arises. What should a State do if speciated and FRM measurements are collected on different schedules (e.g. once every 3 days vs. once every 6 days)? The procedure we recommend in the attainment tests is to use composition estimates which reflect measured speciation profiles that have been averaged over a number of days. Thus, States should follow this procedure even if speciated data are only available for a subset of days used to derive the monitored design values for $PM_{2.5}$ used in the attainment tests.

Deciding what is "near" a monitoring site. As noted in section 1.3, secondary particulate matter is likely to constitute an important fraction of $PM_{2.5}$. Further, a major purpose of urban monitoring performed to determine attainment of the NAAQS for $PM_{2.5}$ is to estimate likely exposure to $PM_{2.5}$ concentrations over 24-hour periods. According to 40CFR Part 58,

Appendix D, sites having "neighborhood" or "urban" scales of representativeness are generally most suitable for estimating exposure or compliance with NAAQS. Because of the rather long sampling (exposure) time (i.e., 24 hours) inherent in the $PM_{2.5}$ NAAQS, locations within the lower range of "urban" scale (4-50 km) would seem consistent with a definition of "near" a site. Therefore, we recommend following the same ground rules used for defining "near" as were used in recent draft guidance for ozone modeling (U.S.EPA, 1999b).

States may assume a distance of $\leq$ 15 km to be "near" a monitoring site. However, for ease in computation, States may assume that a monitor is at the center of the model grid cell in which it is located and that this cell is at the center of an array of "nearby" cells. the number of cells considered "nearby" (i.e., within about 15 km of) a monitor is a function of the size of the grid cells used in the modeling. Table 3.5 provides a set of default recommendations for defining "nearby" cells for grid systems having cells of various sizes. Thus, if one were using a grid with12 km grid cells, "nearby" is defined by a 3 x 3 array of cells, with the monitor located in the center cell. States may consider presence of topographic features, demonstrated mesoscale flow patterns (e.g., land/sea, land/lake interfaces) or other factors to refine our default definitions for the array of "nearby" grid cells, provided the justification for doing so is documented.

Table 3.5. Default Recommendations For Nearby Grid Cells Used To Calculate RRF's

Size of Individual Cell, km	Size of the Array of Nearby Cells, unitless
$\leq$ 5	**7 x 7**
>5 - 8	**5 x 5**
>8 - 15	**3 x 3**
>15	**1 x 1**

Choosing model predictions to calculate a relative reduction factor ($(RRF)_j$) near a monitor. This issue is to decide which one(s) of the 24-hour predicted concentrations of $PM_{2.5}$ components in cells near a monitor should we use to calculate the component-specific RRF's? We recommend taking the spatially averaged value of the nearby predictions. Each component-specific relative reduction factor $(RRF)_j$ used in the modeled attainment test is computed by taking the ratio of the mean of the spatially averaged daily predictions in the future to the mean of the spatially averaged daily predictions with current emissions. The procedure is illustrated in Example 3.4. For ease of presentation, we illustrate the procedure for only one component (SO_4) for a limited number of days.

Example 3.4

Given: (1) Four primary days have been simulated using current and future emissions;

(2) The horizontal dimensions for each surface grid cell are 12 km x 12 km;

(3) Figure 3.1 shows predicted mass associated with SO_4 in each of the 9 cells "near" a monitor with (a) future emissions, and (b) current emissions.

Find: The site-specific relative reduction factor $(RRF)_j$ for mass associated with SO_4.

Solution:

(1) For each day and for both current and future emissions, identify the 24-hour mass associated with SO_4. Since the grid cells are 12 km, a 3 x 3 array of cells is considered "nearby" (see Table 3.5). The numbers appearing beneath each 3 x 3 array in Figure 3.1 are the spatially averaged mass associated with SO_4 for each day.

(2) Compute the mean spatially averaged mass associated with SO_4 for (a) future and (b) current emissions.

Using the information in Figure 3.1,

(a) $(\text{mean mass associated with } SO_4)_{future} = (7.5 + 7.0 + 7.0 + 7.2)/4 = 7.2\ \mu g/m^3$

(b) $(\text{mean mass associated with } SO_4)_{current} = (8.3 + 8.8 + 8.0 + 8.0)/4 = 8.3\ \mu g/m^3$

(3) The relative reduction factor SO_4 near the site is

$$(RRF)_j = (\text{mean mass associated with } SO_4)_{future}/(\text{mean mass associated with } SO_4)_{current}$$

$$= 7.2/8.3 = 0.87$$

Recommendations. States should choose the period with the highest observed $PM_{2.5}$ design value from 4 candidate 3-year periods: the three 3-year periods which include the year represented by the most recent available inventory and the 3-year period upon which the nonattainment designation is based (if this is not one of the first three periods).

States should use stoichiometric relationships together with measured speciated data to estimate mass associated with each of the 6 components of $PM_{2.5}$ considered in the modeled attainment test. Information presented in Table 3.4 may be used as default assumptions.

"Observed" component-specific design values should be estimated by computing the mean mass associated with each component on all days with speciated measurements (annual NAAQS) or on all modeled days with measurements (24-hr NAAQS). These averages should be divided by the mean mass of $PM_{2.5}$ on these same days. This provides a mean estimate of the composition of $PM_{2.5}$. The "observed"component-specific design values are estimated by multiplying the appropriate $PM_{2.5}$ design value times the mean composition of $PM_{2.5}$.

States should use predictions "near" a monitor to compute site-specific RRF values. "Near" is defined as a location within 15 km of a monitoring site, and may be approximated by considering predictions in an array of grid cells centered on the monitor. The number of cells considered is a function of cell size.

States should spatially average current and future predictions in cells near each monitor. The site-specific RRF should be estimated by taking the ratio of spatially averaged future predictions to spatially averaged current predictions.

3.4 Why Do I Need To Consider A Hot Spot Analysis And How Do I Do It?

Primary particulate matter does not undergo physical/chemical transformation between its being emitted and its arrival at a receptor location. Thus, a relatively lengthy travel time from source to receptor (to enable thorough mixing and chemistry to proceed) is not needed for high concentrations of primary particulate matter to occur. Putting this another way, unlike secondary particulate matter, we would often expect concentrations of primary particulate matter to increase the closer one gets to its source(s) of emissions. Therefore, if a designated nonattainment area contains a few (as opposed to many which are spread out) concentrated sources of primary particulate matter, we would expect there to be some substantial spatial gradients in the primary portion of the organic carbon component and in the inorganic particulate matter (IP) and elemental carbon (EC) components of ambient $PM_{2.5}$. Substantial gradients are most likely to be a potential problem in addressing whether a proposed control strategy is sufficient to attain the 24-hour NAAQS for $PM_{2.5}$. This follows, because orientation of a source's plume varies. For many hours, one might expect to find no impact from such a source at a given location. Occasions with no likely impact tend to balance those large impacts occurring over some periods. As a result, we expect the need for a hot spot analysis to be greatest when evaluating the adequacy of a control strategy for meeting the 24-hour NAAQS.

Given the potential for spatial gradients in the primary components of $PM_{2.5}$ and the variety of purposes for $PM_{2.5}$ monitoring, it is not likely that monitors will be located so as to detect potential large impacts from every major source of IP, EC and primary OC. Thus, we recommend that States supplement the recommended modeled attainment test with a hot spot analysis when warranted.

Determining whether a hot spot analysis is warranted. This may be done in one of

two ways. The first approach is to review the spatial distribution of estimated primary $PM_{2.5}$ emissions for grid cells within a designated nonattainment area for $PM_{2.5}$. Note the location of grid cells in which primary emissions (i.e., from point, area and mobile sources) are well above (e.g., > 20% higher) those "near" any monitoring site for $PM_{2.5}$ within the designated nonattainment area. Major sources within such cells are candidates for a hot spot analysis. The second approach for flagging major sources subject to a hot spot analysis makes use of available results from air quality modeling. In this approach, a State should flag major sources in any grid cell for which predicted concentrations of primary particulate matter are substantially (e.g., > 20% greater) and consistently (e.g., $\geq$ 50% of the days modeled) greater than predicted concentrations of primary particulate matter near any site monitoring $PM_{2.5}$.

Size of the area considered in a hot spot analysis. The area over which a hot spot analysis should be performed needs to be determined on a case by case basis. Size of the area considered depends on the nature of a source's plume. For example, analyses for sources whose emissions are released aloft in heated plumes (e.g., combustion sources) will likely need to cover a larger area than is necessary for sources emitting coarse mode particulate matter close to the ground. However, since a hot spot analysis is primarily interested in estimating the maximum impact from a flagged source, the area of coverage is seldom likely to be greater than about 15 - 20 km from the source.

Resolving emissions-related issues. We recommend a focused quality assurance check on emissions from sources flagged for a hot spot analysis. Prior to applying a model, States should review available information to ensure that there are no commonly reported major discrepancies between modeled estimates applying available emission factors and estimated activity levels for similar sources and nearby monitored data for particulate matter. If such discrepancies are commonly reported, those implementing the modeling protocol (i.e., see Section 10.0) should consult with the appropriate U.S. EPA regional office to reach agreement on what, if any, adjustments should be made to the emissions estimates for the source subject to a hot spot analysis.

The next emissions-related issue is, what emissions level should a State assume (i.e., one corresponding to maximum allowable emissions or one corresponding to a typical "actual" level)? For the purpose of attainment demonstrations addressed in this guidance, emissions of greatest interest are future estimates made for the required attainment date. Estimating future emissions for sources subject to hot spot analyses is an issue which needs to be resolved on a case by case basis, as the most reasonable course may depend on the particular source in question. Generally however, for hot spot applications related to the 24-hour NAAQS, we recommend using maximum potential emissions levels. If a source is subject to an operating permit, maximum emissions allowed by that permit should ordinarily be used as its future emissions in a hot spot analysis. For applications related to the annual NAAQS, it is not reasonable to assume that maximum emissions will be sustained throughout the year. Therefore, we recommend using future expected "actual" emissions levels for these applications.

Model application and interpretation. Once outstanding emissions-related issues are resolved, States should apply a point, line or area source model (e.g., a Lagrangian (Gaussian) model) for every day of the 3-year period corresponding to the current observed $PM_{2.5}$ design value used in the modeled attainment test. There is room for some judgment on the part of those implementing the modeling protocol as to whether it is necessary to model every flagged source. For example, if mass and configuration of emissions and background concentrations are similar for two flagged sources, it may only be necessary to model one of them.

If the hot spot analysis is being performed to supplement the recommended test for the 24-hour NAAQS, note the 4th highest concentration in each of the three years at each receptor site. Take the arithmetic mean of these three values at each site. The result is an array of estimated mean 98th percentile impacts of the flagged source on IP, EC and primary portions of the OC component of $PM_{2.5}$. If an analysis is related to the annual NAAQS, compute the arithmetic mean concentration for each quarter of each year at each receptor site. Use this information to calculate the annual arithmetic mean concentration of primary components at each receptor for each of the three years. Take the average of these three arithmetic mean values calculated at each site. The result is an array of estimated average arithmetic mean impacts of the flagged source on the primary components of $PM_{2.5}$.

Superimposing estimated impacts on other ambient estimates. Once a State has estimated the impact of a flagged source, it needs to superimpose this impact on an estimated "ambient" value reflecting secondary components of $PM_{2.5}$ plus other sources of primary particulate matter. If the hot spot analysis is related to the annual NAAQS, this ambient value should be obtained by spatially interpolating the average annual mean $PM_{2.5}$ concentrations projected at monitored locations in the modeled attainment test. Unless there is some geographical feature or other documented reason for not doing so, we recommend using an objective technique for doing the interpolation (e.g., such as weighing the projections according to the inverse square of their distance from the site of the flagged source's maximum projected impact). If the sum of the maximum impact and interpolated ambient value exceeds 15.0 $\mu g/m^3$, some remedial action is needed.

If the hot spot analysis is related to the 24-hour NAAQS, choice of an appropriate ambient value is more difficult. Ideally, States could perform a correlation analysis between the maximum source impact predicted on each day with monitored observations and a background value interpolated from the available monitored data. If there were no statistically significant correlation or if the correlation coefficient were significant but low (e.g., < 0.50), it is appropriate to assume the same background concentration used in the hot spot analysis related to the annual NAAQS. If there were a perfect (positive) correlation between background values interpolated from the monitored data and the estimated maximum impacts, then a background value which is interpolated from the projected 98th percentile $PM_{2.5}$ concentrations at monitoring sites would be appropriate. Given the disparity between sources which may be responsible for the other components of $PM_{2.5}$ and those responsible for primary particulate emissions, we believe a default approach which uses interpolated mean (rather than 98th percentile) projections for

ambient values is appropriate. If there is another source flagged for a hot spot analysis which is within 20 km, it may need to be modeled with a point/line/area source model as well. Resulting estimated impacts from the "nearby"modeled sources should be estimated at each receptor site used for the source which is subject to the hot spot analysis. If the sum of the maximum impact calculated at any receptor site and the interpolated ambient value exceeds 65 $\mu g/m^3$, remedial action is needed.

Projected emissions from nearby flagged sources. This issue is, should I assume projected maximum allowable or projected typical "actual" emissions for nearby sources which are explicitly modeled in a hot spot analysis addressing the 24-hour NAAQS? The question needs to be addressed on a case by case basis. In cases where there is no reason to believe that maximum emission levels among the sources are likely to coincide, we recommend using projected actual emissions for flagged "nearby" sources. If some coincidence of maximum activity is likely (e.g., as might possibly happen for nearby utilities on a hot summer day), States may need to use maximum allowable emissions for projected estimates from nearby sources.

<u>Recommendations</u>. States should apply a supplementary hot spot analysis for major sources of primary particulate matter. Size of the area considered for a hot spot analysis depends on the nature of the modeled source, but is generally less than 20 km from the source.

Prior to modeling, a quality assurance check should be performed on emissions estimated from the source and corrections made as necessary. For hot spot analyses related to the 24-hour NAAQS, States should assume future activity levels which correspond to maximum potential or allowable activity, whichever is less. For applications related to the annual NAAQS expected future activity levels may be assumed.

States should apply a point/line/area source model for every day in the 3-year period corresponding to that of the current $PM_{2.5}$ design value used in the modeled attainment test. For the 24-hour NAAQS, the mean 98th percentile primary particulate matter (IP + EC + primary portions of OC) concentration should be estimated at each receptor site. For the annual NAAQS, States should estimate the average arithmetic mean primary PM concentration at each site. States should note the maximum mean 98th percentile and the maximum average mean impacts.

States should superimpose the maximum impact on the annual mean or 98th percentile daily concentration on ambient concentrations. "Ambient" concentrations should be estimated by interpolating projected annual mean design values obtained in the modeled attainment test at monitoring sites. If there are nearby sources of primary particulate matter which have also been flagged for hotspot analyses, point/line/area modeled concentrations estimated at the receptor sites should also be added to impacts estimated from the source which is the subject

of a particular hot spot analysis. If the sum of a source's maximum impact, the impact of nearby source(s) at that location and the interpolated ambient concentration exceeds either of the two NAAQS for $PM_{2.5}$, remedial action is warranted.

4.0 What Is The Recommended Modeled Test For Reasonable Progress Reducing Regional Haze?

Our recommended modeled test to demonstrate reasonable progress has many similarities to the tests, described in Section 3.0, to demonstrate attainment of the NAAQS for $PM_{2.5}$. They share the following features:

- monitored data should be used to define current air quality;

- monitored concentrations of $PM_{2.5}$ are divided into major components (e.g., SO_4, NO_3, OC, EC, IP);

- it is necessary to estimate monitored and modeled *mass associated with* each major component;

- models are used in a relative sense to develop relative reduction factors (RRF) between future and current predicted concentrations of each component;

- component-specific relative reduction factors are multiplied by current monitored values to estimate future component concentrations;

- estimates for future component concentrations are consolidated to provide an estimate of future air quality which can be related to a goal for regional haze;

- future estimated air quality is compared to the goal to see if a simulated control strategy results in the goal being met.

There are also several differences between the reasonable progress and attainment tests:

- it is not necessary to include a component for unattributed mass (U) of $PM_{2.5}$ in the modeled test for reasonable progress;

- the modeled test for reasonable progress needs to consider coarse particulate matter, not just particles as small or smaller than 2.5 μm;

- for the reasonable progress test, it is acceptable to assume that all measured sulfate is in the form of ammonium sulfate ($(NH_4)_2(SO_4)$).

Regional haze is calculated by estimating light scattering and absorption by components of $PM_{2.5}$. It is not necessary for the sum of these components to equal gravimetric measurements obtained with a Federal Reference or equivalent method for measuring $PM_{2.5}$. Hence, it is not necessary for us to introduce a component "U" to reconcile sum of the components with gravimetric measurements of $PM_{2.5}$. Coarse particles (aerodynamic diameters > 2.5 μm but $\leq$ 10

μm) scatter light. Although light scattering on a per mass basis is less efficient for coarse than for fine particles, in some Class I areas measured coarse mass is comparable or even greater than measured fine mass. Thus, we cannot ignore coarse mass in the modeled test for reasonable progress. In some parts of the U.S. (most notably in the East), reviews of data collected in Class I areas suggest that ammonium bisulfate (NH_4HSO_4) may represent mass associated with sulfate better than does ammonium sulfate. However, in the modeled test for reasonable progress, we are interested in the *product* of mass associated with sulfate and light scattering efficiency. This product is very similar for ammonium sulfate and ammonium bisulfate (i.e., the lower mass of ammonium bisulfate is compensated for by its greater light scattering efficiency) (Tang, 1996). This enables us to simplify the modeled test for reasonable progress by assuming that measured sulfate is in the form of ammonium sulfate throughout the United States.

Goals for reducing regional haze are defined in U.S. EPA , 1999. There are two goals for reducing regional haze.

- Mean regional haze on the 20% of days with worst current visibility should be reduced by an agreed upon relative amount between a base period (2000-2004) and 2018, and

- mean regional haze on the 20% of days with best visibility during the base period should not increase.

The "agreed upon" relative amount of improvement on worst visibility days needs to consider the long term goal stated in the rule, which is to reduce regional haze to "natural background" levels by 2064. Thus, initially, States may be asked to assume that the difference in light extinction on worst visibility days during the base period and natural background is reduced by an amount which is proportional to the number of years between the base period and 2018 vs. the number of years between the base period and 2064. For example, if a Class I area's mean worst visibility during the 2000-2004 base period were 35 *deciviews* and natural background were estimated to be 8 *deciviews*, the goal for 2018 would be to reduce light extinction on the worst days by 6.3 *deciviews* (i.e., [35 -8][(2018-2004)/(2064-2004)]). The rule provides States with some flexibility to establish the goal for improving visibility on worst visbility days between the base period and 2018 in the event this initial goal appears to be infeasible.

In essence, the goals for regional haze between now and 2018 set requirements for *trends* in air quality rather than for meeting some absolute air quality level. The U.S. EPA is developing guidance on how measurements should be used to determine these trends. This is expected to be completed during the first half of 2001. In Section 4.1, we describe how it now appears that ambient trends in regional haze will be determined. The modeled test for reasonable progress needs to provide a good predictor for what the trend in observed air quality will be determined to be in 2018. This is most likely to happen if we reproduce the test which will be applied to the ambient observations as closely as practical. In Section 4.2, we present and recommend a modeled test which closely replicates the anticipated ambient test. Section 4.3 addresses how to develop key inputs for the recommended modeled test for reasonable progress.

4.1 How Is Reasonable Progress Measured?

Regional haze is measured by an *extinction coefficient* (b_{ext}) which represents light attenuation resulting from scattering and absorption of light from ambient particulate matter plus scattering of light due to gas molecules in the air (i.e., Rayleigh scattering). Although b_{ext} can be estimated differently, the regional haze rule requires that it be estimated using measured ambient particulate matter. This follows since, for a given set of meteorological conditions, visibility can be improved by reducing concentrations of ambient particulate matter. Thus, deriving b_{ext} in this manner provides a direct link between regional haze and related pollutant concentrations. Equation (4.1) may be applied in each Class I area to estimate b_{ext} (Sisler, 1996, Malm, et al., 2000)[14]. The equation reflects empirical relationships derived between measured mass of particulate matter components and transmissometer measurements of b_{ext} at monitoring sites in 39 Class I areas within the IMPROVE network.

$$\mathbf{b_{ext} = 3((\mathit{f(rh)})[SO_4] + 3((\mathit{f(rh)})[NO_3] + 4(\mathit{f'(rh)})[OC] + 10[EC] + 1[IP] + 0.6[CM] + b_{rayleigh}} \qquad \mathbf{(4.1)}$$

where

the numerical coefficients on the right hand side of the equation represent the light scattering or absorption efficiency, m^2/gm of the corrsponding component of particulate matter,

f(rh), *f'(rh)* are relative humidity adjustment factors applied to the light scattering efficiency (to be described in greater detail shortly), dimensionless,

SO_4 is the mass associated with sulfates, $\mu g/m^3$,

---

[14]We recommend that States use Equation (4.1) as described. However, States may derive a similar, but different equation which is specific for a given Class I area. We do not recommend this, as it is likely to be expensive and time consuming to develop a sufficient data base to develop such an equation. A State would need to collect non-routine data at the Class I location over a sustained, multi-seasonal period. Data include transmissometer measurements, size distribution measurements for fine as well as coarse particles, composition data, relative humidity measurements and an assessment of the chemical form of sulfate particles at the location in question. This implies need for an ambient nephelometer, a $PM_{2.5}$ nephelometer, a PM_{10} nephelometer, a variable humidity nephelometer, continuous absorption measurements (multi-wavelength aethelometer), full ion measurements as a function of size (MOUDI sampler) as well as the usual compliment of IMPROVE aerosol measurements for comparison purposes. Resulting data would then need to be analyzed and input into various aerosol/optical models as described in Malm, et al., (1994).

NO_3 is the mass associated with nitrates, $\mu g/m^3$,

OC is the mass associated with organic carbon, $\mu g/m^3$,

EC is the mass associated with elemental carbon, $\mu g/m^3$,

IP is inorganic primary particulate matter (excluding primary sulfate and nitrate particles) with aerodynamic diameter $\leq$ 2.5 μm, $\mu g/m^3$,

CM is coarse particulate matter with aerodynamic diameter > 2.5 μm, but $\leq$ 10 μm, $\mu g/m^3$,

$b_{rayleigh}$ is light-scattering attributable to Rayleigh scattering, Mm^{-1} (i.e., inverse "mega-meters"), assumed to be 10 Mm^{-1}.

b_{ext} is the estimated extinction coefficient, Mm^{-1}.

The regional haze rule stipulates that reasonable progress is to be measured in terms of changes in "deciviews". Deciviews are defined as the natural logarithm of the ratio of the extinction coefficient to Rayleigh scattering (Pitchford and Malm, 1994)

$$\textbf{Deciview} = 10 \ln(b_{ext}/10) \quad (4.2)$$

Where the units of b_{ext} and light scattering due to Rayleigh scattering (i.e., the "10" in the denominator of the logarithmic expression) are both expressed in Mm^{-1}.

A *change* in deciviews, which is how we are to assess reasonable progress, is given by Equation (4.3). A 1 deciview change is equivalent to about a 10% change in b_{ext}.

$$\textbf{dv} = 10 \ln ((b_{ext})_{future} / (b_{ext})_{baselinet}) \quad (4.3)$$

Estimating Mass Associated With Components Of Particulate Matter. Sisler (1996) has developed a set of default assumptions for mass associated with each of the components of particulate matter on the right hand side of Equation (4.1). These are presented in Table 4.1. Information in Table 4.1 is very similar to that in Table 3.4 used for modeled attainment demonstrations of NAAQS for $PM_{2.5}$. Notice however, that fine inorganic particulate matter (IP) is assumed to consist entirely of crustal material. As discussed previously, there is no term "U" for unattributed mass. Finally, there is a term for coarse particulate matter included in Table 4.1 and in Equation (4.1)[15].

[15]Footnote has been deleted.

Table 4.1. Default Assumptions Used To Derive Aerosol Species Concentrations For Estimating Extinction Coefficients

(1) Species	(2) Formula	(3) Assumptions
Sulfate	4.125 (measured sulfur)	All elemental sulfur is from sulfate, & all sulfate is from ammonium sulfate
Nitrate	1.29 (measured nitrate)	Denuder efficiency is ~100% & all nitrate is from ammonium nitrate
EC	high + low temperature EC	All high temperature carbon is elemental
OC	1.4 (organic carbon)	Average organic molecule is 70% carbon
IP consists entirely of soil	2.2(Al) + 2.49(Si) + 1.63(Ca) + 2.42(Fe) + 1.94(Ti)	(Soil K)=0.6(Fe), FeO & Fe_2O_3 are equally abundant, a factor of 1.16 is used for MgO, Na_2O, H_2O & CO_3
$PM_{2.5}$	measured gravimetrically	Represents dry ambient fine aerosol mass for continental sites
CM (coarse mass)	$(PM_{10}) - (PM_{2.5})$	Consists only of insoluble soil particles

Considering effects of relative humidity (rh). Presence of high relative humidity increases the light scattering efficiency of fine particulate matter, particularly SO_4 and NO_3 species[16]. Thus, it is necessary to include a relative humidity adjustment factor, *f(rh)*, for these two components of $PM_{2.5}$ on the right hand side of Equation (4.1). Table 4.2 is information provided by Malm (2000a) which reflects data reported by Tang (1996). It denotes the relationship between relative humidity and the relative humidity adjustment factor used to relate b_{ext} to components of particulate matter.

The relative humidity adjustment factor in Equation (4.1) is applied to a 24-hour concentration of particulate matter. Therefore, we need to compute a daily average value for *f(rh)*. This is done by noting hourly values of relative humidity and then using Table 4.2 to estimate adjustment factors for each hour. For a given day, the *f(rh)* value used in Equation (4.2)

[16]As a result of field studies reported in Malm, et al., (2000), we assume the relative humidity adjustment factor (*f'(rh)*) for organic carbon (OC) is equal to "1.00".

Table 4.2. Relative Humidity Adjustment Factors As A Function Of Relative Humidity[17]

(1) Relative Humidity (%)	(2) Relative Humidity Adjustment Factor (f(rh))	(3) Relative Humidity (%)	(4) Relative Humidity Adjustment Factor (f(rh))	(5) Relative Humidity (%)	(6) Relative Humidity Adjustment Factor (f(rh))
≤ 36%	1.00	**57**	1.58	**78**	2.70
37	1.02	**58**	1.62	**79**	2.78
38	1.04	**59**	1.66	**80**	2.86
39	1.06	**60**	1.70	**81**	2.94
40	1.08	**61**	1.74	**82**	3.03
41	1.10	**62**	1.79	**83**	3.12
42	1.13	**63**	1.83	**84**	3.22
43	1.15	**64**	1.88	**85**	3.33
44	1.18	**65**	1.93	**86**	3.45
45	1.20	**66**	1.98	**87**	3.58
46	1.23	**67**	2.03	**88**	3.74
47	1.26	**68**	2.08	**89**	3.93
48	1.28	**69**	2.14	**90**	4.16
49	1.31	**70**	2.19	**91**	4.45
50	1.34	**71**	2.25	**92**	4.84
51	1.37	**72**	2.31	**93**	5.37
52	1.41	**73**	2.37	**94**	6.16
53	1.44	**74**	2.43	**95**	7.40
54	1.47	**75**	2.50	**96**	9.59
55	1.51	**76**	2.56	**97**	14.11
56	1.54	**77**	2.63	**98**	26.37

[17]Malm (2000a)

is the arithmetic mean of 24 hourly adjustment factors. This is illustrated in Example 4.1.

Example 4.1

Given: The following diurnal humidity information for a particular day.

Hour	Relative Humidity	Humidity Adjustment Factor (from Table 4.2)
1	75	2.50
2	80	2.86
3	80	2.86
4	90	4.16
5	95	7.40
6	98	26.37
7	98	26.37
8	98	26.37
9	95	7.40
10	90	4.16
11	80	2.86
12	60	1.70
13	40	1.08
14	30	1.00
15	30	1.00
16	30	1.00
17	30	1.00
18	35	1.00
19	50	1.34
20	50	1.34
21	60	1.70
22	60	1.70
23	70	2.19
24	70	2.19
		Total = 131.55

Find: The appropriate relative humidity adjustment factor for this day.

Solution: The appropriate adjustment for humidity is the mean relative humidity adjustment factor for the day. Adding the hourly adjustment factors, we get "131.55". Divide this total by "24" to get the mean daily relative humidity adjustment factor.

$$f(rh) = (131.55) / (24) = 5.48$$

If a State were interested in estimating the visibility extinction coefficient for the day in example

4.1, it would need to multipy the extinction attributable to sulfates and nitrates under dry conditions by “5.48".

Normalizing trends in regional haze. It is clear from the preceding example and from Equation (4.1) that relative humidity can have a substantial effect on estimated extinction coefficients, as well as on the relative importance of changes in different components of particulate matter can have on trends in regional haze. On the day in Example 4.1 for instance, light scattering efficiency of sulfates and nitrates increases from 3 m^2/gm to 16.4 m^2/gm. This magnifies the effect of any change in sulfates or nitrates on the estimated extinction coefficient. Because of the importance of relative humidity as a determinant of regional haze, it is necessary to normalize any apparent trend in the estimated extinction coefficient for differences in relative humidity. This enables us to assess whether an emissions control strategy will lead to reasonable progress reducing regional haze, without confounding effects of different relative humidity during the base and future periods.

There are two obvious potential ways to normalize trends in visibility for changes in relative humidity. The first is to assume that the same day to day changes in relative humidity which were observed in the base period calculations will occur in future years. Thus, one would use the relative humidity observations made on a specific day together with measured components of particulate matter on that day to compute the day specific visibility extinction coefficient on that day. Subject to the uncertainties posed by the empirically derived coefficients in Equation (4.1), this approach is the most likely to identify the 20% best and worst visibility days during the base period at a Class I site. However, the approach could lead to misleading conclusions if humidity observations were missing for some days or if the humidity observations are atypical in some way. Further, if a State or regional planning organization wanted to perform visibility calculations in a number of Class I areas, they would need to obtain hourly relative humidity data for each area.

The second approach to normalize trends in the extinction coefficient is to review relative humidity data over a long period of record to derive climatological estimates for relative humidity adjustment factors. These climatological estimates would then be used in Equation (4.1) to estimate visibility extinction coefficients. These estimates are more likely to reflect “typical” relative humidity at different times of year and, thus, expected future visibility extinction. Further, if the climatological estimates are made beforehand and published, this approach simplifies the estimates for States. Recently, the U.S. EPA has sponsored a project to examine measured hourly relative humidity data over a 5-year period within the United States to derive month-specific climatological mean humidity correction factors for each Class I area (U.S. EPA, 2000a). A potential downside of the “climatological” vs. “day-specific” approach could occur if there is a strong correlation between concentrations of hygroscopic components (i.e., SO_4 and NO_3) and measured relative humidity. If such were the case, use of average month- and location-specific climatological estimates could understate the importance of hygroscopic components as contributors to light extinction.

The U.S. National Park Service has examined data for each of 12 months from 20 Class I areas where relative humidity measurements are made. In nearly all cases, no statistically significant correlations were found between measured concentrations of SO_4, NO_3 and [SO_4 + NO_3] vs. daily values of relative humidity in a large majority of months. Results are displayed in Table 4.3. Further, deciview calculations have been made using day-specific vs. climatological values for the relative humidity adjustment factor for each of 10 years in 15 Class I areas. In 14 of the 15 areas, little, if any, difference is observed in the year to year calculations for the mean deciview values for the 20% worst and 20% best days nor in the trends. Some difference in the mean deciview value for the worst 20% days was observed in one Class I area. However, the overall trend in the mean worst and best deciview values for this site is similar using the two methodologies for deriving *f(rh)*.

Table 4.3. Correlations Between Hygroscopic Components of PM and Relative Humidity

# of Months with stat.significant positive correlation between hygroscopic species & rel.humidity	# of Class I Areas in this category for SO_4 vs. rel.humidity	# of Class I Areas in this category for NO_3 vs. rel.humidity	# of Class I Areas in this category for [SO_4+NO_3] vs. rel.humidity
0	**3 out of 20**	**6 out of 20**	**2 out of 20**
1-2	**9 out of 20**	**5 out of 20**	**6 out of 20**
3-5	**5 out of 20**	**8 out of 20**	**10 out of 20**
$\geq$ 6 (max.= 8)	**3 out of 20**	**1 out of 20**	**2 out of 20**

Results to date suggest there is a relatively weak correlation between hygroscopic components of PM and relative humidity and that choice of a "climatological" vs. "day-specific" method for computing *f(rh)* has little apparent effect on observed trends in visibility. As of this writing (January 2001), we are examining information to see whether use of climatological vs. day-specific values for *f(rh)* affects the identity of days chosen as the 20% best and 20% worst visibility and, if so, whether this has any effect on conclusions about future visibility obtained using the recommended modeled test for reasonable progress. At this time however, the climatological approach for estimating *f(rh)* appears viable, and it is the approach we recommend. Thus, States should use month- and Class I area-specific climatological relative humidity adjustment factors in in Equation (4.1), to rank visibility based on measured concentrations of major components of particulate matter. The resulting mean values of b_{ext} for the worst 20% and best 20% of sampled days are those which need to be improved (worst 20%) or not made worse (best 20%) in order for reasonable progress to be demonstrated.

<u>Recommendations</u>. States should characterize visibility in a Class I area by

estimating values for the light extinction coefficient (b_{ext}). The extinction coefficient should be estimated using monitored mass of particulate matter associated with SO_4, NO_3, OC, EC, IP and CM in Equation (4.1). States should use relative humidity adjustment factors in Equation (4.1) which reflect climatological mean surface relative humidity observations which are specific to the Class I area being analyzed and to the month of the year in which an observation has occurred.

4.2 How Do I Apply A Modeled Test To Demonstrate Reasonable Progress?

The purpose of a modeled test to demonstrate reasonable progress is to provide assurance that a proposed control strategy will result in reasonable progress goals being met when measured concentrations of particulate matter are used to estimate b_{ext} at some future date. The test we recommend has 6 steps.

1. For each Class I area, rank visibility on each day with observed speciated $PM_{2.5}$ data plus PM_{10} data for each of the 5 years comprising the base period.

2. For each of the 5 years comprising the base period, calculate the mean extinction coefficient for the 20% of days with worst and 20% of days with best visibility. For each Class I area, calculate the average mean extinction coefficients for worst and best days from the five year-specific values.

3. Use an air quality model to simulate base period emissions and future emissions. Use the resulting information to develop relative reduction factors for each component of particulate matter identified on the right hand side of Equation (4.1).

4. Multiply the relative reduction factors times the measured PM data during the base period to compute future daily values of the extinction coefficients. Rank these to determine the 20% best and 20% worst visibility for each of the five base years.

5. Using the results in Step 4, recalculate the future mean extinction coefficients for the 20% best and 20% worst visibility days in each of the five base years. Compute the future average mean extinction coefficients for the worst and best days by averaging the means from each of the 5 "future" years.

6. Convert the difference in average mean extinction coefficients between the future and base periods to a difference in deciviews. Compare the difference in deciviews to the goals for reasonable progress.

We describe each of these steps more fully below. States should follow the outlined procedure for each Class I area subject to a reasonable progress review. We conclude this subsection with an example illustrating the recommended modeled test for reasonable progress.

Step 1. Using monitored data, rank current visibility for each day with PM_{10}, $PM_{2.5}$ and speciated $PM_{2.5}$ measurements within a Class I area

Ranking should be performed separately for each of the 5 years comprising the base period. The extinction coefficient, b_{ext}, should serve as the basis for ranking. This coefficient should be calculated using Equation (4.1). Day-specific observations for mass associated with SO_4, NO_3, OC, EC, IP and CM, as defined in Table 4.1, should be used to calculate b_{ext} for each day. The appropriate month- and area-specific climatological mean value estimated in U.S. EPA (2000a) should be used for the relative humidity adjustment factor (*f(rh)*) in Equation (4.1).

Step 2. Calculate the average baseline extinction coefficient for the 20% of days with worst and the 20% days with best visibility.

For each of the 5 years in the base period, order all days considered in Step 1 from worst (highest extinction coefficient) to best (lowest extinction coefficient) visibility. For each year, note the 20% of days with worst and the 20% of days with best visibility. Calculate the arithmetic mean extinction coefficient for the identified 20% worst- and best visibility days in each year. Average the resulting 5 yearly mean extinction coefficients reflecting worst visibility. This represents the value subject to improvement (i.e., reduction) to meet the reasonable progress goal for regional haze. Average the 5 yearly mean extinction coefficients reflecting mean visibility on the days with best visibility. The resulting average mean extinction coefficient may not increase if reasonable progress is to be demonstrated.

Step 3. Estimate relative reduction factors (RRF) for each component of $PM_{2.5}$ and for CM.

This should be done by using emissions during the base period in air quality model simulations performed for a large number of days[18]. Air quality corresponding with future emissions (reflecting effects of growth and a control strategy under consideration) should be simulated for the same days. Take the (temporal) arithmetic mean concentration for each $PM_{2.5}$ component computed near the Class I monitoring site with future emissions and divide this by the corresponding arithmetic mean concentration for each component obtained with current emissions. The resulting quotients are the component-specific RRF's. Baseline and future $PM_{2.5}$ and PM_{10} are the predicted temporal quarterly arithmetic means of the predicted daily values near the Class I monitoring site with current and future emissions. The relative reduction factor for coarse mode (CM) particulate matter is given by the following expression.

$$(RRF)_{CM} = (PM_{10} - PM_{2.5})_{future} / (PM_{10} - PM_{2.5})_{baseline}$$

As we discuss in Section 4.3, we recommend that States derive separate sets of RRF values for "worst" and "best" visibility days identified in step 2.

---

[18]How many and which days to simulate is discussed in Section 4.3 and in Section 12.0.

Step 4. Multiply the relative reduction factors derived in Step 3 times measured daily values for each component of $PM_{2.5}$ and CM to get future daily estimates of $PM_{2.5}$, its components and CM on "worst visibility" and "best visibility"days.

These multiplications produce future estimates for SO_4, NO_3, OC, EC, IP and CM for each of the previously selected "worst" and "best" visibility days.

Step 5. Use the information developed in Step 4 to compute future values for b_{ext} for each of the previously identified "worst" and "best" days with monitored data. Compute mean values of b_{ext} for the sets of "worst" and "best" days in each year. Average the 5 mean extinction coefficients for "worst" and "best" visibility days.

Apply Equation (4.1) using the future predicted PM component concentrations for each of the previously identified 20% "worst" and "best" days. Use the area- and month-specific climatological mean default value for the humidity adjustment factor (*f(rh)*). Applying Equation (4.1) in this manner produces an estimate of the future value of b_{ext} for each of the previously identified "worst" and "best" visibility days for a Class I area. For days included in a given year of the base period, compute the arithmetic mean future extinction coefficient for the "worst" and "best" visibility days. This leads to 5 future estimated mean extinction coefficients for the "worst" and 5 future estimated mean extinction coefficients for the "best" visibility days. Compute the arithmetic mean of the 5 mean values for b_{ext} on the "worst" days, and the arithmetic mean of the 5 mean values estimated for the "best" visibility days. The resulting average mean values for b_{ext} on "worst" and "best" visibility days are used in Step 6.

Step 6. For "worst" visibility days and for "best" visibility days, convert the difference in future and current values of average mean values of b_{ext} to a difference expressed in deciviews. Compare these differences with the two previously established goals for reasonable progress to see if reasonable progress is demonstrated.

Apply the information produced in Step 5 in Equation (4.3) to estimate a difference in deciviews for previously identified days with future "worst" visibility and future "best" visibility. If the resulting change in deciviews is a negative number, this represents an improvement in visibility. The reasonable progress goal for "worst" visibility days is met if Equation (4.3) yields a negative number and the absolute value of this negative number is greater than or equal to the goal established for "worst" visibility days. The reasonable progress goal for "best" visibility days is met if the deciview value resulting from Equation (4.3) is ≤ 0. If goals for "worst" and "best" visibility days are both met, the modeled test for reasonable progress is passed.

We use Example 4.2 to illustrate the modeled test for reasonable progress. For ease of presentation, we assume there are only 10 speciated samples for PM in the first of 5 years comprising the base period. Since sampling will occur every third day, we anticipate a usual sample size of about 120 days per year. We go through the calculations for the first base year

and then furnish information regarding mean extinction coefficients for the other four base years to illustrate subsequent steps in the test.

Example 4.2

Given:

(1) Ten days have measured components of PM in a Class I area during the first year of a 5-year base period. The table below shows the measurements (in $\mu g/m^3$) for each of the 10 days. The table also shows the date of each measurement and the corresponding climatological relative humidity adjustment factor (made up for this example) for the appropriate month and area.

Day	Date	*f(rh)*	SO_4	NO_3	OC	EC	IP	CM	$(b_{ext})_{baseline}$	Rank
1	2/15	1.7	4.5	2.2	3.3	0.8	0.3	7.3		
2	3/15	2.9	5.1	1.7	2.9	0.7	0.4	9.1		
3	4/15	3.5	5.6	1.2	2.4	0.5	0.4	10.2		
4	5/15	3.7	6.5	1.1	2.8	0.5	0.6	9.8		
5	6/15	4.3	7.4	0.9	3.4	0.6	0.9	10.9		
6	7/15	4.6	8.3	0.7	3.5	0.6	1.0	11.3		
7	8/15	4.4	8.7	0.8	3.2	0.6	0.8	10.7		
8	9/15	4.1	8.0	0.9	3.1	0.7	0.7	8.2		
9	9/30	4.1	6.8	1.1	3.2	0.8	0.4	8.8		
10	10/30	2.7	5.3	1.4	3.2	0.7	0.6	8.1		

(2) Using procedures identical to those followed for year 1, the following table shows mean current values for b_{ext} for "worst" and "best" visibility days in each of the other 4 years comprising the base period.

Year	Mean $(b_{ext})_{baselint}$, Worst Visibility Days, Mm^{-1}	Mean $(b_{ext})_{baseline}$, Best Visibility Days, Mm^{-1}
2	190.3	62.1
3	170.1	87.4
4	180.6	75.3
5	165.0	82.6

(3) The following table shows component-specific relative reduction factors estimated using an air quality model to simulate future and current emissions. Component-specific relative reduction factors are computed as described in Section 4.3. The table also shows predicted mean future and current concentrations of $PM_{2.5}$ and of PM_{10} predicted near the Class I area's monitoring site. The current and future mean $PM_{2.5}$ and PM_{10} estimates are needed to calculate the RRF for CM.. For simplicity in this example, we are assuming that RRF values are similar for "worst" and "best" visibility days.

Modeled Output	SO_4	NO_3	OC	EC	IP	Est. Baseline Conc.	Est. Future Conc.
RRF	**0.85**	**0.95**	**1.00**	**1.00**	**0.85**	**-**	**-**
Mean PM_{10}	**-**	**-**	**-**	**-**	**-**	**28 µg/m^3**	**25 µg/m^3**
Mean $PM_{2.5}$	**-**	**-**	**-**	**-**	**-**	**11 µg/m^3**	**8 µg/m^3**

(4) The following reasonable progress goals have been established: (a) mean visibility on the 20% of days with worst visiblity should be improved by 2.0 deciviews; (b) mean visibility on the 20% of days with best visibility should not deteriorate.

Find: Is the control strategy simulated in this model analysis sufficient to meet the two goals for reasonable progress reducing regional haze in this Class I area?

Solution:

Step 1. Rank days according to their baseline extinction coefficients derived from measured PM data.

First, estimate the extinction coefficient for each day with the needed PM measurements. This is done using the information in "given #1" with Equation (4.1). For day 1 in year 1, the current extinction coefficient is:

$$b_{ext} = (3)(1.7)[4.5] + (3)(1.7)[2.2] + (4)(1)[3.3] + (10)[0.8] + (1)[0.3] + (0.6)[7.3] + 10$$

$$b_{ext} = 70.1\ Mm^{-1}$$

Current extinction coefficients for the remaining 9 days with monitored data in year 1 are calculated in a similar manner. The days are then ranked. The day with the highest extinction coefficient (i.e., worst visibility) is given a rank of "1". The results of these calculations are

Day	Date	*f(rh)*	SO_4	NO_3	OC	EC	IP	CM	$(b_{ext})_{baseline}$	Rank
1	**2/15**	**1.7**	**4.5**	**2.2**	**3.3**	**0.8**	**0.3**	**7.3**	***70.1***	***10***
2	**3/15**	**2.9**	**5.1**	**1.7**	**2.9**	**0.7**	**0.4**	**9.1**	***93.6***	***8***
3	**4/15**	**3.5**	**5.6**	**1.2**	**2.4**	**0.5**	**0.4**	**10.2**	***102.5***	***7***
4	**5/15**	**3.7**	**6.5**	**1.1**	**2.8**	**0.5**	**0.6**	**9.8**	***117.0***	***6***
5	**6/15**	**4.3**	**7.4**	**0.9**	**3.4**	**0.6**	**0.9**	**10.9**	***144.1***	***4***
6	**7/15**	**4.6**	**8.3**	**0.7**	**3.5**	**0.6**	**1.0**	**11.3**	***162.0***	***1***
7	**8/15**	**4.4**	**8.7**	**0.8**	**3.2**	**0.6**	**0.8**	**10.7**	***161.4***	***2***
8	**9/15**	**4.1**	**8.0**	**0.9**	**3.1**	**0.7**	**0.7**	**8.2**	***144.5***	***3***
9	**9/30**	**4.1**	**6.8**	**1.1**	**3.2**	**0.8**	**0.4**	**8.8**	***133.7***	***5***
10	**10/30**	**2.7**	**5.3**	**1.4**	**3.2**	**0.7**	**0.6**	**8.1**	***89.5***	***9***

displayed in the last two columns (shown in italics) of the table above. Based on these rankings, days 6 and 7 comprise the 20% of days with worst visibility. Days 1 and 10 comprise the 20% of days with best visibility.

Step 2. Calculate the average baseline extinction coefficient for the 20% of days with worst visibility and the 20% days with best visibility.

For year 1, mean worst visibility = (162.0 + 161.4) / 2 = 161.7 Mm^{-1}, and

mean best visibility = (70.1 + 89.5) / 2 = 79.8 Mm^{-1}.

Mean worst and best visibility for years 2-5 is provided in "given #2".

The table below summarizes mean worst and best visiblity for each of the 5 years in the base period.

Year	Mean $(b_{ext})_{current}$, Worst Visibility Days, Mm^{-1}	Mean $(b_{ext})_{current}$, Best Visibility Days, Mm^{-1}
1	**161.7**	**79.8**
2	**190 .3**	**62.1**
3	**170.1**	**87.4**
4	**180.6**	**75.3**
5	**165.0**	**82.6**

The "average mean baseline worst and best visibility", needed for reasonable progress calculations, is obtained by taking the arithmetic average of the mean worst and best visibility for the 5 years. Thus, the average mean worst visibility is given by

$$(b_{ext})_{baseline} = (161.7 + 190.3 + 170.1 + 180.6 + 165.0) / 5 = 173.5 \text{ Mm}^{-1}$$

The average mean best visibility is

$$(b_{ext})_{baseline} = (79.8 + 62.1 + 87.4 + 75.3 + 82.6) / 5 = 77.4 \text{ Mm}^{-1}$$

Step 3. Apply a model to develop component specific RRF's for SO_4, NO_3, OC, EC, IP and for coarse mode particulate matter (CM).

Using model results to develop values for the component-specific RRF's is discussed further in Section 4.3. Information in "given #3" provides RRF's for SO_4, NO_3, OC, EC and IP. An RRF value for CM can be derived from the information in "given #3" as follows.

$$(RRF)_{CM} = (25 - 8) / (28 - 11) = 1.00$$

Note that an RRF value for CM of "1.00" simply means that the strategy being simulated focused on reducing fine particle concentations rather than coarse particulate matter.

Summarizing, the RRF values for the components of $PM_{2.5}$ and CM are:

RRF	SO_4	NO_3	OC	EC	IP	CM
RRF	0.85	0.95	1.00	1.00	0.85	1.00

Step 4. Estimate future daily concentrations for each component of $PM_{2.5}$ and for CM for each of the previously identified "worst" and "best" visibility days by multiplying component-specific RRF's times the corresponding *observed* component concentration on each day.

In year 1, we previously identified days 6 and 7 as those included in the 20% of days with worst visibility (i.e., see Step 1). Similarly, days 1 and 10 are the 20% of days with best visibility. In this step, we need to estimate future concentrations for components of $PM_{2.5}$ and for CM for these two sets of days. This is done using information shown in tables presented in Steps 1 and 3.

Worst Days

Day 6: $[SO_4]_{future} = (RRF)_{SO4}\ [SO_4]_{baseline} = (0.85)\ [8.3] = 7.1\ \mu g/m^3$
$[NO_3]_{future} = (0.95)\ [0.7] = 0.7\ \mu g/m^3$
$[OC]_{future} = (1.00)\ [3.5] = 3.5\ \mu g/m^3$
$[EC]_{future} = (1.00)\ [0.6] = 0.6\ \mu g/m^3$
$(IP)_{future} = (0.85)\ [1.0] = 0.9\ \mu g/m^3$
$[CM]_{future} = (1.00)\ [11.3] = 11.3\ \mu g/m^3$

Day 7: $[SO_4]_{future} = (0.85)\ [8.7] = 7.4\ \mu g/m^3$
$[NO_3]_{future} = (0.95)\ [0.8] = 0.8\ \mu g/m^3$
$[OC]_{future} = (1.00)\ [3.2] = 3.2\ \mu g/m^3$
$[EC]_{future} = (1.00)\ [0.6] = 0.6\ \mu g/m^3$
$[IP]_{future} = (0.85)\ [0.8] = 0.7\ \mu g/m^3$
$[CM]_{future} = (1.00)\ [10.7] = 10.7\ \mu g/m^3$

Best Days

Day 1: $[SO_4]_{future} = (0.85)\ [4.5] = 3.8\ \mu g/m^3$
$[NO_3]_{future} = (0.95)\ [2.2] = 2.1\ \mu g/m^3$
$[OC]_{future} = (1.00)\ [3.3] = 3.3\ \mu g/m^3$
$[EC]_{future} = (1.00)\ [0.8] = 0.8\ \mu g/m^3$
$[IP]_{future} = (0.85)\ [0.3] = 0.3\ \mu g/m^3$
$[CM]_{future} = (1.00)\ [7.3] = 7.3\ \mu g/m^3$

Day 10: $[SO_4]_{future} = (0.85)[5.3] = 4.5\ \mu g/m^3$
$[NO_3]_{future} = (0.95)\ [1.4] = 1.3\ \mu g/m^3$
$[OC]_{future} = (1.00)\ [3.2] = 3.2\ \mu g/m^3$
$[EC]_{future} = (1.00)\ [\ 0.7] = 0.7\ \mu g/m^3$
$[IP]_{future} = (0.85)\ [0.6] = 0.5\ \mu g/m^3$
$[CM]_{future} = (1.00)\ [8.1] = 8.1\ \mu g/m^3$

Similar calculations are performed for each of the "worst" and "best" days in each of the other 4 years in the base period.

Step 5. Use the estimated future PM component concentrations to estimate future values for b_{ext} on each of the "worst" and "best" visibility days. Then estimate the mean "worst" and "best" visibility for each of the 5 meteorological years in the base period. Finally, estimate future values for average mean visibility on "worst" visibility days and average mean visibility on "best" visibility days.

Using future concentrations of PM components obtained in Step 4 in Equation (4.1), we can estimate corresponding values for b_{ext}.

For year 1

Worst Days

Day 6: b_{ext} = (3)(4.6)[7.1] + (3)(4.6)[0.7] + (4)[3.5] + (10)[0.6] + 0.9 + (0.6)[11.3] + 10 = 145.3 Mm^{-1}

Day 7: b_{ext} = (3)(4.4)[7.4] + (3)(4.4)[0.8] + (4)[3.2] + (10)[0.6] + 0.7 + (0.6)[10.7] + 10 = 144.2 Mm^{-1}

Future mean visibility on worst days = (145.3 + 144.2) / 2 = 144.8 Mm^{-1}

Best Days

Day 1: b_{ext} = (3)(1.7)[3.8] + (3)(1.7)[2.1] + (4)[3.3] + (10)[0.8] + 0.3 + (0.6)[7.3] + 10 = 66.0 Mm^{-1}

Day 10: b_{ext} = (3)(2.7)[4.5] + (3)(2.7)[1.3] + (4)[3.2] + (10)[0.7] + 0.5 + (0.6)[8.1] + 10 = 72.1 Mm^{-1}

Future mean visibility on best days = (66.0 + 72.1) / 2 = 69.1 Mm^{-1}

Similar calculations are performed for previously selected "worst" and "best" days in each of years 2-5. To illlustrate the modeled test for reasonable progress, assume these other calculations yield the following estimates for future mean b_{ext} on worst and best visibility days.

Year	Future Mean b_{ext} On Worst Visibility Days, Mm^{-1}	Future Mean b_{ext} On Best Visibility Days, Mm^{-1}
1	**144.8**	**69.1**
2	**137.7**	**62.0**
3	**169.4**	**79.3**
4	**153.0**	**67.8**
5	**154.0**	**69.4**

Using results in the preceding table, we see that the estimated future average mean extinction coefficient for the 20% days with worst visibility is

$(b_{ext})_{future}$ = (144.8 + 137.7 + 169.4 + 153.0 + 154.0) / 5 = 151.8 Mm^{-1}

The estimated future average mean extinction coefficient for the 20% days with best visibility is

$(b_{ext})_{future}$ = (69.1 + 62.0 + 79.3 + 67.8 + 69.4) / 5 = 69.5 Mm^{-1}.

Step 6. Convert the difference between future and baseline estimates of b_{ext} on "worst" visibility days and on "best" visibility days to differences in deciviews. Compare the two resulting differences to the two goals for reasonable progress. If both goals are met, the modeled test for reasonable progress is passed.

Use Equation (4.3) and the results generated in Step 5 to first estimate the difference in deciviews for days with worst visibility and then the difference in deciviews for days with best visibility.

For the 20% days with worst visibility,

$$dv = 10 \ln(151.8 / 173.5) = - 1.3 \text{ deciviews}$$

For the 20% days with best visibility,

$$dv = 10 \ln(69.5 / 77.4) = - 1.1 \text{ deciviews}$$

Note that the reasonable progress goal for days with best visibility is met. This follows, since Equation (4.3) yields a negative value, indicating that visibility is not predicted to deteriorate on these days but, in fact, is estimated to improve. However, the reasonable progress goal for the 20% days with the worst visibility is not predicted to be met. That is, the goal is to decrease light extinction on these days by 2.0 deciviews. The simulated control strategy only leads to a reduction in light extinction of 1.3 deciviews. If the modeled test for reasonable progress is not passed, States may do well to examine the numerical coefficients in Equation (4.1) on the days with "worst" (in this example) baseline visibility. Note that humidity adjustment factors are a part of the numerical coefficient for SO_4 and NO_3. Generally, the most effective strategies for improving visibility on days with poor visibility are those which realize large relative differences in components of particulate matter having large numerical coefficients on many of the base period's "worst" visibility days.

<u>Recommendations</u>. States should apply a modeled test for reasonable progress which has six steps.

1. Use extinction coefficients (b_{ext}'s) estimated from monitored components of particulate matter to rank light extinction on each day with PM measurements in each of five years comprising a base period. Days with the highest estimated extinction coefficients ("worst" days) are ranked first.

2. For each year in the base period, calculate the baseline arithmetic mean value for b_{ext} for the 20% of days in the sample with worst visibility and the 20% of days in the sample with best visibility. Take the arithmetic average of the 5 yearly mean extinction coefficients for days with "worst" and "best" visibility.

3. Use an air quality model to simulate relative reduction factors (RRF) for each of the six components of particulate matter used in Equation (4.1) to derive b_{ext}.

4. For each day included in the sample of 20% "worst" and 20% "best" visibility days, multiply each monitored component concentration times the RRF value specific for that component. The product is a day-specific estimate for the future concentration of each component of particulate matter.

5. Use Equation (4.1) to calculate future daily values for b_{ext} for each of the "worst" and "best" days in each year. For each year, compute the future arithmetic mean value for the "worst" and "best" days. Take the arithmetic average of the 5 yearly mean future extinction coefficients for days with "worst" and "best" visibility.

6. Convert the difference between future and baseline average mean extinction coefficents to deciviews. Do this twice. Once for the 20% of days with "worst" visibility and once for the 20% days with "best" visibility. Compare the two differences in deciviews with the corresponding two goals established for reasonable progress. If the differences meet or exceed the goals, the modeled test for reasonable progress is passed.

4.3 How Do I Select Appropriate Inputs To The Reasonable Progress Test?

In Section 4.2, we described the recommended modeled test for reasonable progress. An important part of the test requires using component-specific relative reduction factors (RRF's), obtained with models, to estimate future concentrations of these components and, subsequently, future visibility. We did not say much about how to estimate the RRF's. In this subsection, we address this issue. Second, there are several assumptions inherent in the recommended modeled test for reasonable progress. We identify these in this subsection and comment on their underlying rationale. More specifically, we address seven issues:

- how to estimate base period air quality in a Class I area without monitored data;
- how to handle a base year without data or with a small sample size;
- how to consider a day with missing data for one or several components of PM;
- use of the same days to estimate changes in visibility for the worst days and a different set of same days to estimate changes in visibility for the best days;
- which predictions to use to derive relative reduction factors;
- how many, what kind of days to use to develop RRF values, and
- use of relative reduction factors which are specific for days with poor vs. days with good visibility

Estimating baseline "worst" and "best" visibility in a Class I area without monitors. This issue will be addressed in guidance for estimating trends in regional haze from ambient data which is being developed by the U.S. EPA . This guidance is expected to be available during the first half of 2001. It is likely that the guidance for estimating regional trends in visibility will suggest using observations from a nearby Class I area, or interpolating from observations made in several nearby Class I areas. It is conceivable that the "observations" which get used will be mean values for b_{ext} for days with "worst" and "best" visibility. If this is the case, the modeled reasonable progress test will differ slightly from that applied in Class I areas with measured components of particulate matter. If the only information available for a Class I area is a mean value for b_{ext} for "worst" and "best" visibility days, the test will apply relative reduction factors

(RRF) to these two single values rather than individually to days in a distribution of "worst" and "best" days. Thus, the test becomes simplified. Derivation of RRF's is discussed in greater detail below.

What if a Class I area is so remote from other Class I areas that no available monitored data appear applicable? In this case, we recommend that a State model every day during one of the years in the base period. If this is not feasible, a State should follow guidance presented in Section 12.0 for selecting days to model for an application related to the annual NAAQS for $PM_{2.5}$. Next, base period emissions should be modeled for the selected days. Equation (4.1) should then be applied to estimate values for b_{ext} on each modeled day. Days should then be ranked according to the resulting values for b_{ext}. Next, note the 20% of modeled days with the worst visibility and the 20% of days with best visibility. Compute mean values of b_{ext} for days with "worst" and "best" visibility. Use the modeled results to estimate typical concentrations of SO_4, NO_3, OC, EC, IP and CM on the 20% best and worst (modeled) visibility days. Estimate average RRF values for the 20% worst and best (modeled) visibility days for each of the 6 components of PM used to derive b_{ext}. Apply component-specific RRF's to estimate future values for each of the 6 PM components needed to estimate future mean worst and best mean b_{ext}. Apply Equation (4.1) to estimate future mean best and worst values for b_{ext}. Finally, apply Equation (4.3) to see if the two reasonable progress goals have been met.

Considering a base year with little or no monitored particulate matter. First, we address a case where there are no monitored observations for one or two of the years in the 5-year base period. If this happens, apply the reasonable progress test as described in Section 4.2. The baseline average mean extinction coefficients should be estimated using the years within the base period with data. The second case is one in which one year has less data than the other years in the base period. If a year has less than half of the number of days with observations than the other years in a base period, do not include it when computing the average mean current extinction coefficient. Similarly, if a year does not have observations during a season in which there are ordinarily many days with poor (good) visibility, do not include it when estimating average mean values for baseline b_{ext} during days with "worst" ("best") visibility.

Considering days on which measurements for one or more of the 6 components of particulate matter needed to compute b_{ext} are missing. If a component is missing on a particular day with measurements, that day should, nevertheless, be considered in ranking daily b_{ext} values preparatory to identifying the 20% worst and 20% best visibility days. We recommend that States compute the 90^{th} percentile (i.e., close to the highest) and 10^{th} percentile (i.e., close to the lowest) concentration for the missing component from days where its measurements are available. If a State is trying to determine "worst" visibility days, assume the missing component's concentration is equivalent to its 10^{th} percentile value (i.e., is low). Conversely, If a State is addressing "best" visibility days, assume the missing component's concentration is equivalent to its 90^{th} percentile value (i.e., is high). Following this procedure will increase the sample size used to determine days with worst and best visibility. At the same time, the procedure ensures that missing data are not likely to play a major role in flagging

"worst" and "best" visibility days. That is, a day is identified as one of the worst or best visibility days *despite* the missing component.

Using a constant sample of days to estimate baseline and future visibility. For a typical Class I area, we anticipate there will be about 120 days per year having measurements needed to estimate $(b_{ext})_{baseline}$ with Equation (4.1). Thus, there should be about 24 "worst" and 24 "best" visibility days for each of the 5 years in the base period. It is conceivable that the identity of these "worst" and "best" days could change if emissions were altered to reflect net effects of controls and growth. The recommended test described in Section 4.2 assumes that the identity of the "worst" and "best" days remains unchanged. This is done primarily to avoid having to perform iterative analyses to identify future worst and best visibility days and to keep the test relatively simple and more readily understood. This assumption could cause improvement in visibility to be overestimated for the "worst" days and could also cause the test to overestimate the difficulty in preventing deterioration of visibility on the "best" days. However, for the reasons described below, we do not believe the effects of this assumption are substantial.

It is unlikely that there would be any wholesale change in the identity of "worst" or "best" days with future vs. current emissions. Analyses performed by Meyer, et al. (1997) have shown that the predicted ranked severity of high ozone days is largely unaffected by simulated controls and growth (i.e., highest days tend to remain the highest days after the effects of growth and conctrols are simulated). There is no reason to expect a different outcome for other secondary pollutants. If there are differences, we would expect these to occur near the borderline between the "worst" days and more moderate days. Meyer, et al. (1997) show that rankings are less stable as one examines days which are not characterized by high concentrations of ozone. However, the reason for this appears to be that the concentrations on such days are so similar to begin with. Thus, even though the identity of "best" days may be more subject to change, the resulting differences in the predicted concentrations on the best days is small, because the variability in these concentrations among different days is very small.

Because the reasonable progress test relies on mean values for b_{ext} on 20 or more "worst" visibility days and most of these days are unlikely to change, we would expect little difference in the outcome of the reasonable progress test. Further, because of the expected shape of the distribution of extinction coefficients, the arithmetic mean value for b_{ext} on the worst days is more heavily influenced by extreme days rather than those on the borderline between "worst" and more moderate light extinction. There could be differences in some "best" visibility days corresponding with pre- and post-control emissions. However, because the differences in concentrations of particulate matter on such days are likely to be relatively low, differences in the computed mean value for b_{ext} on "best" days are likely to be small. Further, any resulting difference in the reasonable progress test for "best" days is likely to be protective of the environment. If our recommended procedure leads to suspected problems in the outcome of a test, a State may perform a more rigorous version of the reasonable progress test (in which the identity of pre-control and post-control days changes) as part of a weight of evidence determination.

Selecting predictions to use in deriving RRF. Relative reduction factors should be developed for each Class I area. When a Class I area contains a monitoring site, the RRF estimates should be derived using predictions which are made "near" that site. "Near" is defined in Section 3.3. For each day, daily average surface predictions of each component of PM made near a monitor should be estimated. These nearby estimates should then be spatially averaged to estimate a spatially representative daily concentration. Spatially representative daily concentrations obtained for each modeled day with monitored data should then be averaged. This final average should be used to compute the RRF. Thus, component-specific RRF values for a Class I area with a monitor are the ratio of the temporally averaged spatial mean of nearby concentrations predicted with future emissions to that predicted with baseline emissions. The recommended procedure is illustrated by Table 3.5, Figure 3.1 and Example 3.3 for a grid whose cells are 12 km on a side. Note that for cells larger than 15 km on a side, no spatial averaging is necessary—States should just use the prediction in the cell containing the monitor.

If a Class I area does not contain a monitoring site, States should derive component-specific RRF values using a procedure which differs only slightly from that used in Class I sites with monitors. The difference occurs in how spatially averaged component concentrations are calculated for each modeled day. In this case, daily spatially averaged component concentrations should be estimated by taking the arithmetic mean value of predicted concentrations in all surface grid cells which include any part of the Class I area. Once this is done, the recommended procedure is identical to that for Class I areas with monitors. Note that days used to compute the RRF correspond to those modeled days for which "monitored concentrations" have been derived from measurements outside the Class I area.

Selecting days to derive RRF values. RRF values should be estimated by taking the ratio of future predictions averaged over several days to current predictions averaged over the same several days. It may often happen that a regional planning organization or a group of States decides to model effects of a strategy for numerous Class I areas simultaneously. As we note in Section 12.0, this may make it adviseable to simulate a full year so that relative reduction factor (RRF) values for each Class I area is based on a substantial number of observed "best" and "worst" days. For the "worst" days in the chosen year, the RRF for a component of PM should be estimated as the ratio of its arithmetic mean predicted value on the 20% worst days with future emissions to that with baseline emissions. Thus, the RRF should reflect values averaged over ~ 24 "best" and "worst" days in that year.

If it is not feasible to model an entire year, or if only a small number of Class I areas is to be considered, a State should examine when worst visibility (derived using Equation (4.1)) is observed to occur. Choose a sample of days from each quarter in which an incident of "worst" visibility occurs and calculate a RRF estimate for each component of PM. The appropriate RRF value would be applied to monitored "worst visibility" days. There is not, as yet, a good basis for suggesting a minimum number of days to choose for this purpose. However, information presented in Section 12.0 suggests that this number should be $\geq$ ~10 days.

Since meteorological conditions and/or emissions may be markedly different on "best" visibility vs. "worst" visibility days, we believe it is prudent to develop a separate set of RRF values for "best" visibility days. As with "worst" days, the preferred approach is to model an entire year and select an RRF value for concentrations averaged over the 20% "best" visibility days for each Class I area. If this is not feasible or only a limited number of Class I areas are to be considered in an analysis, States may review when "best" visibility days are observed to occur in the base period. Model $\geq \sim 10$ days with observed "best" visibility and average predicted baseline and future concentrations for each PM component. The RRF values are the ratios of the future to baseline modeled averages. The appropriate RRF values should then be used in concert with each observed "best" day to estimate future concentrations for each component on each identified "best" day.

<u>Recommendations</u>. States may assume that days which are identified as having the "worst" and "best" visibility with baseline emissions are representative of those with "worst" and "best" visibility with future emissions.

States may use fewer than 5 years of observations to derive present estimates for b_{ext} if this is dictated by available speciated estimates of particulate matter. In order to ensure that as large a sample as feasible is used to identify "worst" and "best" visibility days, States may assume a high value for a missing component of PM on days being ranked to determine "best" visibility. For purposes of determining "worst" visibility days, States may assume a low value for a missing component of PM.

Component-specific relative reduction factors (RRF) should be estimated using procedures similar to those used for the modeled attainment test for the NAAQS for $PM_{2.5}$.

RRF estimates for each Class I area should be based on predictions of components of particulate matter averaged over the 20% observed worst and 20% observed best visibility days in a chosen year. If it is not feasible to model an entire year, RRF values should be derived from modeled future and baseline concentrations averaged over as close to 10 days or more with "worst" and "best" observations.

Separate values for RRF should be derived for days with "worst" and "best" visibility.

5.0 If I Use A Weight Of Evidence Determination, What Does This Entail?

A weight of evidence determination examines results from a diverse set of analyses, including outcomes of the modeled attainment or reasonable progress test and hot spot analysis. Each type of analysis has an outcome or set of outcomes consistent with a hypothesis that a proposed control strategy is sufficient to meet the specified air quality goal within the required time frame. If such an outcome occurs, then results of that analysis support the hypothesis that the proposed strategy is adequate. Each analysis is weighed qualitatively, depending on its ability to address adequacy of a strategy and on the credibility of the analysis. If most (i.e., overall weight of) evidence produced by the diverse analyses supports the hypothesis, then attainment of the goal is demonstrated with the proposed strategy. The end product of a weight of evidence determination is a document which describes analyses performed, data bases used, key assumptions and outcomes of each analysis, and why a State believes that the evidence, viewed as a whole, supports a conclusion that the area will attain the NAAQS or meet the goals for reasonable progress. The further a modeled test (or hot spot analysis) is from clearly showing attainment of the air quality goal(s), the more convincing other evidence needs to be to reach a different conclusion in a weight of evidence determination.

Each weight of evidence determination will be subject to area-specific conditions and data availability. Area-specific factors may affect the types of analyses which are feasible , as well as the significance of each for a nonattainment or Class I area. Thus, decisions concerning which analyses to perform and how much credence to give each needs to be done on a case by case basis by those implementing the modeling/analysis protocol. In Section 5.1, we identify several broad types of analyses which can be used to corroborate one another in a weight of evidence determination addressing the NAAQS. States should utilize each of these types of analysis to ensure that conclusions regarding adequacy of a proposed strategy are based on a variety of independent analyses. In Section 5.2, we note that additional, optional corroborative analyses may be performed which address some of the default assumptions in the modeled attainment test more specifically. We provide several examples of such analyses, and identify conditions which should be met for them to be considered in a weight of evidence determination.

The goals for regional haze between the base period and 2018 relate to *trends* in light extinction rather than to some absolute value, as is the case for the NAAQS. Thus, weight of evidence analyses differ in some respects for visibility-related applications. Section 5.3 describes how to use weight of evidence in reasonable progress determinations.

The types of weight of evidence analyses identified in this section do not represent a comprehensive list. Analyses performed in addition to the recommended modeled attainment or reasonable progress tests will likely depend, in part, on ways to remedy shortcomings in the tests or underlying models perceived by those implementing the modeling/analysis protocol. Prior to expending a substantial amount of resources, it is a good idea for States to consult with the appropriate U.S. EPA regional office regarding the types of analyses they wish to include in a weight of evidence determination.

5.1 What Broad Types Of Analyses Should I Consider In A Weight Of Evidence Determination Related To The NAAQS For $PM_{2.5}$?

At a minimum, a weight of evidence determination should consider the following 3 types of corroborative analyses: application of air quality models, trends in observed air quality trends and estimated emissions and outcome of observational models.

Table 5.1 addresses each of the 3 recommended types of analyses. In the table, we identify factors which might cause those implementing the modeling/analysis protocol to give greater credence to a particular set of results (column (2)). We also identify outcomes for each analysis consistent with a hypothesis that emission reductions implied by a strategy are adequate to demonstrate attainment (column (3)).

We discuss each of the recommended types of analyses in the following subsections.

5.1.1 Air Quality Models

Weight given to results obtained with air quality models depends on how good the model performance is as well as the rigor with which the performance has been tested. Model applications for which an extensive observational data base exists have greater credence, especially if the data base includes monitored species concentrations which can be compared with species-specific model estimates and can be used to derive concentrations of $PM_{2.5}$ components which can be compared with model predictions.

Model results are likely to be more credible if based on predictions made for several rather than one or a small number of days. As explained by Hogrefe, et al. (2000) and in U.S. EPA (1999b), RRF values computed for 8-hour daily maximum ozone concentrations are apparently more stable (and, therefore, more reliably estimated) if they are based on a composite response averaged over about 10 or more modeled days. Hogrefe, et al. (2000) also recommend that model performance evaluation is most meaningful if based on a comparison of predictions and observations averaged over many (e.g., $\geq$ 10) days. The preceding references address ozone rather than $PM_{2.5}$ and its components. However, we believe the reasons given for basing model performance evaluation and RRF values on a composite of several days should hold for $PM_{2.5}$-related model applications as well.

One of the most uncertain inputs to a modeling analysis is the emission projections which must be made to a future year(s) of interest. This uncertainty is reduced if the projection period is short. Hence, weight of evidence provided by modeling is increased with short projection periods. If rigorous quality assurance and review is provided for the model's emissions and meteorological inputs, this may increase confidence that the model is yielding correct answers for the right reasons. Thus, rigor used in preparing model inputs also increases credibility given to the results.

Table 5.1. Recommended Analyses For A Weight Of Evidence Determination, Factors Affecting Their Credibility And Outcomes Consistent With Meeting The NAAQS

(1) Type of Analysis	(2) Factors Increasing Credibility of the Analysis	(3) Outcomes Consistent with Hypothesis That a Candidate Strategy will Lead to Attainment
Air Quality Models	**-good model performance** **-extensive observational data base available** **-RRF are based on a composite response on many days** **-selected episode days have observations near the design value (24-hr NAAQS)** **-short projection periods** **-carefully quality assured inventory** **-confidence in meteorological inputs** **-good ability to pose and address questions about a strategy's adequacy** **-other analyses tend to corroborate conclusions**	**-the modeled attainment test is passed, and** **-hot spot analyses do not produce estimates > 65 µg/m^3 (24-hr NAAQS) or > 15.0 µg/m^3 (annual NAAQS), or** **-the attainment/hot spot tests are nearly passed, the control strategy requires additional (unmodeled) reductions and efforts are underway to subsequently review/refine the strategy** **-commitment is made to deploy monitors at locations not passing the hot spot test** **-substantial modeled improvement in air quality is predicted using several measures described in Section 5.1.1** **-similar conclusions are reached with other peer reviewed models**

Table 5.1. Recommended Analyses For A Weight Of Evidence Determination, Factors Affecting Their Credibility And Outcomes Consistent With Meeting The NAAQS (continued)

(1) Type of Analysis	(2) Factors Increasing Credibility of the Analysis	(3) Outcomes Consistent with Hypothesis That a Candidate Strategy will Lead to Attainment
Analysis of Air Quality and Emissions Trends	**-current or future (air quality model) predicted design value is within a few $\mu g/m^3$ above the concentration specified in the NAAQS** **-extensive monitoring network exists** **- trends are available for all or most of the 6 major components of $PM_{2.5}$, as well as for $PM_{2.5}$ itself** **-statistical model used to normalize trend for meteorological differences explains much variance** **-short projection periods used in the analysis** **-a pronounced, statistically significant downward trend is apparent** **-similar conclusions are reached using several trend parameters** **-continued, comparable relative reductions in emissions are provided for** **-contemplated strategy focuses on components which are similar to those which were addressed by past strategies.**	**-a pronounced downward normalized trend exists in the site-specific design value at all sites with design values greater than the level of the NAAQS.** **-Using projected emissions to extrapolate the air quality trend line to the required attainment date indicates a 24-hr or annual design value $\leq$ the concentration specified in the NAAQS.** **-Other observed air quality trend parameters also show a substantial improvement.**

Table 5.1. Recommended Core Analyses for a Weight of Evidence Determination, Factors Affecting Their Credibility and Outcomes Consistent with Meeting the NAAQS (concluded)

(1) Type of Analysis	(2) Factors Increasing Credibility of the Analysis	(3) Outcomes Consistent with Hypothesis That a Candidate Strategy will Lead to Attainment
Use of Observational Models	**-an extensive monitoring network exists** **-precursor and indicator species are measured using instruments with appropriate sensitivity** **-monitoring sites appear spatially representative** **-data have been quality assured, and results are self-consistent** **-results of a given observational model are stable and plausible physical explanations exist for findings** **-similar conclusions are reached using a variety of observational approaches**	**-Findings indicate sources controlled in the candidate strategy are important causes of observed $PM_{2.5}$.** **-Analysis of indicator species suggests the direction of the strategy (e.g., emphasis on NH_3 or NOx to reduce nitrates) is appropriate.**

Proper selection of episode days increases confidence in the results of a modeled attainment test. For the annual NAAQS this means including a balance of days for each of the four seasons, as well as a balance of days where $PM_{2.5}$ concentrations are high and low. For the 24-hour NAAQS, considering days with concentrations near site-specific design values increases confidence that relative reduction factors, developed for use in the modeled attainment test, are appropriate.

Confidence in the quantitative results of a modeled attainment test is greater if corroborative, more qualitative, analyses support conclusions about appropriateness of a strategy. Finally, of the analyses available, modeling reflects the most comprehensive attempt to integrate emissions and meteorological information with atmospheric chemistry. As such, modeling has the greatest capability to address questions about adequacy of a strategy to meet air quality goals in the future. Thus, States should include modeling results in a weight of evidence determination, and these should ordinarily be very influential in deciding if attainment will occur.

The outcome from modeling supports use of a proposed strategy for attainment if the modeled attainment and hot spot tests, described in Section 3.0, are passed. If the hot spot test is not passed, a commitment to deploy monitors at such locations should be an important consideration in approving an attainment demonstration. In general, the closer modeled output is to passing the attainment and hot spot tests, the easier it is for other analyses to produce evidence which supports attainment. If a modeled attainment or hot spot test is not passed, selection of a strategy which substantially reduces emissions of one or more of the major components of $PM_{2.5}$ and an agreement to perform a subsequent review (based on improved data bases/tools) to refine the strategy, if necessary, can be considered in a weight of evidence determination if other modeling outputs and other analyses support a conclusion that the current selected strategy may lead to attainment. Other model-produced indicators that a proposed strategy may be adequate are (1) model predictions show major improvements in air quality using a variety of metrics, and (2) other peer reviewed atmospheric simulation models suggest attainment occurs using modeled attainment test results or other outputs.

We recommend that at least 3 additional model outputs be examined in weight of evidence determinations to provide assurance that passing or nearly passing the recommended modeled attainment test indicates attainment. Like the test, and for similar reasons, each of these additional outputs reflects *relative* changes in predicted air quality. States may use other model outputs (not described herein) in a weight of evidence determination as well.

1. Compute the relative change in frequency of surface grid-days > 65 μg/m^3 (or 15.0 μg/m^3 for annual applications) in the nonattainment area.

This output reflects the frequency with which predicted daily average concentrations exceed the concentration specified in the NAAQS for $PM_{2.5}$. Such a measure is not directly related to the form of the NAAQS. Further, if current and future predictions are subject to a systematic bias, this output could be misleading. However, if modeled episodes are chosen to represent a variety of meteorological conditions under which concentrations specified in the NAAQS are exceeded at one or more monitoring sites, and the model performs well using the measures described in Section 16.0, a large reduction in the frequency of predicted daily concentrations greater than 65 μg/m^3 (or 15.0 μg/m^3) is consistent with a conclusion that a proposed strategy would meet the NAAQS.

2. Compute the relative change in the number of grid cells in the nonattainment area with predicted 24-hr (annual) $PM_{2.5}$ > 65 μg/m^3 (15.0 μg/m^3)

This output estimates reduction in the pervasiveness of estimated $PM_{2.5}$ concentrations in excess of the concentrations specified in the NAAQS. It is subject to the same caveats as the preceding output. One additional complication may occur if there are not many surface grid cells in which current emissions lead to concentration estimates exceeding those specified in the NAAQS. This measure should not be used if this number is small.

3. Compute the relative change in the total difference ($\mu g/m^3$ -day) of daily predictions > than the concentrations specified in one or both of the NAAQS for $PM_{2.5}$.

Although not the same, this output is similar in concept to the change in the "dosage" to concentrations greater than those specified in the NAAQS. For applications related to the 24-hour NAAQS, we recommend using Equation (5.1). A similar equation can be used for applications related to the annual NAAQS by replacing "65" in the numerator and denominator with "15.0".

(5.1)

Where

RD = Relative Difference
C_{ij} = Predicted 24-hour concentration for day i and grid cell j greater than 65 $\mu g/m^3$
N = Total number of days
G = Total number of surface grid cells.

This metric is subject to the same caveats as the preceding two metrics.

5.1.2 Analysis Of Air Quality And Emissions Trends

Using trend information in a weight of evidence determination requires two preliminary steps: choosing a trend parameter(s) and normalizing fluctuations in the chosen parameter for year to year differences in meteorology.

Choosing a trend parameter. The preferred approach is to look at trends in the components which comprise most of $PM_{2.5}$ (i.e., usually, SO_4, OC, IP and/or NO_3). If these measurments are not available, then observed trends in measured $PM_{2.5}$ is probably the next best choice. As of 2000, the period of record for measurements of $PM_{2.5}$ and/or its components is not substantial in many locations. Thus, initially, States may wish to use visibility (e.g., measured extinction coefficients) or PM_{10} as a surrogate trend parameter. If a surrogate trend parameter is used, States should take precautions (identified shortly) in interpreting the results in a weight of evidence determination.

A second decision which needs to be made in choosing a trend parameter is, which member in a temporal distribution of observed concentrations to track? That is, do I choose a mean value, a 50^{th} percentile value, a 98^{th} percentile value, or what? We recommend that States

choose trend parameters which are as close as possible to the entities used in the modeled attainment test. Thus, for applications related to the annual NAAQS, the ideal is the observed quarterly mean concentration averaged over 3 consecutive years. For the 24-hour NAAQS, the ideal is the observed quarterly 98th percentile concentration averaged over 3 consecutive years. For a number of reasons however, the ideal may not be practical (e.g., insufficient number of observations, large number of resulting trends leading to difficulties in intepretation, etc.). Thus, for practical reasons, it is acceptable to look at fewer, more robust indicators (e.g., annual 50th percentile concentrations, 90th (rather than 98th) annual percentile concentrations). There are, undoubtedly, additional trend parameters which might provide useful insights for some applications. Additional trend parameters may be considered for use on a case by case basis.

Normalizing observed trends for meteorological differences. The second step needed prior to interpreting results of trend analysis is to normalize changes in the chosen trend parameter(s) for year to year fluctuations in meteorology. There are several approaches which have been used to do this. These include (a) using peaks averaged over multiple years (California Air Resources Board, 1993), (b) filtering techniques (Rao and Zurbenko, 1994, Flaum, et al., 1996, Milanchus, et al., 1997, 1998, Hogrefe, et al., 2000), (c) using a probability distribution of meteorological severity based on climatological data (Cox and Chu, 1993, 1996), and (d) using CART analysis to identify meteorological classes and selecting days from each year so that the underlying frequency of the identified meteorological classes remains the same (Stoeckenius, 1990, Deuel and Douglas, 1996). A brief summary of the strengths and weaknesses of these and other techniques will be available in U.S. EPA's *$PM_{2.5}$ Data Analysis Workbook* (U.S. EPA, 2001), expected to be completed early in 2001.

Interpreting a normalized trend. A curve is fit through the normalized trend and extrapolated to the year in which the air quality goal is to be met. Extrapolations are made by considering past trends as well as past and projected emission reductions. If the trend is statistically significant, the extrapolated value for the attainment year is at or below the air quality goal, and projected relative emission reductions are comparable or greater than reductions occurring during the period for which the trend is constructed, results of the trend analysis suggest a strategy will be adequate. This procedure is illustrated in the following example.

> **Example.** A proposed control strategy focuses on reducing sulfate concentrations. Estimate the relative reduction in emissions (e.g., SO_2 or NH_3) occurring during the period corresponding to the observed normalized trend in the 98th (or 90th if data limitations are a problem) percentile sulfate concentration in the nonattainment area. Use the estimated emission trend in concert with the normalized air quality trend to determine an "emission reduction sensitivity factor" (e.g., ($\mu g/m^3$)/(percent emission reduction)). Multiply the sensitivity factor times the percent reduction in emissions projected between the current period and the required attainment date. Note the resulting estimated difference in sulfate concentrations. Since the strategy focuses on reducing sulfates (and other components are not much affected), subtract the computed difference in the sulfate concentration from the current design value for $PM_{2.5}$ to get a projected design value. If

the projected design value is # to the concentration specified in the NAAQS for $PM_{2.5}$, the trend analysis supports a hypothesis that a proposed control strategy will suffice to reach attainment by the required date. Note that the procedure works even if a trend parameter is used which differs from the form of the design value (e.g., 90^{th} rather than 98^{th} percentile for 24-hr NAAQS), provided the slopes in the trend parameters are similar.

Some precautions. The trend analysis we suggest assumes that a linear extrapolation of observed past correspondence between monitored design values and estimated emission changes will accurately describe future air quality. This assumption probably works best if the current design value is already close to meeting the air quality goal. Weight given to trend analyses depends on several other factors as well. The more air quality data available and the greater variety of trend parameters which show major improvements, the more credible the results. Weight of evidence produced by trend results is higher if the procedure used to normalize the trend for meteorological differences explains much of the variability attributable to these differences. In addition, trend analysis is more believable if it is not necessary to extrapolate very far into the future. Trend analysis is most credible if the contemplated strategy is similar to a past strategy (e.g., both strategies focus on reducing sulfates). For example, if a past strategy focused on reducing sulfates, but a future one envisions controlling OC, there is no guarantee that ambient OC will respond similarly to changes in emissions.

Finally, we previously noted that extra precaution is needed in interpreting results of trend analysis when a surrogate for $PM_{2.5}$, such as PM_{10}, is used as a trend parameter. The methodology described in the preceding example is only valid if the past emission reductions corresponding with the observed air quality trend consisted largely of efforts to reduce the fine fraction (i.e., ≤ 2.5 μm) of PM_{10}. We would not expect a past trend in PM_{10} to be indicative of what might happen to future concentrations of $PM_{2.5}$ if past emission reductions focused on the coarse fraction of PM_{10}.

5.1.3 Use Of Observational Models

Observational models take advantage of monitored data to draw conclusions about the relative importance of different types of $PM_{2.5}$ emissions or precursors as factors contributing to observed $PM_{2.5}$, as well as inferences which might be drawn about the effectiveness of various strategies to reduce $PM_{2.5}$ concentrations. There are at least 3 approaches with potential for doing this: source apportionment (i.e., "receptor") models, hybrid approaches in which air quality observations are combined with meteorological observations in some manner, and indicator species approaches.

Receptor models. There is a large body of literature describing the theory and use of receptor models to identify and/or apportion sources which may be contributing to monitored air quality. Our intent is merely to highlight major approaches and to illustrate how they might be used in a weight of evidence determination. A review by Seigneur, et al. (1997), which is also summarized in Seigneur, et al. (1999), contains a more complete description of the major

approaches, summarizes findings obtained in a number of applications, and provides an extensive list of references. Receptor models are most useful for identifying contributions of various source categories to observed *primary* components of particulate matter.

There are two major types of receptor models. The first type is the chemical mass balance model (CMB). A description and user's guide is available for the CMB model (U.S. EPA, 1990). This model assumes that the user already has a good idea of what source categories potentially contribute to observations at a monitoring site. Speciated emissions profiles for all source categories are then compared with speciated air quality measurements at a monitor on a given day. A combination of source category contributions which minimizes observed differences between calculated and observed concentrations for a set of PM species (i.e., "fitting elements") is derived. This "optimum" combination is derived using statistical software which calculates the square of the differences between observations and predictions of all of the fitting elements. Each resulting difference is weighted according to the uncertainty attendant with measuring or estimating the corresponding fitting element. It is the sum of these weighted differences which gets minimized. The resulting solution represents each source category's estimated contribution to measured $PM_{2.5}$ at a monitor on the day in question. States wishing to apply CMB to address the annual NAAQS for $PM_{2.5}$ should calculate source contributions for a representative sample of days and use the resulting information to estimate the average relative importance of each of the source categories. We address the issue of what is a "representative sample of days" in Section 12.0.

One key assumption of the CMB approach is that there is no substantial chemical transformation of particulate matter between the point of emissions and the site of the ambient measurements. Thus, the technique is limited when a large portion of measured $PM_{2.5}$ is secondary particulate matter. If such is the case, CMB can still be used to provide insights in several ways. First the analysis can focus on apportioning only those components of $PM_{2.5}$ which largely consist of primary particulate matter (i.e., IP, EC and that portion of OC which is identified as being emitted as primary organic particulate matter[19]). A second approach is to continue to apportion total $PM_{2.5}$ using "fictitious" source categories whose "source profiles" consist entirely of "sulfate", "nitrate" and "secondary OC". The resulting solution, by itself, will not lead to any insights about the sources of "sulfates", "nitrates" or "secondary OC", but insights regarding contributors to the primary components of $PM_{2.5}$ (IP, EC , primary OC) are possible. As we describe later, CMB estimates, when combined with meteorological analyses in hybrid models, could provide insights into possible sources of secondary particulate matter as well.

Multi-variate statistical models comprise the second type of receptor model. Unlike with CMB, States do not have to make a priori assumptions about contributing source categories. Another distinction from the CMB approach is that multi-variate models look at day to day

---

[19]Gray and Cass, (1986) and Schauer, et al., (1996) describe procedures to (a) distinguish primary from secondary OC, or (b) identify organic species appropriate to use as fitting elements in a CMB model.

variations in speciated observations rather than focusing on individual days. The object of this is to identify sets of chemical species which track one another well. Statistical methods like cluster analysis or principal component analysis are used to identify groups of days or groups of chemical species which are associated with one another. Knowing which days are assoicated with one another can lead one to examining other attributes of these days to provide clues about significant sources. In particular, insights are possible when identified clusters of days are considered in hybrid models incorporating meteorological variability. Knowing which species have high temporal correlations with one another provides a clue about the possible identity of significant source categories. One outcome of these analyses is an identification of statistically-identified "sources" accounting for variability in the monitored data. An essential part of the analysis is to relate these "sources" to actual source categories in the area being studied. This could be done by comparing each identified "source" with available source speciation profiles or with other attributes of the area. Once actual, physically meaningful sources are identified, a CMB-type approach can be used to provide an estimate of the contributions of identified source categories to measured particulate matter on specific days.

Results obtained with multi-variate statististical methods, like cluster or principal component analysis, can sometimes produce physically unrealistic solutions (e.g., negative "factor loadings"for identified "marginal" sources). Recent multi-variate models include constraints to avoid physically unrealistic solutions. The Positive Matrix Factorization (PMF) technique (Paatero and Tapper, 1994) is one such technique. A user's manual (including a tutorial) for this approach is available for downloading at ***ftp://rock.helsinki.fi/pub/misc/pmf***. Papers by Huang, et al. (1999) and Polissar, et al. (1998) provide good examples of how PMF can be used. UNMIX is a second recently available multi-variate statistical model which incorporates constraints to avoid physically unrealistic solutions. Henry (1997) has prepared a conceptual description of the principles underlying this model. A user's manual is available (Henry, 2000). UNMIX and PMF have recently been applied to a data base developed in Phoenix. Information relating to this and other applications is described in Willis (2000). In addition to identifying "sources", UNMIX and PMF provide estimates of "source"contributions to monitored concentrations of components of particulate matter. Unlike with CMB estimates, these estimates represent "average" rather than day-specific contribution estimates. As such, this output of PMF and UNMIX may be particularly useful for applications related to the annual NAAQS.

Hybrid models. Hybrid models use results obtained with CMB or multi-variate receptor models in concert with other information. These techniques are needed to provide insight about contribution of potential sources to observed secondary particulate matter. Other information used in hybrid models includes some or all of the following: location of sources of emissions, composition of emissions associated with different identified sources, spatial correlation of $PM_{2.5}$ or one or more of its components measured at different monitoring sites, correlation of high $PM_{2.5}$ or its composition with wind direction or other meteorological variables, and use of trajectory analysis.

Location and composition of potential sources of $PM_{2.5}$ or its precursors is needed if one is to make use of meteorological information to infer likely important contributors to secondary particulate matter. Associations of high observed $PM_{2.5}$ or secondary components of $PM_{2.5}$ with a subset of wind directions/speeds, suggests a locus for likely important contributors to secondary particulate matter. If there are certain primary tracer species of particulate matter which seem to track high observed secondary particulate matter, one could search locations along the wind track suggested by the wind data to locate areas where sources of the tracer and sources of precursors of secondary particulate matter exist. Relating wind direction to high secondary particulate matter observed at different monitoring sites allows one to locate potential important source areas for secondary particulate matter with greater confidence through triangulation. Henry, et al. (1994) and Henry (1997a, 1997b and 1997c) have published several papers which present an approach utilizing data from several monitoring sites, and show how it can be used to infer sources which likely contribute to observed sulfate concentrations.

Trajectory models can be used together with receptor models to identify potential regional sources of secondary particulate matter. One trajectory model commonly used for similar purposes with ozone-related applications is available through the National Oceanic and Atmospheric Administration (NOAA, 1999). Trajectory models are used to trace likely previous locations of pollutant laden air sampled at monitoring sites. Thus, they are referred to as, "back-trajectories". Because we can only have a limited amount of confidence in the accuracy of any single back-trajectory, and a number of potentially important factors are ignored in computing a trajectory, these methods require trajectories to be constructed for many days. Similar trajectories can then grouped to see whether each group is characterized by distinctive differences in measurements of $PM_{2.5}$ or its components. An alternative approach for using trajectory models and receptor models together is to note days with high measured secondary components of $PM_{2.5}$ (i.e., SO_4, NO_3 or secondary OC) and contrast this with trajectories computed for other days. If days with high measurements are characterized by trajectories which suggest longer residence times over certain areas, this provides some evidence that sources of PM precursors in such areas may contribute to high levels of measured secondary particulate matter. Much use has been made of trajectory models to estimate possible need for making regional reductions in ozone precursors. The methods used appear equally applicable for identifying possible sources of secondary particulate matter. (Poirot, 1998, Schichtel and Husar, 1995, Schichtel, 1999).

Indicator species approaches. Indicator species approaches are based on the predicted sensitivity of a secondary pollutant's concentration to changes in different precursors for that pollutant. Predicted sensitivity is determined from simulations performed with the chemical kinetics mechanism used in air quality simulation models. Based on these simulations, it is possible to identify ratios of certain species which are good indicators of whether a secondary pollutant is sensitive to reductions in precursor A or precursor B. The approach is an observational one, because the next step is to examine monitored data to calculate observed ratios of the indicator species. Values of observed ratios are then compared to previously identified ranges for which the secondary pollutant is sensitive to reductions in precursor A and ranges for which reductions in precursor B appears most effective. Although results may

sometimes fall into a "gray area" where more than one precursor may affect ozone or secondary particulate matter, the indicator species approach provides a potential, quasi-independent means for assessing whether a proposed strategy to reduce a secondary pollutant is likely to be efficient in doing so.

As of late 2000, nearly all applications of indicator species approaches have addressed ozone-related problems. Sillman (1995), Sillman (1998) and Lu and Chang (1998) provide good descriptions of the method, identify key ratios and illustrate application of the approach. Even though the preceding ratios are oriented toward ozone, they can provide insight into why modeled concentrations of secondary particulate matter are sensitive to changes in VOC, or NOx emissions (Pun and Seigneur, 1999).

Ansari and Pandis, (1998) have developed an indicator ratio of species and applied it to several combinations of secondary particulate matter present under different environmental conditions. They use this ratio to predict how mass of particulate matter will respond to reductions in sulfate, nitrate and ammonia. Blanchard, et al. (2000) have also examined how indicator species might be used to assess whether particulate nitrate concentrations are limited by NOx or by ammonia emissions using mechanisms which incorporate reactions dealing with secondary particulate matter. These authors identify two ratios of indicator species which appear potentially useful for identifying limiting precursors for secondary nitrate particulate matter: (1) the ratio of particulate ammonium plus gas-phase ammonia over the sum of nitric acid plus particulate nitrate plus particulate sulfate, and (2) the ratio of particulate to total nitrate. It is likely that additional indicator species approaches will be identified as the user community gains more experience with chemical mechanisms incorporating secondary particulate formation and more speciated particulate and gas phase ambient measurements become available.

Potential utility of observational models in weight of evidence determinations. Observational models are potentially useful for assessing whether a proposed strategy is oriented toward source categories whose emissions appear to be associated with current observed $PM_{2.5}$. Thus, they are useful for confirming whether a modeled strategy (a) is addressing likely important source categories of primary particulate matter, and/or (b) is addressing likely sources or areas in which precursors for secondary particulate matter originate. However, their ability to estimate *how much* control is needed is limited, unless one can justify assuming an approximately linear relationship between $PM_{2.5}$ or precursor emissions and observed concentrations of $PM_{2.5}$ and/or its components. Thus, observational approaches are suitable to corroborate results obtained for $PM_{2.5}$ or its components with more quantitative techniques, like air quality models. Observational models can be used to examine days which have not been modeled with an air quality model, as well as days which have been modeled. The resulting information may be useful for drawing conclusions about the representativeness of the responses simulated with the air quality model for a limited sample of days.

In summary, if conclusions drawn with one or more observational models suggest that the types of sources to be controlled under a proposed strategy are those that appear associated with

high $PM_{2.5}$ and/or are those to which observed $PM_{2.5}$ or its major components are sensitive, this supports a hypothesis that the strategy is effective.

Interpreting results from observational models requires familiarity with the location being studied, and requires considerable subjective judgment. Strength of the evidence produced by observational models is increased if an extensive monitoring network exists and at least some of the monitors in the network are capable of measuring pollutants to the degree of sensitivity required by the methods. Results are more believable if they do not vary substantially in response to different assumptions made in applying the statistical algorithms underlying many of the models. Evidence produced by observational models is also more compelling if several techniques are used which complement one another and produce results which are consistent. Results are also more believable if plausible underlying physical/chemical explanations for them can be developed. Indications of a strong quality assurance analysis of collected data and measurements that are made by a well trained staff also lend credence to the results.

Recommendations. Weight of evidence determinations are best performed using a variety of diverse analyses in a corroborative fashion. Prior to its application, each selected analysis should have identified outcomes consistent with concluding a proposed strategy is adequate. At a minimum, States should consider the following corroboratory analyses in a weight of evidence determination: (1) output from air quality model(s) (i.e., modeled attainment test results plus other indicators), (2) air quality and emission trend analysis, and (3) interpretation of results obtained with observational models. If it is not feasible to include one or more of the recommended corroboratory analyses, the reasons should be documented.

5.2 What If I Want To Consider Additional Corroborative Analyses?

The list of analyses in Section 5.1 is not an exhaustive one. A State may use other types of analyses in a weight of evidence determination to supplement those identified in Section 5.1. To have another type of analysis considered, a State should identify why it believes the analysis will produce information which has a bearing on attainment of the NAAQS. In addition, the procedure to be used in applying the method and the extent of the data base available to support it should be identified. Finally, prior to application of the method, a State should identify outcomes which would be consistent with a hypothesis that a proposed emission reduction strategy will lead to attainment.

Choice of additional corroborative analyses is, in part, a function of the available data base and analytical tools, as well as questions posed by the outcomes of the analyses described in Section 5.1. For purposes of illustration, we identify some additional analyses of the sort which States might consider. None of these is required, and States may well choose to consider other optional analyses or no optional analyses at all.

Quantifying uncertainty associated with air quality model estimates. In this guidance, we recommend that "uncertainty" be accounted for using a modeled attainment test which uses models in a "relative" sense and by recognizing that use of corroboratory analyses may be desirable in a weight of evidence determination. Thus, we account for uncertainty in a qualitative way, without actually estimating it.

States may find it useful to quantify estimates of uncertainty and then use these results qualitatively in a weight of evidence determination. In Section 17.0, we identify three sensitivity tests which may be useful for this purpose. The first of these is one which has been proposed by Reynolds, et al., (1996). This test is to prepare "alternative base case" emission estimates, reflecting reasonable alternative assumptions about current emissions which lead to comparable or better model performance. Note differences in projected design values from these alternative current emissions. A second test is to assume alternative (reasonable) growth assumptions. This could reflect using differing growth rates or placement of new sources in different, equally probable, locations. Note the differences in projected design values for the different growth assumptions. Combinations of the first two tests are also possible. A third test is one in which a control strategy under consideration is simulated with an alternative grid resolution or with different (reasonable) meteorological assumptions.

Other approaches for estimating uncertainty have been described in the literature (Gao, et al., (1996) and Yang, et al., (1995)). Many of these approaches also assess sensitivity of model predictions to uncertainties in input variables. For outcomes to be most relevant to the way we recommend models be applied in attainment demonstrations, it is preferable that such procedures focus on sensitivity of estimated relative reduction factors (RRF) and resulting projected design values to the variations in inputs or model formulations.

Once a range in projected design values is obtained using tests like those previously described, a qualitative assessment can be made of how likely it is that a strategy will lead to attainment of the NAAQS. For example, in applications related to the 24-hour NAAQS, if most of the results lead to projected design values $\leq 65\ \mu g/m^3$, this supports a conclusion that if the strategy is implemented, the NAAQS will be attained. Choice of tests and interpretation of the outputs should be agreed upon beforehand in concert with the appropriate U.S. EPA Regional Office.

Compare monitored design values for the current period used in the test with those measured in other periods. In Section 3.3, we suggested choosing the highest monitor-specific design values from a limited sample of recent 3-year periods. This is intended to provide some protection against basing the modeled attainment test on a relatively mild period. However, using such a limited number of periods to choose from opens the possibility that we have inadvertantly chosen an extreme period upon which to base the test. The objective of this analysis is to assess whether current design values used in the modeled attainment test are atypically high or low due to natural or meteorological conditions. If the current design values are lower (or much higher) than normal, the tests would yield overly optimistic (pessimistic)

results.

An analysis of current design values is complicated by trends in emissions as well as by a limited period of observations. For example, one would expect design values measured several years previously to be higher than current values if there has been an ongoing program to reduce $PM_{2.5}$ or precursor emissions. This does not necessarily imply that the current design value is atypically low. The problem of emissions trends can be addressed by examining statistical relationships between meteorological conditions and daily $PM_{2.5}$ concentrations to see whether conditions corresponding to high $PM_{2.5}$ occurred more or less frequently than usual during the current period. Approaches which are analogous to that described by Cox and Chu, (1996) or by Deuel and Douglas, (1998) might be tried to see how the current 3-year period ranks with other 3-year periods in terms of its frequency of days with meteorological conditions which appear to correspond with high observed $PM_{2.5}$.

Assess appropriateness of default compositional assumptions for major components of $PM_{2.5}$. This analysis examines available speciated PM observations to note proportions of elements or compounds. These proportions are compared with default composition assumptions for major components of $PM_{2.5}$ noted in Section 3.3 to check for consistency. If large inconsistencies occur, alternative assumptions regarding measured mass associated with one or more components may be proposed. This, in turn, could revise the estimated current composition of $PM_{2.5}$ used in the modeled attainment test. To illustrate, suppose measured samples in a nonattainment area suggest that measured organic molecules typically consist of 60% carbon, rather than the 70% we assume as a default value (see Table 3.4). This would be consistent with using a multiplication factor of "1.7" rather than "1.4" for estimating mass associated with organic carbon (OC). A State could use this information to recalculate the mix of the major components of currently measured $PM_{2.5}$. The resulting changes could lead to refined estimates regarding attainment using the modeled attainment test.

<u>Recommendations</u>. Optional analyses may be considered in addition to the 3 recommended analyses identified in Section 5.1. To use an optional analysis in a weight of evidence determination, a State should (1) explain the rationale for the analysis, (2) identify the data base underlying the analysis, (3) describe the methodology to be used in applying the analysis, and (4) identify outcomes which would be consistent with a hypothesis that a proposed control strategy will suffice to attain the NAAQS.

5.3 What Role Should Weight Of Evidence Play In Visibility-Related Analyses?

We believe a weight of evidence analysis is an appropriate option for States to use if the modeled test described in Section 4.2 does not clearly show reasonable progress. Unlike the NAAQS, reasonable progress goals address *trends* in air quality rather than some absolute level of air quality. Thus, the focus of weight of evidence analyses differs from those performed in NAAQS attainment demonstrations. In this subsection, we note some potential uses for

additional air quality modeling, observed ambient and emissions trends and observational models. Trend analyses and observational models have already been discussed in detail in Section 5.2. Therefore, in Section 5.3 we only identify some potential uses for these analyses. We then identify several refinements to our recommended modeled test for reasonable progress if a State or regional planning organization believes these are warranted for a weight of evidence determination. We conclude by noting some potential concerns about the ability of models to address days with very good visibility and identify some analyses which a State might wish to consider in addition to or instead of the modeled test for reasonable progress for "best" visibility days.

Additional air quality modeling. Sensitivity tests could be performed to see if conclusions about trends in "worst" and "best" visibility are robust. One example of such an analysis is applying a model with and without a more finely resolved nested grid near one or more Class I areas. A purpose of this would be to see whether conclusions are affected by the degree of detail in which nearby sources are considered A second example of an analysis would be to consider alternative current emissions and/or differing growth rate assumptions.

One or more additional models could be used to develop relative reduction factors. These RRF's could then be used in the test described in Section 4.0 to see if consistent conclusions are reached regarding trends in "worst" and "best" visibility days. used in the moded test for reasonable progress.

Further, visibility extinction can be estimating using a procedure which differs from the approach of "reconstructing" light extinction from mass concentrations of components of particulate matter described in Section 4.0. For example, models like the Community Multiscale Air Quality (CMAQ) model can also utilize information about size distribution of fine particles and associated light extinction efficiency factors to estimate light extinction in an independent way (U.S. EPA, 1999c). Visibility trends which are calculated in different ways could be compared to see if they corroborate each other.

If trends in visibility for "worst" and/or "best" days are similar using sensitivity tests, alternative models and/or alternative modeling approaches, this supports conclusions reached in the modeled test for reasonable progress.

Review of trends. This includes some qualitative comparisons between past trends in reconstructed visibility and estimated changes in emissions (e.g., early '90's to early '00's). This information could be used to confirm that more control measures on previously reduced emissions of a component or its precursors is likely to be useful. It may also be used to see whether certain components are becoming increasingly important sources of light extinction.

Observational models. Observational models are potentially useful for flagging potential importance of local sources, if any, in influencing measurements made in a Class I area. This could lead to a refined treatment of a local source, either through a more finely resolved

nested grid model application or plume-in-grid modeling. Use of nested models and plume-in-grid algorithms are discussed in greater detail in Section 13.0. Observational models may also be useful for identifying the types of meteorological conditions most often corresponding to observed "worst" and "best" visibility in various Class I areas. This, in turn, may enable States to draw inferences about the orientation of areas containing sources most likely to influence visibility in a Class I area on days with "poor" and "good" visibility.

Refinements to the recommended modeled test for reasonable progress. If a strategy for meeting reasonable progress goals appears generally successful, but the modeled test for reasonable progress is not passed in a limited number of Class I areas, States may consider refining the recommended test for reasonable progress in some manner. Refinements are best made if they are based on local observations/analyses which suggest that some of the underlying assumptions in the recommended test may not be applicable. We list some potential refinements which could be considered. The list is intended to illustrate types of additional analyses which could be performed. The list is not comprehensive and, in general, the effort needed to perform the listed analyses would be substantial.

- Rerank future estimated light extinction (i.e., b_{ext} values) for all days with current measurements and recompute mean future "best" and "worst" visibility (i.e., do not assume that the identity of baseline "best" and "worst" days remains the same).

- Use day specific estimates for relative humidity and relative humidity adjustment factors (*f(rh)*) to compute light extinction attributable to hygroscopic compounds.

- Perform area specific analyses relating monitored visibility extinction to locally applicable composition of OC, IP and SO_4. Use results to develop an area-specific version of Equation (4.1).

- Background levels themselves are not constant from day to day. Available speciated data and other information may be reviewed to see whether the outcome of the test is being influenced by including several days with extraordinary events (e.g., a nearby major forest fire lasting a number of days). If convincing arguments can be made that the event is a "natural" one, excluding these days from the calculations should be discussed with the appropriate U.S. EPA regional office.

Concerns about modeling days with "best" visibility. In some parts of the United States, concentrations of the components of particulate matter used in visibility calculations may be within a $\mu g/m^3$ or two of background levels on days with "best" visibility. Measurements and model estimates may be subject to more relative uncertainty (i.e., lower signal to noise ratio) on days where observed concentrations of particulate matter are very low (and light extinction is also low). Utililty of weight of evidence determinations is heightened in such cases. If a State has reason to believe that an atmospheric simulation model's ability to estimate concentrations of components of particulate matter is limited on such days, performance tests described in Section

16.0 should be applied to the extent feasible for the specific Class I area in question. Next, a State should see whether a model's inability to accurately predict one or more individual components of particulate matter has a substantial effect on the extinction coefficient calculated with Equation (4.1). If it does, and diagnostic tests (also described in Section 16.0) are unable to resolve a performance problem, a State may address the goal for "best" visibility days in the particular Class I area(s) without using results from a model. We anticipate that observational models may be helpful in such cases. They may be used to characterize meteorological conditions under which "best" visibility occurs and may be used to identify zones where increases in emissions may need to be limited in order to meet a goal of not degrading visibility on "best" visibility days.

Recommendations. States may use additional analyses in a weight of evidence determination to corroborate or refine a strategy for which the modeled test for reasonable progress has been applied. The decision of whether to perform a weight of evidence analysis in a reasonable progress demonstration is best made for a limited number of individual Class I areas on a case by case basis. If the ability of air quality models to reproduce very low concentrations of particulate matter on "best" visibility days is shown to be highly uncertain for a particular Class I area, a State may consider a reasonable progress demonstration for that area in which an atmospheric simulation model is not used to determine if the goal for "best" visibility days will be met.

(This page is left blank intentionally)

6.0 What Should Be My Data Gathering Priorities?

A key purpose of this document is to identify and prioritize data collection activities which may increase the credibility of modeled demonstrations of attainment or reasonable progress. We recognize that a State's data gathering priorities depend on what data are already available. Thus, data gathering priorities need to be established on a case by case basis by those implementing the modeling/analysis protocol for the attainment or reasonable progress demonstration. In this section, we identify key data needs implied by the recommended modeled tests for attainment or reasonable progress. We also identify needs for obtaining reliable model results (prior to applying the tests) and for corroboratory analyses used in weight of evidence determinations. Section 6.1 identifies data and analytical capabilities which are important for these purposes. We recognize that it may not be feasible to collect all the data desirable during the timeframes required for developing a revision to a State implementation plan (SIP). Thus, in Section 6.2 we identify factors to consider in performing "mid-course reviews" of approved SIP's.

6.1 What Data And Other Needs Are Implied By The Modeled Demonstrations For $PM_{2.5}$ And Regional Haze?

There are three major activities included in demonstrations of attainment or reasonable progress for which a good data base is needed:

- generating and evaluating model results which are subsequently used in the modeled tests for attainment or reasonable progress (described in Part II of this guidance);

- applying modeled tests for attainment or reasonable progress (described in Sections 3.0 and 4.0);

- performing weight of evidence determinations (described in Section 5.0).

We list key needs for each of these three major activities. Highest priority needs are listed first. Of course, priorities for collecting data in any given geographical area are affected by the data and capabilities which are already available. Further, those implementing the modeling protocol may need to make case by case decisions regarding tradeoffs among the different needs to meet budgetary constraints. Planning documents for a study in California's San Joaquin Valley provide an example of how to balance theoretical needs with existing data bases and budgetary constraints when developing priorities for collecting data (California Air Resources Board, 1999).

6.1.1 Key Needs For Generating Credible Model Results Underlying Attainment And Reasonable Progress Tests

1. Develop accurate emission factors, seasonal and diurnal activity levels for major sources

and source categories of $PM_{2.5}$, SO_2, NH_3, NOx, VOC and CO

These are the most important inputs to an air quality model, because they are the ones that get altered in a control strategy. The effects of altering emissions are suspect if the emissions are poorly characterized. Priorities for improving emissions estimates for different source categories depend on the amount of uncertainty associated with current estimates, the potential for large variability in a source category's emission factors or activity levels among geographical areas and the relative importance of related monitored components of $PM_{2.5}$.

2. Make speciated measurements of ambient $PM_{2.5}$, and refine speciated source profiles so that they are applicable locally

Speciated ambient measurements are needed to help evaluate performance of an air quality model. Clearly, performance is likely to be enhanced for primary particulate matter (IP, primary OC, EC) if speciated emission profiles for $PM_{2.5}$ are applicable to the area and time of the simulation. As we note in Section 15.0, speciated ambient measurements are potentially useful as a means for quality assuring emission estimates to assess their applicability to the local situation. Speciated ambient measurements and speciated emission profiles are also keys to developing an accurate conceptual description of an area's $PM_{2.5}$ or visibilitiy problem. As we discuss in Section 9.0, the conceptual description of a problem is instrumental in choosing a modeling approach, making decisions about detail needed for various inputs and choosing strategies to investigate with a model.

Speciated measurements should be collocated with $PM_{2.5}$ mass measurements. Highest priority should be given to making measurements at sites violating or likely to violate the NAAQS for $PM_{2.5}$ and in Class I areas. We want to be sure that the model performs well here. In order to improve chances of developing an accurate conceptual picture of an area's nonattainment problem, it is also important to make some ambient speciated measurements in rural areas to characterize regional levels and composition of particulate matter. Availability of some rural mass and speciated data also makes it possible to better evaluate whether a model can accurately predict buildup of particulate matter in urban areas.

3. Make upper air measurements

These measurements should include wind velocity, temperature, dew point and pressure. Ideally, they should be made continuously. If this is not feasible, we recommend that measurements be made 3-4 times per day—shortly after sunrise, midday, at twilight and during the night. Similar surface measurements should be made so that a complete vertical profile can be developed for these variables. As we note in Section 14.0, these measurements are used to ensure that meteorological input generated for air quality models by dynamic meteorological models is realistic. Meteorological models, in turn, develop the wind fields which define source-receptor orientations on any given day. As described in Section 12.0, meteorological factors may also help in choosing appropriate, representative periods to model.

If possible, short duration special studies should be performed, in which speciated air quality measurements are made aloft. These measurements should include speciated $PM_{2.5}$ (i.e., to derive SO_4, NO_3, OC, EC, IP), other oxidized forms of nitrogen, SO_2 and ozone). These measurements are useful for evaluating model performance, as well as for developing a conceptual description of an area's problem.

4. Make short term measurements of $PM_{2.5}$ and its species

"Short term", in this context, refers to the sampling time. Measurements of $PM_{2.5}$ and its components should be made using sampling times no longer than 6 hours duration. Measurements of this nature are very important to meaningfully evaluate model performance. Comparing predictions and observations over 24-hour periods could result in failure to detect poor performance in reproducing diurnal patterns of observed particulate matter. Poor performance of this nature could result in source/receptor orientation being mischaracterized on a particular day. Collection of 4 or more samples per day obviously greatly increases needed resources and data management problems. It suffices to make measurements of this nature only for limited duration field studies. Such short term studies should also include measurements of air quality aloft. For purposes of evaluating model performance predicting relatively non-volatile components of PM (e.g., SO_4 , EC, IP) continuous surface measurements should be considered. As we describe in Section 16.0, efforts to evaluate model performance are best directed to these periods with intensive data bases.

5. Develop accurate estimates for natural sources of VOC, primary OC and ammonia

Relative importance of natural emissions as a source of $PM_{2.5}$ (especially of secondary or primary OC) is not yet clear. We anticipate that the U.S. EPA will continue its efforts to enhance/improve the BEIS emissions model (U.S. EPA, 1997a). However, States could focus on items like distribution of land use and environmental factors (e.g., temperature) so that the national model furnishes estimates which are as reliable as feasible.

6. Procure hardware, data storage capabilities and develop expertise to apply a regional photochemical grid model with aerosol capabilities

As we discuss in Section 1.0, secondary particulate matter likely constitutes a major portion of measured $PM_{2.5}$, and is a major source of light extinction. In order to address reasonable progress, it will be necessary to use such models. In most instances, regional scale photochemical grid models will also be needed to address $PM_{2.5}$ nonattainment problems. While the tasks of running and interpreting results from such models can be performed externally, it is important for State personnel to be sufficiently familiar with related issues to direct and critically review these studies. If the work is to be done inhouse or with inhouse facilities, then necessary hardware and data storage facilities are required.

7. Make PM_{10} as well as $PM_{2.5}$ mass measurements in Class I areas

This needs to be done to ensure that the model performs adequately in predicting observed coarse particulate matter (CM). Changes in predicted concentrations of coarse particulate matter affects predicted changes in light extinction (see Equation (4.1)).

8. Measure surface meteorological conditions routinely

Measurements should include, wind velocity, relative humidity, temperature and pressure. They should prove useful in helping to identify distinctive types of episodes to model. They may also prove useful in developing a conceptual description of an area's $PM_{2.5}$ problem and provide needed inputs for hot spot analyses (see Section 3.4).

9. Analyze speciated data to ascertain the importance of water as a component of measured mass of particulate matter and to distinguish between primary and secondary parts of organic carbon (OC)

Analyzing for presence of water may be needed to explain any apparent underprediction of measured $PM_{2.5}$ by an air quality model. Further, if water associated with measured sulfates and nitrates can be estimated, a State could reduce the fraction of unattributed mass (U) in the modeled attainment tests and instead associate this mass with sulfates and/or nitrates. Distinguishing between secondary and primary OC should help develop a better conceptual description of an area's $PM_{2.5}$ or visibility problem. It should also facilitate a more complete assessment of model performance.

6.1.2 Key Needs For Performing Modeled Tests For Attainment Or Reasonable Progress

1. Develop accurate emission factors and seasonal/diurnal activity levels for major anthropogenic sources and source categories of $PM_{2.5}$, SO_2, NH_3, VOC and NOx.

These estimates are needed to determine accurate relative reduction factors (RRF) and to choose strategies likely to pass modeled tests for attainment or reasonable progress on an informed basis.

2. Make ambient speciated measurements of $PM_{2.5}$

These measurements are needed to establish current concentrations of the major components of $PM_{2.5}$. The attainment and reasonable progress tests estimate future concentrations of major components by multiplying current component concentrations times component specific relative reduction factors. Because we need to account for all measured mass of $PM_{2.5}$ with the major components (i.e., this is done through use of a component "U", which accounts for differences in the measured mass and the sum of mass associated with the remaining components), it is best to collocate speciated measurements with FRM or equivalent monitors.

Priorities for making these measurements follow:

- make measurements at monitoring sites violating or likely to violate NAAQS for $PM_{2.5}$;

- make measurements in rural areas so that as many nonattainment areas as possible have at least one relatively nearby background site with speciated measurements.

3. Measure PM_{10} and $PM_{2.5}$ in Class I areas

These measurements are needed to derive current estimates for coarse particulate matter (CM). CM is used in Equation (4.1) to relate measured components of particulate matter to the extinction coefficient used to characterize light extinction.

4. Develop post-processing software to facilitate application of model results and monitored data in the modeled tests for attainment and reasonable progress

Software is needed to convert speciated measurements into mass associated with each major component of PM. Information presented in Tables 3.4 and 4.1 should be used as the basis for such software. In addition, it is clear from Examples 3.1, 3.2, 3.3 and 4.2 that the modeled tests we recommend require substantial manipulation of model-generated and monitored data. Further, likely sample sizes will be much larger than those we have used for illustrative purposes in the examples. Automated procedures will be needed to make the required data analysis tractable. We plan to develop and provide States with such software.

5. Develop speciated emissions profiles for major source categories of particulate matter

These estimates are more important for generating model results and for applying weight of evidence analyses. However, they are also useful in applying modeled tests for attainment or reasonable progress. The information developed in the profiles can be used in concert with monitored data to help focus control efforts on sources which appear to be contributing to monitored exceedances of the NAAQS or to light extinction. This could be done using observational models like those described in Section 4.1.3.

6. Develop accurate estimates for biogenic/geogenic sources of VOC, primary OC and ammonia

These estimates are needed to improve accuracy of relative reduction factors associated with strategies to reduce concentrations of $PM_{2.5}$ and regional haze. The more important the contribution that emissions of biogenic/geogenic sources make to a component of $PM_{2.5}$, the less responsive the RRF value for that component will be to control measures. We anticipate that the U.S. EPA will continue its efforts to enhance/improve the BEIS emissions model (U.S. EPA, 1997a). However, States could focus on items like distribution of land use and environmental

factors (e.g., temperature) so that the national model furnishes estimates which are as reliable as feasible.

7. Make ambient $PM_{2.5}$ and wind measurements at locations which are likely hot spots.

Measurements of $PM_{2.5}$ are intended to confirm that an identified hot spot is a potential concern for meeting the NAAQS. Wind measurements are useful to confirm that the monitor/source orientation is consistent with high $PM_{2.5}$ measurements. If preliminary measurements of $PM_{2.5}$ mass suggests a problem is likely, States may wish to augment the mass and wind measurements with speciated measurements. Availability of speciated measurements will enable the hot spot to be considered more readily using the recommended modeled attainment test in a mid-course review or subsequently.

8. Deploy routine instrumentation to estimate relative humidity at speciated monitoring sites in Class I areas

As we note in Sections 4.1 and 4.2, the monitored and modeled tests for reasonable progress utilize "climatological" estimates for humidity adjustment factors (*f(rh)*). However, not every Class I area has relative humidity measurements which can be used to derive the climatological values for *f(rh)*. In these cases, relative humidity measurements made elsewhere have been used after corrections for differences in altitude are made. As Examples 4.1 and 4.2 make clear, effectiveness of various control strategies to reduce regional haze can be significantly affected by assumptions about relative humidity. The purpose of obtaining this information is to develop more accurate climatological estimates of relative humidity which could be considered subsequently in a mid-course review or in designing the next iteration of reasonable progress.

9. Analyze FRM data to determine water content associated with hygroscopic components of particulate matter.

The purpose of this analysis is to reduce the amount of measured $PM_{2.5}$ mass which is assigned to component "U" (e.g., unattributed mass). Recall from Section 3.1 that mass assigned to component "U" remains unaffected by a simulated control strategy. This analysis would provide a rationale for assigning some currently unattributed mass of measured $PM_{2.5}$ to components which can be reduced by a control strategy (e.g., SO_4 and NO_3).

6.1.3 Key Needs For Performing Weight Of Evidence Determinations

Our discussion is limited to the three types of analyses discussed in Section 5.1 (i.e., other air quality model outputs, trend analysis and use of observational models).

1. Monitor speciated concentrations of $PM_{2.5}$

Highest priority should be given to making these measurments at $PM_{2.5}$ monitoring sites

violating the NAAQS or likely to violate the NAAQS. A second priority is measurements at locations which are often "upwind" from major source areas. Ideally, speciated and total mass measurements should occur on the same days and at the same frequency. These measurements are needed to use multivariate statistical receptor models. They are also one required input for a chemical mass balance receptor model. Further, trend analysis can be more readily interpreted if speciated ambient air quality data are available. Finally, these data may be potentially useful for implementing indicator species approaches.

2. Develop speciated emission profiles for major source categories of $PM_{2.5}$

These are prerequisites for using the chemical mass balance receptor model. They also serve as a means for relating the outcome of multivariate statistical receptor models to source categories likely affecting monitored concentrations of $PM_{2.5}$.

3. Make an accurate assessment of the spatial configuration of major source categories of $PM_{2.5}$, SO_2, NH_3, NOx and VOC

This information is needed to be able to use hybrid receptor models. Hybrid models combine results of statistically based receptor models (e.g., CMB, multivariate statistical approaches) with meteorological information. They are potentially useful for identifying potentially important sources of measured secondary as well as primary components of particulate matter.

4. Make upper air meteorological measurements

Continuous or at least three or 4 such measurements should be made daily. Ideally, the measurements should be made at at least one location per urban area where speciated $PM_{2.5}$ measurements are made. Measurements should be used to characterize winds aloft, as well as temperature and moisture so that vertical profiles describing wind strength and direction, turbulence, mixing heights and conditions suitable for cloud formation can be identified. Pressure measurements need to be included so that the other measurements can be related to altitude above ground level. Wind information can be used to drive trajectory models or, alternatively, to guide predictions made with dynamic meteorological models used to provide wind data for trajectory models[20]. Trajectory models can be used as described in Section 5.1.3 in concert with receptor models to help draw inferences about potential sources of observed secondary particulate matter. Information regarding mixing and/or cloud cover may be useful in explaining variability in observed concentrations of components of particulate matter which accompanies a seemingly similar set of trajectories.

5. Make surface meteorological measurements at sites measuring speciated concentrations of particulate matter

---

[20]See Section 14.0 for a more complete discussion of dynamic meteorological models.

Wind, temperature, moisture and pressure measurements should be included in the ground level measurements. Surface measurements such as these are potentially useful for identifying distinctive clusters of days. Each cluster can then be examined to see if measured speciated profiles have attributes in common or if any particular cluster corresponds to high concentrations of $PM_{2.5}$. Surface meteorological data are also potentially useful in helping to normalize observed trends in $PM_{2.5}$ or its components for meteorological differences.

6. Develop software to facilitate post-processing of air quality model outputs

In Section 5.1.1, we identified several air quality modeling outputs (e.g., number of grid cell-days greater than some threshold value such as 15.0 $\mu g/m^3$) which could prove useful in making "weight of evidence" arguments. Calculating these outputs generally is not feasible unless computer software is developed for doing so.

6.1.4 Overall Priorities

All of the previously mentioned data and capabilities are important. However, the list is overwhelming. In this subsection, we attempt to identify the highest priority efforts. In establishing these priorities, we have considered the following factors:

- presence as a priority for all three of the major activities (generating model results, applying the tests and performing weight of evidence analyses);

- fundamental nature of the need—do subsequent activities depend on it?

- is assessing attainment or reasonable progress practical without this need being met?

Using the preceding criteria, we judge the following activities to have the highest priority.

1 (tie). Develop accurate emission factors and seasonal/diurnal activity levels for major anthropogenic sources and source categories of $PM_{2.5}$, SO_2, NH_3, VOC, NOx and CO, and

1 (tie). Make speciated ambient measurements of $PM_{2.5}$ at mass monitoring sites likely to exceed the NAAQS and at a limited number of rural locations.

3. Develop locally applicable source speciation profiles for primary $PM_{2.5}$ and VOC.

4. Make continuous or 4 or more upper air meteorological measurements/day and ensure these are accompanied by surface meteorological measurements at the same site.

5. Develop software to post process modeling and monitored data so that the modeled tests for attainment and reasonable progress can be applied.

6. Arrange a limited duration study in which $PM_{2.5}$ and its components are measured continuously or sampled over periods not exceeding 6 hours at selected locations with routine 24-hour sampling.

<u>Recommendations</u>. States should enhance available data bases and capabilities so that the following activities can be performed credibly: generating model results for use in the tests, applying the modeled tests for attainment and reasonable progress, and performing weight of evidence analyses. A number of important priorities are identified in this Section. Of these, we believe the following have the highest priority.

1. (Tie) Improve emission factors and activity estimates for major source categories of $PM_{2.5}$, SO_2, NH_3, VOC, NOx and CO, and

1. (Tie) Make ambient speciated measurements of $PM_{2.5}$ routinely at sites with mass measurements which are likely to violate one or both NAAQS and in Class I areas.

3. Develop locally applicable source speciation profiles.

4. Make continuous or at least several (e.g., 4) upper air meteorological measurements per day and corresponding surface measurements.

5. Develop software so that the recommended modeled tests for attainment and reasonable progress can be readily applied.

6. Perform a limited duration study in which $PM_{2.5}$ and its components are sampled continuously or discretely for periods not exceeding 6 hours.

6.2 Why Is It Desirable To Plan For A Subsequent Review?

States should anticipate the need for a subsequent review of their SIP after its initial approval by the U.S. EPA. As we note in Section 6.1, there are many data-related needs which, if met, would enhance the credibility of a strategy to meet air quality goals for $PM_{2.5}$ and regional haze. Resources are limited and some of these needs will likely have to be met over a sustained period of time. Second, our understanding of the science underlying formation, transport and deposition of $PM_{2.5}$ is evolving. Increased understanding of these processes will be reflected in periodic updates of models and other analytical tools used to support the demonstrations of attainment or reasonable progress. Third, formation and transport of secondary particulate matter is closely related to processes which are important in the formation and transport of ozone. Thus, it makes sense for programs designed to reduce $PM_{2.5}$ and improve visibility to be cognizant of programs to reduce ozone and vice versa. This can best be done after there has been an opportunity to review changes in air quality resulting from implementation of plans to reduce $PM_{2.5}$, regional haze and ozone.

We recommend that States consider the following in planning for subsequent reviews. First, continue to improve data bases. Second, retain the means to perform modeling/analysis. Third, make emission and air quality projections to some intermediate year which is about the time a subsequent review is anticipated. These projections may subseqently prove useful in assessing whether a "mid-course correction" to an emission reduction strategy is needed. Finally, retain modeling input and output files used to simulate base emissions, emission projections for the previously discussed intermediate year and projected emissions at the time of required attainment or at the end of the first iteration of the plan to make reasonable progress reducing regional haze. Only the files reflecting the strategy approved in the approved SIP revision need be retained.

> **<u>Recommendations</u>. States should maintain modeling and analysis capabilities after approval of their initial SIP revisions addressing NAAQS for $PM_{2.5}$ and regional haze. Efforts should continue to enhance data bases which will underlie modeling performed to support subsequent reviews and adjustments to the initial SIP revisions. Emissions and air quality projections should be made for an intermediate year to compare with monitored data during a "mid-course review" of an approved plan. To facilitate such a review, States should archive modeled input and output files associated with current emissions and with projections reflecting the control/growth assumptions in the approved SIP revision.**

7.0 What Documentation Do I Need To Support My Attainment Demonstration?

Recommendations regarding documentation for demonstrations showing that a SIP will meet goals for $PM_{2.5}$ or regional haze are similar to those for the 8-hour ozone NAAQS found in U.S. EPA (1999b). We have added one requirement for demonstrations related to $PM_{2.5}$ or regional haze. That is, the documentation should contain a conceptual description of an area's $PM_{2.5}$ or regional haze problem. This description includes a summary of the observational data bases examined and the rationale that was used for focusing the modeling efforts in the way that they were. We believe this requirement is useful, because of particulate matter's nature as a mixture. Attempting to reduce concentrations of a mixture of pollutants provides more potential directions for modeling efforts and strategy development than would otherwise be the case. This increases the possiblity of focusing on tangential issues and strategies. A conceptual description provides some assurance that the modeling efforts and strategy are likely addressing the most significant attributes of an area's $PM_{2.5}$ or regional haze problem.

> **<u>Recommendations</u>. States should address the 11 subject areas shown in Table 7.1 in the documentation accompanying an attainment or reasonable progress demonstration. The documentation should contain a summary section which addresses issues shown in the table. More detailed information should be included in appendices, as necessary.**

Table 7.1. Recommended Documentation For Demonstrating Attainment Of Air Quality Goals For $PM_{2.5}$ Or Regional Haze

Subject Area	Purpose of Documentation	Issues Included
Conceptual Description of $PM_{2.5}$ or Regional Haze Problem	**To provide a short narrative of observational evidence which led to the selected modeling approach and strategies which were investigated.**	**Measurements used** **Analyses performed** **Resulting rationale used to support the modeling approach and strategies investigated**
Modeling/Analysis Protocol	**Communicate scope of the analysis and document stakeholder involvement**	**Names of stakeholders participating in preparing and implementing the protocol** **Types of analyses performed** **Steps followed in each type of analyses** **Days and domain considered**
Emissions Preparations and Results	**Assurance of valid, consistent emissions data base and that appropriate procedures are used to derive emission estimates needed for air quality modeling**	**Data base used and quality assurance methods applied** **Data processing used to convert data base to model-compatible inputs** **Deviations from existing guidance and underlying rationale** **Emissions models used and justification for choice of models** **$PM_{2.5}$, SO_2, NH_3, VOC, NOx, CO emissions by State/county for major source categories**

Table 7.1. Recommended Documentation For Demonstrating Attainment Of Air Quality Goals For $PM_{2.5}$ Or Regional Haze (continued)

Subject Area	Purpose of Documentation	Issues Included
Air Quality/Meteorology Preparations and Results	**Assurance that representative air quality and meteorological inputs are used in analyses**	**Extent of data base and procedures used to derive & quality assure inputs for analyses used in the weight of evidence determination** **Departures from guidance and their underlying rationale** **Meteorological model(s) used** **Performance of meteorological model if used to generate meteorological inputs to the air quality model**
Performance Evaluation for Air Quality Model (and Other Analyses)	**Show decision makers and the public how well the model (or other analyses) reproduced observations or otherwise performed on the days selected for analysis**	**Summary of observational data base available for comparison** **Identification of performance tests used and their results for components of PM, mass of $PM_{2.5}$ and key gas phase species** **Ability to reproduce observed temporal and spatial patterns for components of PM, mass of $PM_{2.5}$ and key gas phase species** **Overall assessment of what the performance evaluation implies**
Diagnostic Tests	**Ensure rationale used to adjust model inputs or to discount certain results is physically justified and the remaining results make sense.**	**Results from application prior to adjustments** **Consistency with scientific understanding and expectations** **Tests performed, changes made and accompanying justification** **Short summary of final predictions.**

Table 7.1. Recommended Documentation For Demonstrating Attainment Of Air Quality Goals For $PM_{2.5}$ Or Regional Haze (continued)

Subject Area	Purpose of Documentation	Issues Included
Description of the Strategy Demonstrating Attainment (or Reasonable Progress)	**Provide the EPA and the public an overview of the plan selected in the attainment or reasonable progress demonstration.**	**Qualitative description of the selected strategy** **Reductions in $PM_{2.5}$, SO_2, NH_3, CO, VOC, and/or NOx emissions from each major source category for each State/county from current (identify) emission levels** **Clean Air Act mandated reductions and other reductions** **Show relative reduction factors estimated for each major component of $PM_{2.5}$ and (for Regional haze applications) CM** **Show predicted site-specific future $PM_{2.5}$ design values for the selected control scenario and identify any location for which a hot spot analysis was needed together with results of the hot spot analysis** **Identification of authority for implementing emission reductions in the selected strategy** **Evidence that emissions will remain at or below projected levels throughout the 3-year period used to determine future attainment for $PM_{2.5}$-related applications, and the 5-year period used to determine if reasonable progress has occurred for visibility-related applications**

Table 7.1. Recommended Documentation For Demonstrating Attainment Of Air Quality Goals For $PM_{2.5}$ Or Regional Haze (continued)

Subject Area	Purpose of Documentation	Issues Included
Data Access	**Enable the EPA or other interested parties to replicate model performance and simulation results for the selected strategy, as well as results obtained with other analyses.**	**Assurance that data files are archived and that provision has been made to maintain them** **Technical procedures for accessing input and output files** **Identify computer on which files were generated and can be read, as well as software necessary to process model outputs** **Identification of contact person, means for downloading files and administrative procedures which need to be satisfied to access the files**
Weight of Evidence Determination ($PM_{2.5}$ NAAQS-related applications)	**Assure the EPA and the public that the strategy is likely to produce attainment of the NAAQS within the required time.**	**Description of the modeled attainment test and observational data base used** **Identification of air quality model(s) used** **Identification of other analyses performed** **Outcome of each analysis, including the modeled attainment test** **Assessment of the credibility associated with each type of analysis in this application** **Narrative describing process used to conclude the overall weight of available evidence supports a hypothesis that the selected strategy is adequate to attain the NAAQS**

Table 7.1. Recommended Documentation For Demonstrating Attainment Of Air Quality Goals For $PM_{2.5}$ Or Regional Haze (concluded)

Subject Area	Purpose of Documentation	Issues Included
Weight of Evidence Determinations (related to reasonable progress goals to reduce regional haze)	**Assure the EPA and the public that the strategy is likely to meet regional progress goals to reduce regional haze**	**Description of the modeled test for reasonable progress and observational data base used** **Identification of air quality model used** **Identification of other analyses performed** **Outcome of each analysis, including the modeled test for reasonable progress** **Assessment of the credibility associated with each type of analysis in this application** **Narrative describing process used to conclude the overall weight of available evidence supports a hypothesis that visibility on the days with best visibility will not deteriorate and that goals for days with poor visibility will be met**
Review Procedures Used	**Provide assurance to the EPA and the public that analyses performed in the attainment or reasonable progress demonstration reflect sound practice**	**Scope of technical review performed by those implementing the protocol** **Assurance that methods used for analysis were peer reviewed by outside experts** **Conclusions reached in the reviews and the response thereto**

Part II

How Should I Generate Modeling Results To Use In Attainment Or Reasonable Progress Tests?

(This page is left blank intentionally)

8.0 How Do I Apply Air Quality Models?–An Overview

In Part I of this guidance, we described how to estimate whether a proposed control strategy will lead to attainment of the NAAQS for $PM_{2.5}$ or reasonable progress goals to reduce regional haze. We noted that air quality models play a major role in making this determination. We assumed that modeling had been completed, and discussed how to use the information produced. We now focus on how to apply models to generate the information used in the modeled tests for attainment and reasonable progress. The procedure we recommend consists of 9 steps:

1. formulate a conceptual description of an area's nonattainment or regional haze problem;
2. develop a modeling/analysis protocol;
3. select one or more appropriate air quality models to use;
4. select appropriate days to model;
5. choose a modeling domain with an appropriate number of layers and appropriately sized grid cells;
6. generate meteorological and air quality inputs to the air quality model;
7. generate emissions inputs to the air quality model;
8. evaluate performance of the air quality model and perform diagnostic tests;
9. simulate and evaluate prospective control strategies.

In this Section (Section 8.0), we briefly describe each of these steps to better illustrate how they are interrelated. Because many of these steps require considerable effort to execute, States should take care to keep the appropriate U.S. EPA regional office(s) informed as they proceed. This will increase the likelihood of having an approvable attainment demonstration when the work is completed. Steps outlined in this Section are described in greater depth in Sections 9.0 - 17.0.

1. Formulate a conceptual description of an area's nonattainment or regional haze problem. A State needs to have an understanding of the nature of an area's problem before it can proceed with a modeled attainment or reasonable progress demonstration. For example, it would be difficult to identify appropriate stakeholders and develop a modeling protocol without insights into whether resolution of the problem requires close coordination and cooperation with other nearby States.

In the case of a nonattainment problem, the State containing the designated nonattainment area is expected to initially characterize the problem. For regional haze-related model applications, the initiative for developing a conceptual description may be assumed by a State containing a Class I area or, if its member States agree, by a regional planning organization formed to address regional haze. Initial characterization of the $PM_{2.5}$ or regional haze problem provides a starting point for addressing steps needed to generate required information by those implementing the protocol. Several examples of issues addressed in the initial description of a problem follow. Is it a regional or local problem? Are factors outside of the nonattainment area

likely to affect what needs to be done locally? Which components of $PM_{2.5}$ seem to constitute a major portion of the measured mass of particulate matter? What seems to be the relative importance of primary vs. secondary particulate matter? Are monitoring sites observing violations located in areas where meteorology is complex or where there are large emission gradients? How has observed air quality responded to past efforts to reduce emissions of PM or its precursors? Are there any ambient measurements suggesting which precursors and sources are important to further reduce secondary particulate matter? What priorities exist for gathering additional air quality, meteorological and emissions data? What information might be needed from potential stakeholders? As many of the preceding questions imply, an initial conceptual description may be based largely on a review of ambient air quality data. Sometimes, methods described in Section 5.0 (e.g., trend analysis, observational models) may be used. Other times, these types of analyses may be deferred until after a team is in place to develop and implement steps following a modeling/analysis protocol. The initial conceptual picture may be based on less resource intensive analyses of available data.

2. Develop a modeling/analysis protocol. A protocol describes how modeling will be performed to support a particular attainment or reasonable progress demonstration. Its direction and participating stakeholders are influenced by the previously developed conceptual description of the problem to be resolved. The protocol outlines methods to be used to perform the subsequent 7 steps needed to generate the modeling results and to subsequently apply the modeled attainment/reasonable progress tests and hot spot analyses, as well as other corroborating analyses in a weight of evidence determination. In addition, the protocol defines procedures which will be followed in performing the analyses that support attainment or reasonable progress demonstrations. These procedures include: a) identifying those responsible for implementing the modeling, b) identifying those who will review each step as it occurs, c) identifying procedures to be used to consider input/suggestions from those potentially affected by the outcome (i.e., "stakeholders"), and d) outlining how decisions will be made concerning technical analyses needed to complete each step in the modeling procedure. In short, the protocol defines the "game plan" and the "rules of the game".

3. Select one or more appropriate models for use. This step includes reviewing air quality data to gain insight about the nature of a nonattainment area's measured concentrations of particulate matter and its components, reviewing rules established in the *Guideline for Air Quality Models* (U.S. EPA, 2000b), and considering experience/expertise of those performing the modeling. Identifying an air quality model to be used is an early step in the process, since it may affect how emissions and meteorological information are input to the model. It could also affect size of the area modeled and choice of the horizontal/vertical resolution considered. Major components of particulate matter include primary particulate matter (i.e., inorganic particulate matter, elemental carbon and some organic carbon) and secondary particulate matter (i.e., sulfates, nitrates and some organic carbon). Because modeling requirements may differ significantly for primary and secondary particulate matter and for performing hot spot analyses, it is conceivable that a State may choose to use more than one model to support its attainment demonstration.

4. Select appropriate days to model. Like the preceding step, this step requires review of available air quality data. It also requires familiarity with the forms of the annual and 24-hour national ambient air quality standards for particulate matter and the goals for reasonable progress reducing regional haze (U.S. EPA, 1997, 1999). It is also important to thoroughly understand the modeled tests described in Sections 3.0 and 4.0. In addition, it requires a review of meteorological conditions accompanying monitored exceedances of the concentration specified in the 24-hour NAAQS (65 $\mu g/m^3$) as well as those accompanying good and poor visibility. The object of these reviews is to select periods which a) include days with observed concentrations exceeding site-specific design values (24-hour NAAQS), b) select a representative mix of days for each quarter in applications dealing with the annual NAAQS and c) select days which are representative of those corresponding with good and poor visibility.

5. Choose a modeling domain with an appropriate number of layers and appropriately sized grid cells. Appropriate domain size is influenced by the air quality goal being addressed (visibility or NAAQS for $PM_{2.5}$), whether the model is being applied to address primary or secondary particulate matter and choice of days modeled. Presence of topographical features or mesoscale meteorological features (e.g., land/sea breeze) near or in the nonattainment area of principal interest are factors to consider in choosing the number of required layers and size of individual grid cells for that portion of the modeling grid. Other factors affecting choice of grid cell size are the proximity of nearby sources and available spatial detail in the emissions data used as input to an emissions model. Finally, feasibility of managing large data bases and resources needed to estimate meteorological inputs and air quality in many grid cells are factors which cannot be ignored in choosing size of a domain and its grid cells. Using a domain with "nested" grids (i.e., fine grid cells near an area of principal interest, larger cells further away) is a potential means for reconciling needs with available resources.

6. Generate meteorological and air quality inputs to the air quality simulation model. Unlike emissions, meteorological inputs remain constant during "performance-related"[21], "current" and "future" periods simulated with the air quality model. Nevertheless care needs to be taken in specifying these, as they may affect relationships predicted between particulate matter and emissions. Modeling applications which need to consider secondary particulate matter may have to consider large geographical areas in many instances. Further, it is conceivable that meteorological conditions aloft can have an important effect on predicted concentrations of secondary particulate matter. Finally, meteorological monitoring is relatively sparse outside of cities and, especially, aloft. Thus, we recommend that for applications addressing strategies to reduce concentrations of secondary particulate matter, meteorological models ordinarily be used to generate meteorological inputs. Application of meteorological models and choice of model grid resolution in the preceding step are closely related. Meteorological conditions near the area which is the focus of a modeled demonstration may dictate the required spatial resolution. On the other hand, cost and data management difficulties increase greatly for finely resolved grids. Thus, those implementing the protocol will likely be faced with a tradeoff between

[21]See step 7 for definitions of these terms.

cost/feasibility of running air quality and meteorological models and resolution at which it might be most desirable to treat dispersion of nearby emissions. The problem can be resolved in part by considering primary and secondary components of particulate matter separately. Modeling for primary components of particulate matter may be performed over a much more limited domain, but with a greater degree of spatial resolution.

Air quality inputs consist of initial conditions and boundary conditions to the model domain. Importance of initial conditions should be diminished by beginning a simulation at a time prior to the period which is of interest. Nature of boundary conditions is an important factor in deciding how large to make the size of the area modeled. The most satisfactory way to generate future boundary conditions is through use of a regional air quality model. Therefore, those implementing the protocol will once again be faced with a tradeoff between cost/feasibility of data base management vs. a desire to limit the importance of an arbitrarily specified input to the modeling exercise.

7. Generate emissions inputs to the air quality simulation model. Emissions are the central focus in a modeled attainment or reasonable progress demonstration. That is, they are the only input to an air quality model which those implementing the protocol can control. Hence, they are the major input which gets changed between the present and future. Emissions which are input to a complex, multi-source air quality model are generated using an emissions model. Applying such a model is as complicated as the air quality model itself, and demands at least as much attention. In current emissions models, emissions from some of the major source categories of precursors for secondary particulate matter are affected by meteorological conditions. This requires an interface between meteorological inputs and emissions. Emissions which are input to an air quality model are also affected by the latter's horizontal/vertical resolution and, of course, the size of the area modeled. In short, treatment of emissions is a central and complex one which, itself, involves several steps. These include deriving emission inventories, quality assuring results, applying results in an emission model(s), and (again) quality assuring results.

Emissions inputs may be needed for as many as 3 periods: (1) a "performance-related period" corresponding to the time in which air quality observations used to evaluate an air quality model's performance are made, (2) a "current period", corresponding to that represented by the current monitored design value (or base period from which reasonable progress is measured), and (3) a future period, corresponding to a time two years prior to the required attainment date or 2 years prior to when the next assessment of reasonable progress is required. If a model's performance evaluation is based on comparisons between predictions and observations on a small number of days (e.g., with an intensive data base), performance-related emissions should reflect day-specific conditions to the extent possible. For performance measures which compare observed and predicted concentrations averaged over numerous days, emissions needs are similar to those for the "current period" considered in the modeled tests for attainment or reasonable progress. "Current period" emissions may ordinarily reflect quarterly average estimates. However, if seasonal differences in a location's emissions do not closely correspond to calendar quarters, States should use monthly average emission estimates to estimate "current emissions".

8. Evaluate performance of the air quality model and perform diagnostic tests. To an important extent, credibility of a modeled attainment or reasonable progress test's results and other modeled outputs is affected by how well the model replicates observed air quality. Evaluating model performance and conducting diagnostic tests depend on prior definition of the modeling exercise and specification of model inputs. Hence, this is generally the last step prior to using the model identify strategies which are likely to meet the air quality goal(s).

At first, it may seem that the obvious way to evaluate model performance is to simply compare predicted and observed concentrations of $PM_{2.5}$ or predicted/observed values of extinction coefficients. It is not so simple, however. Because particulate matter is a mixture, it is possible to get seemingly good agreement between observed and predicted values of $PM_{2.5}$ but poor performance predicting individual components. Badly mischaracterizing the mix of secondary components of $PM_{2.5}$ has the potential for deriving erroneous component-specific relative reduction factors. Further, models needed to simulate formation, transport and deposition of particulate matter often have many inputs. It is possible to get similar predicted concentrations of $PM_{2.5}$ or an individual component with different combinations of these inputs. There is no guarantee that $PM_{2.5}$ or values of b_{ext} will respond the same way to controls with these different combinations of inputs. Further, limited availability of observed samples for periods shorter than 24 hours presents difficutlites in evaluating model performance. Thus, we place greater emphasis on additional kinds of tests. These include ability to predict observed ozone, use of precursor observations, ability to predict components of $PM_{2.5}$, use of indicator species, use of corroborative analyses with observational models and use of retrospective analyses.

Performance tests which are most closely related to how we believe models should be used in attainment/reasonable progress analyses should receive greatest emphasis. Thus, tests which compare spatially paired observations/predictions which have been averaged over numerous days are important, because they reflect how relative reduction factors should be calculated. To the extent possible, States should also include performance tests which evaluate accuracy of a model's *response* to changes in emissions.

Diagnostic tests are separate simulations which are performed to determine the sensitivity of a model's predictions to various inputs to the model. As the name implies, these tests are performed to diagnose why a model is performing poorly or to provide assurance that seemingly good performance is not just a serendipitous outcome. These tests can be performed for a variety of purposes, including prioritizing inputs needing greatest quality assurance and assessing uncertainty associated with model predictions. In performing such tests, States should remember how model results are used in the modeled attainment and reasonable progress tests recommended in Sections 3.0 and 4.0. With the exception of hot spot analyses, model results are used in a relative rather than absolute sense. In particular, the modeled tests require use of relative reduction factors (RRF), generated by models. Thus, diagnostic tests should be used to consider how RRF, as well as absolute predictions of primary particulate matter (in hot spot analyses), are affected by changes to model inputs.

9. Simulate and evaluate prospective control strategies. This step is to identify potential control strategies which appear consistent with a previously developed conceptual description. Several of these may be tested to confirm that previous implications about what might be an effective strategy to meet an air quality goal are correct. The intent of this step is to narrow down and eventually select a strategy to propose in a SIP.

The step entails modeling future control strategies, applying the recommended modeled attainment or reasonable progress tests (or hot spot analyses) to see whether the air quality goal(s) is met. This step includes a series of sensitivity tests performed with models. Nature of these tests depends on the amount of confidence one has in the conceptual description of an area's nonattainment or regional haze problem. For example, if a State has reservations about this description, initial sensitivity tests may be relatively coarse ones (e.g., do I control local emissions only or is a regional approach needed?) Future predictions are, of course, subject to uncertainty. Thus, this step also includes sensitivity tests to see the extent to which conclusions about the ability of a strategy to meet a future air quality goal are subject to uncertainties in the modeling analysis. Outcome of these sensitivity tests could serve as one factor which helps a State decide which among several candidate strategies to choose. For example, if one apparently successful (relatively inexpensive[22]) strategy depends on making reductions in a component of particulate matter having large associated uncertainties (e.g., OC) and another, more expensive strategy, depends on reducing a component with less associated uncertainty, (e.g., SO_4) these tests could help a State make a choice among strategies.

<u>Recommendations</u>. States should follow nine steps in applying models to generate information required for use in modeled attainment demonstrations.

1. Formulate a conceptual description of an area's nonattainment or regional haze problem.
2. Develop a modeling/analysis protocol.
3. Choose one or more appropriate models.
4. Choose appropriate days to model.
5. Choose a modeling domain with an appropriate number of layers and appropriately sized grid cells.
6 .Generate appropriate meteorological and air quality inputs.
7. Generate quality assured emissions inputs.
8. Evaluate model performance and undertake diagnostic tests.
9. Simulate and evaluate prospective control strategies.

Execution of steps 3-9 should be performed in accordance with procedures identified in the protocol. Rationale and outcome of the steps should be documented

---

[22]We are not implying that controlling OC is less expensive than controlling SO_4. We are merely using this as an example to illustrate a potential use for sensitivity tests prior to choosing a strategy for a State implementation plan.

as described in Section 7.0. To increase the likelihood of an approvable demonstration, States should carefully coordinate development and execution of steps with the appropriate U.S. EPA regional office(s).

9.0 How Do I Develop And Use A "Conceptual Description"?

The State which contains a nonattainment or Class I area should lead efforts to develop a conceptual description of the related $PM_{2.5}$ or visibility problem. If mutually agreed upon, leadership may be delegated to the appropriate regional planning organization or another group. A State should start developing information to support a modeled attainment demonstration by assembling and reviewing available air quality, emissions and meteorological data. Current design values should be calculated at each $PM_{2.5}$ monitoring site, as described in Section 3.0. Speciated data should be reviewed to get a sense of what component(s) might be contributing most significantly to nonattainment or light extinction. If past modeling has been performed, the emission scenarios examined and air quality predictions may also be useful. Readily available information should be used by a State to develop an initial conceptual description of the nonattainment or regional haze problem in the area which is the focus of a modeled demonstration. A conceptual description is instrumental for identifying potential stakeholders and for developing a modeling/analysis protocol. It may also influence a State's choice of air quality model, modeling domain, grid cell size, priorities for quality assuring and refining emissions estimates and choice of initial diagnostic tests to identify potentially effective control strategies. In general, a conceptual description is useful for helping a State to identify priorities and allocate resources in performing a modeled demonstration. An example of a detailed conceptual description can be found in Pun and Seigneur (1998) and in Pun and Seigneur (1999).

In this Section, we identify key parts of a conceptual description. We then present examples of analyses which could be used to describe each of these parts. We note that initial analyses may be complemented later by additional efforts performed by those implementing the protocol, and that many of the analyses we describe would be more convincing with improved data bases. Thus, a conceptual description is likely to evolve during the course of analyses underlying attainment or reasonable progress demonstrations.

9.1 What Is A "Conceptual Description"?

A "conceptual description" is a qualitative way of characterizing the nature of an area's nonattainment or regional haze problem. It is best described by identifying key components of a description. To illustrate, we pose two sets of questions below. The first illustrates what might be asked in assessing the nature of a nonattainment problem.. The second focuses on preparing for a reasonable progress demonstration. The examples are not necessarily comprehensive. There are other features of an area's problem, such as data gathering priorities, which are also important. For purposes of illustration later in the discussion, we have answered each of the example questions posed below. Our responses appear in parentheses.

Example nonattainment application (annual NAAQS)

-1. Is the nonattainment problem primarily a local one, or are regional factors important?

(Surface measurements suggest that only design values in or immediately downwind of

the city violate the NAAQS. However, other nearby design values come close to the concentration specified in the NAAQS)

-2. What is the relative importance of measured primary and secondary components of $PM_{2.5}$ measured at sites violating the NAAQS?

(Secondary components (i.e., SO_4, NO_3, OC) constitute about 80% of the measured mass of $PM_{2.5}$).

- 3. What are the most prevalent components of measured $PM_{2.5}$?

(The most important components in ranked order are mass associated with SO_4, OC and inorganic primary particulate matter (IP)).

- 4. Does the measured mix of PM components appear to roughly agree with mix of emission categories surrounding the monitoring sites?

(No. Relative importance of measured crustal material (IP) appears less than what might be inferred from the inventory).

- 5. Do there appear to be any potential hot spots for primary particulate matter in unmonitored areas?

(Cannot really tell for sources of crustal material until we resolve the preceding inventory/monitoring discrepancy. There are no other obvious major sources of primary particulate matter which may pose a hot spot problem).

- 6. Is there any indication of what precursor might be limiting formation of secondary particulate matter?

(No indicator species analyses have been performed. Past analyses performed for ozone-related SIP revisions suggest that ozone in this area may be limited by availability of VOC).

- 7. Do monitored violations occur at locations subject to mesoscale wind patterns (e.g., at a coastline) which may differ from the general wind flow?

(No.)

- 8. Have there been any recent major changes in emissions of PM or its precursors in or near the nonattainment area? What?

(Yes, 4 measures believed to result in major reductions in VOC have been implemented in the last 5 years. Reductions in SO_2 emissions have resulted from the national program to reduce acid deposition.)

- 9. Are there discernible trends in design values or other air quality indicators which have accompanied a change in emissions?

(The period of record is insufficiently long to tell).

- 10. Is there any apparent spatial pattern to the trends in design values?

(No.)

- 11. What past modeling has been performed and what do the results suggest?

(A regional modeling analysis has been performed for ozone. Two emission scenarios were modeled: current emissions and a substantial reduction in NOx emissions throughout a regional domain. Reduced NOx emissions led to substantial predicted reductions in 8-hour daily maximum ozone in most locations, but changes near the most built-up area in the nonattainment area in question were small or nonexistent.)

-12. Are there any distinctive meteorological measurements at the surface or aloft which appear to coincide with occasions with $PM_{2.5}$ concentrations in excess of 15.0 $\mu g/m^3$?

(Other than routine soundings taken twice per day, there are no measurements aloft. There is no obvious correspondence with meteorological measurements other than daily maximum temperatures are always $\geq$ 85F on days with the highest $PM_{2.5}$ observations.)

- 13. Do periods with high measured particulate matter or components of particulate matter appear to track each other or any other measured pollutant?

(There appears to be some correspondence between measured high concentrations of SO_4 and ozone).

Using responses to the preceding questions in this example, it is possible to construct an initial conceptual description of the nonattainment area's ozone problem. First, responses to questions 1, 2 and 3 suggest there is a significant regional component to the area's nonattainment problem. Second, responses to questions 1 and 3 indicate there is a local component to the problem.. The responses to questions 11,12 and 13 suggest that there may be a link between reducing ozone and reducing particulate matter. Thus, it may be appropriate to assess effects of previously committed to strategies to reduce ozone before simulating additional control measures. The responses to questions 4 and 5 suggest that it is premature to determine whether one or more hot spot analyses will be needed. The response to question 7 suggests that it may not be necessary to model with very small grid cells, at least for the secondary components of $PM_{2.5}$.

The preceding conceptual description implies that the State containing the nonattainment

area in this example will need to involve stakeholders from other, nearby States to develop and implement a modeling/analysis protocol. It also suggests that a nested regional modeling analysis will be needed to address the problem.

Example reasonable progress application

- 1. What components of particulate matter appear to have high concentrations on days with poor visibility?

(Mass associated with SO_4 and coarse particulate matter (CM) seem to have the highest concentrations on most such days).

- 2. What are typical values for the humidity adjustment factor during the times of year when most of the days with poor visibility occur?

(Typical values appear to be about “4.0").

- 3. Does visibility appear to track well among nearby Class I areas?

(Can’t tell yet).

- 4. Does poor visibility seem to occur under any specific meteorological conditions?

(This information is not readily available).

- 5. Does poor visibility seem to coincide with high observed concentrations of any particular other pollutant?

(There seems to be some correspondence with high regional ozone concentrations)

- 6. What components of particulate matter appear to have relatively high concentrations on days with good visibility?

(Coarse particulate matter and OC)

- 7. What are typical values for the humidity adjustment factor during times of year when most of the days with good visibility occur?

(About “2.3")

- 8. Does good visibility appear to occur under any specific meteorological conditions?

(Don’t know.)

Answers to the preceding questions suggest that strategies to reduce sulfate concentrations and, perhaps, regional ozone concentrations might be effective in reducing light extinction on days when visibility is currently poor. The responses suggest that a strategy which focuses on this alone should first be tried for the days with good visibility as well. Even though sulfate concentrations appear low on such days, the fact that sulfates scatter light efficiently (see Equation (4.1)) and relative humidity is still high enough to enhance this effect is worth considering. Responses suggest that further meteorological analyses would be worthwhile prior to selecting strategies to simulate with a resource intensive regional model.

It should be clear from the preceding two examples that the initial conceptual description of an area's nonattainment problem draws on readily available information and need not be detailed. It is intended to help launch development and implementation of a modeling/analysis protocol in a productive direction. It will likely be supplemented by subsequent, more extensive modeling and ambient analyses performed by or for those implementing the modeling/analysis protocol discussed in Section 10.0.

<u>Recommendations</u>. States should begin an analysis to support a modeled attainment or reasonable progress demonstration by developing a conceptual description of an area's nonattainment or regional haze problem. This description is based on use of readily available air quality, meteorological and emissions information. It may be refined later as additional analyses are performed by those implementing the modeling/analysis protocol.

9.2 What Sorts Of Analyses Might Be Useful For Developing And Refining A Conceptual Description?

Questions like those posed in Section 9.1 can be addressed using a variety of analyses ranging in complexity from an inspection of air quality data to sophisticated mathematical analyses. We anticipate the simpler analyses will often be used to develop the initial conceptual description. These will be followed by more complex approaches or by approaches requiring more extensive data bases as the need later becomes apparent. In the following paragraphs, we revisit key parts of the conceptual description identified for the nonattainment problem discussed in Section 9.1. We note analyses which may help to develop a description of each part. The list serves as an illustration. It is not necessarily exhaustive.

1. Is the nonattainment problem likely to be most affected by regional or local control measures?

- Note spatial distribution of observed design values.
- Note spatial correlation among day to day changes in observed $PM_{2.5}$.
- Note spatial correlation in day to day changes in each of the components of $PM_{2.5}$.
- Note typical amount of each component of $PM_{2.5}$ relative to measured mass of $PM_{2.5}$—what is the relative importance of known secondary components (SO_4, NO_3) vs. known primary components (IP, EC).
- Apply a receptor model like CMB or a multi-variate model–are any local source categories

frequently identified as major contributors to observed $PM_{2.5}$?

2. What control strategies appear potentially promising to simulate?

- Note typical relative importance of primary vs. secondary components—does this relative importance differ on days with measured mass > 15.0 $\mu g/m^3$?
- Apply one or more receptor models to identify source categories which appear to impact monitoring sites frequently.
- Review inventory information to identify potentially important sources of $PM_{2.5}$ and its precursors.
- Map spatial configuration of sources.
- Develop pollution roses for $PM_{2.5}$ and its components to see whether wind orientation on days with high observations yields clues about origins of high observed $PM_{2.5}$ or components.
- If secondary components appear important, combine receptor model results with trajectory models or other meteorological analyses to provide clues about the origins of precursors to secondary components.
- If secondary components appear important, apply indicator species methods to assess precursor(s) which most likely formation of secondary particulate matter.

Resolving questions like the two preceding ones should provide further insights about the scope of needed model simulations. For example, if the analyses suggest that the problem is a regional one in which secondary components are important, this tells us that using a regional scale model which considers atmospheric chemistry will likely be necessary. This, together with emission maps, tells us things about needed domain and grid cell size. If, on the other hand, the conceptual description suggests that the nonattainment problem can be addressed by controlling emissions of primary particulate matter, scope and focus of the modeling can be quite different.

<u>Recommendations</u>. States should analyze ambient air quality, meteorological and emissions data in concert with an air quality modeling analysis. These analyses perform at least 3 functions. First, they are needed to help develop a conceptual description of a nonattainment area's problem. Second, they help guide application of a model in an air quality modeling analysis. Third, analysis of air quality, meteorological and emissions data generates corroborative information which may confirm conclusions drawn with an air quality model or cause some of the underlying assumptions in the modeling to be reexamined.

10.0 What Does A Modeling/Analysis Protocol Do, And What Does Developing One Entail?

Developing and implementing a modeling/analysis protocol is a very important part of an acceptable modeled attainment demonstration. Much of the information in U.S. EPA (1991) regarding modeling protocols remains applicable. States should review the 1991 guidance on protocols. In this document, we have revised the name of the protocol to "Modeling/Analysis Protocol" to emphasize that the protocol needs to address all types of analyses considered in a weight of evidence determination, not just modeling. The nature of a potential $PM_{2.5}$ nonattainment problem may be more diverse than one for ozone. This means that a State may need to apply more than one model to design a strategy which will attain the NAAQS. If so, the protocol will need to address how best to apply two or more models which are different in their approach and coverage.

10.1 What Is The Protocol's Function?

The most important function of a protocol is to serve as a means for planning and communicating how a modeled attainment or reasonable progress demonstration will be performed *before* it occurs. The protocol serves as a vehicle for review of substantive aspects of the planned analyses as they occur, and establishes processes by which affected groups may provide inputs to the analyses. The protocol is also the means by which States and other stakeholders can assess applicability of default recommendations described herein and develop alternatives. A good protocol should lead to widespread participation in developing the demonstration. It should also reduce risk of spending time and resources on efforts which the appropriate U.S. EPA Regional Office(s) believes are unproductive or inconsistent with Agency policy.

The protocol also serves several important, more specific functions. First, it identifies methods and procedures used to support the demonstration. Second, it identifies who will be helping the State or local air quality agency (generally the lead agency[23]) to undertake or evaluate analyses needed to support a defensible demonstration (i.e., the stakeholders). Third, it identifies how communication will occur among stakeholders (e.g., States/tribes, environmental groups, those in the private sector potentially affected by control strategies) to develop consensus on various issues. Fourth, the protocol describes the review process applied to key steps in the demonstration. Fifth, it describes how changes in methods and procedures or in the protocol itself are agreed upon and communicated with stakeholders and the appropriate U.S. EPA regional office(s). Major steps taken in implementing the protocol should be discussed with the appropriate U.S. EPA regional office(s) as they are being decided. States should update the protocol as major decisions are made concerning forthcoming analyses.

---

[23]For reasonable progress demonstrations, the "lead agency" for modeling/analysis may be a regional planning organization if that is agreeable to its member States.

10.2 What Subjects Should Be Addressed In The Protocol?

There are two broad categories of subjects which should be addressed in a protocol. The first of these is "substantive topics". These include such things as the analyses which will be performed, how they will be used, etc. The second category is "procedural topics". This category includes a discussion of how issues will be communicated and resolved as they arise. We list important substantive and procedural topics below.

(a) Substantive topics

1. Indicate choice of the air quality simulation model(s) to be used and how the model(s) meets requirements in 40CFR51, Appendix W for using "alternative" models.

2. Provide assurance that proposed modeling procedures have been scientifically peer reviewed and the protocol includes plans for technical review of how procedures are used in the specific application and the resulting outputs.

3. Identify the emissions, air quality and meteorological data bases to be used to support air quality modeling and other types analyses used in a weight of evidence determination.

4. Identify criteria/goals to be used to select periods to model and the process used to select the specific candidate periods to be modeled.

5. Note and discuss the underlying rationale for the domain size and spatial resolution to be used.

6. Identify methods used to quality assure meteorological and emissions inputs.

7. Note the rationale to be used to choose methods for generating meteorological inputs and for selecting appropriate models to convert emissions into forms compatible with the chosen air quality model(s).

8. Identify model performance evaluation procedures and performance-related diagnostic tests which are planned.

9. Identify sensitivity and other tests to be used to select a mix of control measures needed to meet the air quality goal(s).

10. Identify the types of analyses to be included in a weight of evidence determination should those implementing the protocol decide to use weight of evidence.

11. Identify outcomes for each selected weight of evidence analysis which would be consistent with showing that a selected strategy will meet the air quality goal(s).

(b) Procedural topics

1. Identify stakeholders participating in the process.

2. Discuss management/communication procedures to be used.

3. Identify methods for resolving conflicts.

4. Identify procedures for updating the protocol as new findings come to light during the analyses.

5. Discuss procedures to be followed if a model's performance is shown to be inadequate for simulating a particular period or strategy.

6. Identify specific deliverables and the schedule for delivery to the appropriate U.S. EPA regional office.

7. Identify procedures to be used to archive, document and report results consistently with our guidance in Section 7.0.

> **<u>Recommendations</u>. States should prepare a modeling/analysis protocol as part of an acceptable demonstration of attainment or reasonable progress. The protocol should address substantive and procedural issues. Generally, procedures recommended in the 1991 guidance and followed for the 1994 ozone SIP revisions are also appropriate for $PM_{2.5}$- and visibility-related model applications. These procedures should be augmented to include a discussion of all analyses to be included in the weight of evidence determination, not just modeling. The protocol should also include provision for review of key parts of the analysis and data base underlying the attainment or reasonable progress demonstration. The protocol should be kept up to date to reflect major changes in initial plans.**

(This page is left blank intentionally)

11.0 What Should I Consider In Choosing Air Quality Model(s)?

We begin this section by identifying the types of models which are most suitable for addressing secondary and primary components of particulate matter. We then relate this to choice of model types to address PM NAAQS- and regional haze-related applications. We next identify a set of general requirements which an air quality model should meet to qualify for use in an attainment or reasonable progress demonstration. We then identify several factors which will help in choosing among qualifying air quality models for a specific application. We conclude by identifying several air quality models which are available for use in attainment or reasonable progress demonstrations.

Models are, in reality, *modeling systems* which integrate an emissions model or procedures for deriving emissions, a meteorological model or procedures for deriving meteorological inputs and a chemical transport model. Meteorological and emissions models are discussed in Sections 14.0 and 15.0, respectively. In this guidance, the term "air quality model" means "air quality modeling system". Some modeling systems are modular, at least in theory. This means that it is possible to substitute alternative emissions or meteorological models within the modeling system. Often however, choice of an emissions or meteorological model or their features is heavily influenced by the chosen air quality model (i.e., an effort is needed to develop software to interface combinations of emissions/meteorological/air chemistry models differing from the modeling system's default combination). Thus, choice of an appropriate air quality model(s) is among the earliest decisions to be made by those implementing the modeling/analysis protocol.

To better understand our recommendations concerning choice of models, it is useful to summarize key features of the conceptual description for $PM_{2.5}$ presented in Section 1.3. $PM_{2.5}$ consists of "fine mode" as well as some "coarse mode" particles. In $PM_{2.5}$ attainment demonstrations, the principal distinction between the two modes of particles is their origins. Fine mode particles are a mix of primary $PM_{2.5}$ emissions (arising mainly from combustion) and secondary particulate matter. Components of this mix of fine mode particles (i.e., mass associated with sulfates, nitrates, secondary and primary organic carbon and elemental carbon) can act independently of each other in some instances and interact in others. Coarse mode particles (e.g., mass associated with "crustal" material and, in coastal areas, sea salt) occur as a result of physical activity (e.g., grinding, crushing, resuspension). Coarse mode particles act independently of fine mode particulate matter. While most mass associated with coarse mode particles is associated with particles larger than 2.5 μm, some are $\leq$ 2.5 μm. All of these smaller coarse mode particles result from primary emissions of $PM_{2.5}$. Because primary emissions of $PM_{2.5}$ undergo limited atmospheric transformations, resulting ambient concentrations of this portion of $PM_{2.5}$ are likely to be characterized by more pronounced spatial gradients than is the case for secondary particulate matter.

The discussion in the preceding paragraph suggests that it may be cost-effective to model primary and secondary components of $PM_{2.5}$ differently (e.g., using different spatial resolution, domain sizes and, possibly, different models). Thus, primary components of particulate matter

(IP and EC) may be modeled differently than secondary components (SO_4, NO_3). The remaining major component of $PM_{2.5}$ (OC) consists of *both* primary and secondary particulate matter. Ideally, a State should breakup measured OC into primary and secondary origins and then model the primary part of the OC consistently with other primary particulate matter and secondary OC consistently with other secondary particulate matter. However, it is difficult, at present, to distinguish that portion of ambient OC which is of primary vs. secondary origin. Since secondary components of PM are *not* independent of one another, it is important to include treatment of secondary OC in modeling which addresses secondary components of particulate matter. Thus, if a distinction cannot be made between measured primary and secondary OC at a monitoring site, we recommend including OC of primary and secondary origin in modeling which addresses "secondary particulate matter". In addition, it may be necessary in some hot spot analyses to treat primary emissions of OC from a limited number of sources using modeling approaches which are appropriate for addressing primary particulate matter.

States should use a photochemical grid model to simulate effects of strategies to reduce secondary components of particulate matter (i.e., mass associated with SO_4, NO_3 and secondary OC). Because of the regional nature of "regional haze" and relatively high efficiency with which secondary particulate matter scatters light, we believe that photochemical grid models will be needed to perform reasonable progress demonstrations. Based on its conceptual description of a $PM_{2.5}$ nonattainment problem, a State could conclude that the problem can be addressed by reducing primary components of measured $PM_{2.5}$. If this is the case, the State need not use a photochemical grid model in their attainment demonstration if it can present convincing qualitative arguments that an increase in the secondary components of PM will not cause reductions in the primary components to be insufficient to meet the NAAQS.

Greater flexibility is possible in choosing a modeling approach to address primary components of $PM_{2.5}$ (i.e., IP, EC and primary portions of OC) and coarse particulate matter (i.e., needed for regional haze-related applications) than is true for secondary components. That is, it is not necessary to use a model which considers atmospheric chemistry in addressing changes in primary components. Either a numerical grid or a Lagrangian (such as a Gaussian) model may be used. If a Lagrangian model is used to estimate RRF's for primary particulate matter, these estimates should be made at the monitoring location. In some cases, it may be acceptable to use source apportionment results obtained with a receptor model plus rollback to estimate future concentrations of primary PM. This latter approach may be appropriate if (a) sources contributing to monitored observations are ubiquitous (and subject to uniform control measures) or are uniformly mixed, or (b) there is a single source or a single cluster of sources of observed primary particulate matter which is confined to a small, well-defined area. Thus, a State may use a regional photochemical grid model to address secondary components of particulate matter and a less resource intensive inert model applied over a more limited domain with a finer horizontal/vertical resolution to address primary components.

In regional haze-related applications it may suffice when modeling primary particulate matter to use grid cells which are the same size as those used to model secondary components of particulate matter. If there is no reason to believe that there are major individual sources of

primary PM within about 50 km which affect the monitor site in a Class I area, primary components can be considered using the same coarse grid used for the photochemical grid model. This is generally a good default assumption.

<u>Recommendations</u>. States should use a regional scale photochemical grid model to estimate effects of a control strategy on secondary components of particulate matter. Changes in primary components may be estimated using a numerical grid model (with no chemistry), a Lagrangian model or, in some cases, a receptor model/rollback approach.

Modeling used to support attainment demonstrations for the $PM_{2.5}$ NAAQS may consist of the following combinations:

- **photochemical grid model (changes in secondary components) plus finer scale grid model w/o chemistry (changes in primary components);**
- **photochemical grid model (secondary components) plus Lagrangian model(s) or receptor/rollback model (primary components);**
- **fine scale grid model or Lagrangian model or receptor/rollback model (primary components), no modeling of secondary components.**

If the last option is chosen it should include a justification why an increase in secondary components is not likely to lead to a violation of the NAAQS despite the prescribed reduction in primary particulate matter.

A regional scale photochemical grid model should be used to support demonstrations of reasonable progress reducing regional haze. Primary components may be modeled using the same spatial resolution used for the regional analysis. Exceptions may be required on a case by case basis.

11.1 What Prerequisites Should An Air Quality Model Meet To Qualify For Use In An Attainment Or Reasonable Progress Demonstration?

A model should meet several general criteria for it to be a candidate for consideration in an attainment or reasonable progress demonstration. These general criteria are consistent with proposed requirements in 40CFR Part 51, Appendix W (i.e., the "*Model Guideline*") (U.S.EPA, 2000b). Note that we do not recommend a specific model for use in the attainment or reasonable progress demonstrations. At present, there is no single model which has been extensively tested and shown to be clearly superior or easier to use than several alternatives. Thus, the proposed revision to 40CFR Part 51 Appendix W does not identify a "preferred model" for use in attainment demonstrations of the NAAQS for $PM_{2.5}$ or in reasonable progress demonstrations. Using language in 40CFR Part 51 Appendix W, models used for these purposes should meet requirements for "alternative models".

"Alternative models" may be used if they are non-proprietary. A "non-proprietary" model is one whose source code is available for free or for a reasonable cost. Further, the user

must be free to revise the code to perform diagnostic analyses. If such revisions are made, the user needs to show that the introduced changes do not affect the model's air quality estimates. For a model to be "non-proprietary", users should also be able to revise the code in order to improve the model's ability to describe observations in a credible manner. If these latter revisions are done, they need to be documented. Further, the rationale for any such changes needs to be defended in a technical memorandum to the appropriate U.S. EPA regional office(s). If the U.S. EPA believes the changes are substantial, they should be subject to a scientific peer review just as would any other "alternative model".

Several additional prerequisites should be met for an "alternative model" to be used to support an attainment or reasonable progress demonstration.

(1) It should have received a scientific peer review.

(2) It should be applicable to the specific application.

(3) It should be used with a data base which is adequate to support its application.

(4) It should have performed in past applications in such a way that estimates are not likely to be biased low.

(5) It should be applied consistently with a protocol on methods and procedures.

An air quality model may be considered to have undergone "scientific peer review" if the basic approach followed in each of the major components of the modeling system (e.g., chemical transport model, meteorological and emissions models) has been described in the peer reviewed literature. Further, each component should have been tested, the results documented and reviewed by one or more disinterested third parties. We believe that it should be the responsibility of the model developer or group which is applying an air quality model on behalf of a State to document that a "scientific peer review" has occurred. States should then reference this documentation to gain acceptance of an air quality model for use in a modeled attainment or reasonable progress demonstration.

Should the U.S. EPA identify a "preferred model" at some future date, an "alternative model" may still be used in a subsequent application if it is shown to be more appropriate for the specific application. This could be demonstrated by side by side comparisons of predictions obtained with the "preferred" and "alternative" models with observations. While such comparisons may be desirable, they are not required. Criteria described in Section 11.2 may be used to show that an "alternative model" is more appropriate than a "preferred model" for a specific application.

<u>Recommendations</u>. For an air quality model to qualify as a candidate for use in an attainment demonstration for $PM_{2.5}$ or in a reasonable progresss demonstration, a State needs to show that it meets several general criteria.

1. The model has received a scientific peer review.

2. The model can be demonstrated applicable to the problem being addressed.

3. Data bases needed to perform the analysis are available and adequate.

4. Available past appropriate performance evaluations have shown the model is not biased toward underestimates.

5. A protocol on methods and procedures to be followed has been established.

6. The developer of the model must be willing to make the source code available to users for free or for a reasonable cost, and the model cannot otherwise be proprietary.

11.2 What Factors Affect My Choice of A Model For A Specific Application?

States should consider several factors as criteria for choosing a qualifying air quality model to support an attainment demonstration for the $PM_{2.5}$ NAAQS or a reasonable progress demonstration. These factors are: (1) ability to address the air quality problem described in the conceptual description; (2) documentation and past track record of candidate models in similar applications; (3) experience of staff and available contractors; (4) required time and resources vs. available time and resources; (5) in the case of regional applications, consistency with regional models applied in adjacent regions.. The first of these factors is used to identify attributes needed for a model to be chosen. Factors (2)-(5) are used to help choose among candidate models having these attributes. Finally, before results of a selected model can be used in an attainment or reasonable progress demonstration, the model should be shown to perform satisfactorily using the data base available for the specific application.

Ability to address the air quality problem described in the conceptual description. This is the most important criterion for selecting an appropriate model. Prior to selecting a model, we recommend that those implementing the protocol review available air quality, meteorological and emissions data, and take account of the geographic location of the nonattainment or Class I area(s) relative to that of PM and precursor emissions. Section 9.0 identifies some types of analyses which may be useful for developing a conceptual description of an area's nonattainment problem. Once the conceptual description is developed, a State should review the ability of candidate models to address the identified problem(s).

Documentation and Past Track Record of Candidate Models. For a model to be used in an attainment or reasonable progress demonstration, evidence should be presented that it has been found acceptable for estimating concentrations of $PM_{2.5}$ and most of its major components. Preference should be given to models exhibiting satisfactory past performance under a variety of conditions. Finally, a user's guide (including a benchmark example and outputs) and technical

description of the model should be available.

Experience of Staff and Available Contractors. This is a legitimate criterion for choosing among several otherwise acceptable alternatives. The past experience might be with the air quality model itself, or with a meteorological or emissions model which can be more readily linked with one candidate air quality model than another.

Required vs. Available Time and Resources. This is a legitimate criterion provided the first two criteria are met.

Consistency of a Proposed Model with Models Used in Adjacent Regions. This criterion is applicable for regional model applications. If candidate models meet the other criteria, this criterion should be considered in choosing a model for use in a regional or nested regional modeling application.

Demonstration that an "Alternative Model" is Appropriate for the Specific Application. If an air quality model meets the prerequisites identified in Section 11.1, a State may use the factors described in this section (Section 11.2) to show that it is appropriate for use in a specific application. Choice of an "alternative model" for use in a specific attainment or reasonable progress demonstration needs to be reviewed and approved by the appropriate U.S. EPA regional office and by the U.S. EPA Model Clearinghouse.

Satisfactory Model Performance in the Specific Application. Prior to use of a selected model's results, it should be shown to perform adequately in the specific application. Means for evaluating model performance are discussed in Section 16.0.

<u>Recommendations</u>. States should first determine what attributes are needed for a qualifying model to address an area's $PM_{2.5}$ or regional haze problem, and then choose among models possessing these attributes. Five factors should be considered in selecting an air quality model for a specific application. Selection of an air quality model should be concurred with by the appropriate U.S. EPA regional office and U.S. EPA Model Clearinghouse. The five factors are listed approximately in order of importance.

1. Nature of the air quality problem leading to nonattainment of the $PM_{2.5}$ NAAQS or light extinction should first be assessed, and the selected model(s) should have attributes and capabilities which enable it to address the conceptual description of the problem.

2. Availability, documentation and past performance should be satisfactory.

3. Relevant experience of available staff and contractors should be consistent with choice of a model.

4. Time and resource constraints may be considered.

5. Consistency of the model with what was used in adjacent regional applications should be considered.

Prior to using model results in a specific application, a State should show that the model performs adequately in replicating observations available for that application.

11.3 What Are Some Examples Of Air Quality Models Which May Be Considered?

Seigneur, et al., (1997) provide a good review and a more complete discussion of air quality and receptor models which have been used in applications related to particulate matter. Table 11.1 draws upon this information and lists several air quality models (available in 2000) which have been used to simulate ambient concentrations of $PM_{2.5}$ or its components. The list is not intended to be comprehensive. States should review the U.S. EPA's website which lists models and provides short descriptions for updates (***http://www.epa.gov/ttn/scram/***). Exclusion of a model from the list does not necessarily imply that it cannot be used to support a modeled attainment or reasonable progress demonstration. By the same token, inclusion on the list does not necessarily imply that a model may be used for a particular application. States should follow the guidance presented earlier in Section 11 for selecting an air quality model for a specific application.

Table 11.1. Some Air Quality Models Used To Model $PM_{2.5}$ Or Its Components

Model	Reference(s)	Past Applications	Comments
CMAQ	U.S.EPA, 1998, 1999c	Eastern and western halves of continental U.S. by U.S. EPA Office of Research & Development and the Office of Air Quality Planning & Standards, Los Angeles Basin (Seigneur, 2000)	secondary & primary components simulated on regional & urban scales
SAQM-Aero	Dabdub, et al., 1998	San Joaquin Valley, California by California Air Resources Board, Los Angeles Basin (Pai, et al., 2000)	secondary & primary components simulated on regional & urban scales
URM	Kumar, et al., 1996	Southern & mid-Atlantic regions as part of the Southern Appalachian Mountain Initiative (contact Dr.A.G. Russell, GA Inst.of Technology)	secondary & primary components simulated on regional & urban scales
MAQSIP	MCNC, 2000, Odman and Ingram, 1996	Eastern U.S., focused on North Carolina, Los Angeles Basin, (Seigneur, 2000).	secondary & primary components simulated on regional & urban scales

Table 11.1. Some Air Quality Models Used To Model $PM_{2.5}$ Or Its Components (continued)

Model	Reference(s)	Past Applications	Comments
UAM-Aero	Lurmann and Kumar, 1996, Lurmann, et al., 1997 Kumar and Lurmann, 1997	Los Angeles Basin, San Joaquin Valley, CA for SCAQMD and for CARB	secondary & primary components simulated on urban scale
CALPUFF	Strimaitis, et al., 1998	Evaluated with data sets for two power plants	Lagrangian model which considers primary PM for a limited number of sources. Also potentially useful for hot spot analyses.
REMSAD	Systems Applications, International, 1998	Applied for contiguous 48 States by/for U.S. EPA, Office of Air Quality Planning and Standards and Office of Water	primary & secondary PM using condensed chemical mechanism. Most practical for annual applications
AERMOD	Cimorelli, et al., 1998, Paine, et al., 1998	See Paine, et al., 1998	primary particulate matter only, nearby (< 50 km) sources, also useful for hot spot analyses
ISC3	U.S.EPA, 1995, Paine, et al., 1997	Numerous applications	primary particulate matter only, nearby (< 50 km) sources, also useful for hot spot analyses

Table 11.1. Some Air Quality Models Used To Model $PM_{2.5}$ Or Its Components (concluded)

Model	Reference(s)	Past Applications	Comments
CAL3QHCR	Eckhoff and Braverman, 1995	New York City and other locations	line source model, useful for estimating concentrations of primary particulate matter near roadways
CMB (Chemical Mass Balance Model)	U.S. EPA, 1990	Numerous applications (see Seigneur, et al., 1997)	receptor model, most useful for identifying source categories of primary particulate matter impacting monitoring sites
PMF (Positive Matrix Factorization)	Paatero and Tapper, 1994	Narragansett, RI, Alaska, Phoenix	receptor model, most useful for identifying source categories of primary particulate matter impacting monitoring sites
UNMIX	Henry, 2000	Phoenix, AZ	receptor model, most useful for identifying source categories of primary particulate matter impacting monitoring sites

12.0 How Do I Decide Which Periods Or Meteorological Episodes To Model?

How many and which days to model depends primarily on the air quality goal the modeling seeks to address. Therefore, our recommendations differ for the two NAAQS for $PM_{2.5}$ and for the goals reflecting reasonable progress reducing regional haze. Goal-specific criteria are presented in Sections 12.2-12.4. Nevertheless, there are several criteria for choosing days to model which are relevant for all of the air quality goals considered in this guidance. We begin this section by identifying these generally applicable criteria. We then address each of the two NAAQS and reasonable progress goals in turn, and identify additional, goal-specific criteria for selecting days to model. We conclude by identifying some practical considerations in choosing days to model.

12.1 What Are Generally Applicable Criteria For Choosing Days To Model?

There are three such criteria: choose a mix of days reflecting a variety of meteorological conditions; choose days for which an extensive air quality/meteorological data base exists, choose a sufficient number of days so that estimated relative reduction factors do not depend on the outcome on one or a small number of days.

Simulating a variety of meteorological conditions. This criterion is important, because we want to be assured that a control strategy will be effective under a variety of conditions. In the case of the 24-hour NAAQS for $PM_{2.5}$ and the two goals for visibility, greatest priority should be given to ensuring that this variety is drawn from days with high observed concentrations of $PM_{2.5}$ or days with poor (good) current visibility. We believe the most important indicator of variety is differing wind fields. This affects source/source and source/receptor orientations and may sometimes serve as a surrogate for other meteorological characteristics which affect mixing. Maximum daily temperature and moisture content may also have a major influence on the effectiveness of a strategy in reducing concentrations of $PM_{2.5}$ or improving visibility.

Those implementing the modeling/analysis protocol should describe the rationale for distinguishing among episodes which are modeled. If the analysis is to be focused on a limited geographic area, the selection may reflect a number of area specific considerations. Qualitative procedures such as reviewing surface and aloft weather maps, observed or modeled wind patterns may suffice to distinguish episodes with distinctively different meteorological conditions. More quantitative procedures, such as a CART analysis, to identify distinctive groupings of meteorological/air quality parameters may sometimes be desirable. An example of a CART analysis applied to select episodes in a study which addressed regional haze is described by Deuel and Douglas (1998).

Choosing days with extensive data bases. If possible, days selected for modeling should include some with extensive data bases, such as might be collected in a short duration intensive field study. States should try to include days on which $PM_{2.5}$ and speciated particulate matter are collected for sampling periods not exceeding 6 hours (i.e., 4 samples/day). If it is not possible to include any such days for a particular application, a State should provide information

showing that the model being used has performed adequately elsewhere when extensive data bases have been available.

An "extensive" data base should best include vertical profile measurements of wind velocity, temperature, moisture and pressure taken 3-4 times/day or, better still, continuous measurements. Preferably, upper air measurements for SO_4, NO_3, OC, EC, IP, SO_2, NOy, NO, NO_2, and O_3 or as many of these species as possible should be available. Surface measurements for the previously identified meteorological and air quality data should also be available.

Although the tests we recommend place less emphasis on a model being able to predict absolute concentrations of each component accurately, it is important that the model be able to predict the mix of secondary components and ozone in approximately the right proportion. This follows since there is an interdependency among these components which could affect accuracy of estimated RRF's if the model performs poorly. Thus, the need for the chosen model(s) to have performed well on days with good data bases remains an important one.

Choose a sufficient number of days to derive robust estimates for the RRF values. Studies examining the stability of estimated RRF values for ozone suggest that the estimates become more stable if based on an average of ten or more days (Milanchus, et al., 1998, Hogrefe, et al. , 2000). As we will describe shortly, these findings also appear applicable to components of $PM_{2.5}$, but the number of days needed for RRF values to become "stable" may be somewhat larger for PM than for ozone. Modeling a large number of days also increases the number of days with observed $PM_{2.5}$ and speciated data. This, in turn, increases the likelihood of having a large data base to conduct a more meaningful model performance evaluation. Based on our analyses to date, we suggest that States derive their estimated RRF values using modeled concentrations averaged over several (e.g., ~ 15 or more, not including ramp-up) days.[24]

Recommendations. States should observe three criteria which are generally applicable in choosing days to model, plus additional criteria which depend on the air quality goal being addressed. The three general criteria are:

1. Choose a variety of meteorological conditions to model;
2. To the extent possible, include days with extensive data bases;
3. Model a sufficent number of days so that RRF values are based on several (e.g., ≥ ~15) days.

[24]We consider these findings to be preliminary ones. We are performing additional tests to provide more definitive analyses which address the issue of the minimum number of days which have to be modeled in order for the model to provide an adequate indication of the effects of a control strategy.

12.2 Additional Recommendations For The Annual $PM_{2.5}$ NAAQS

(NOTE TO REVIEWERS: we are continuing with our analyses of strategy simulations obtained with the REMSAD model. Prior to finalizing the guidance, we hope to augment these with results obtained using the CMAQ model. The analyses are intended to improve the underlying basis for recommendations on choice of a minimum number of consecutive days to model per quarter to support demonstrations related to the annual and 24-hour NAAQS and the regional haze rule. Recommendations presented herein reflect findings as of December 2000.)

Emissions of $PM_{2.5}$ and its precursors are likely to vary seasonally, as are meteorological conditions. Therefore, we recommend that for applications related to the annual NAAQS, decisions regarding days to model should be made on a calendar quarter by calendar quarter basis. Days chosen from each quarter should reflect a balance of days above and below the observed quarterly arithmetic mean $PM_{2.5}$ concentration, a mix of meteorological conditions and, to the extent possible, include periods with intensive data bases. Approximately an equal number of days should be drawn from each quarter. States may use any one of three approaches as the basis for choosing days to model in applications related to the annual NAAQS. Each of these is described below.

Approach 1. Model every day over a recent 3-year period. This is the method we recommend for performing a hot spot analysis near sources of primary particulate matter. It is also the preferred approach when a Lagrangian (e.g., Gaussian) or other relatively non-resource intensive model is used to estimate relative reduction factors (RRF's) for primary components of particulate matter. One advantage of this approach is that the modeled results can be most readily related to the form of the NAAQS. For example, for each quarter, the relative reduction factor is the ratio of the average of the 3 consecutive mean quarterly future to current predictions for each component of $PM_{2.5}$. The modeling yields these estimates directly. It is not necessary to rely on assumptions about the representativeness of RRF values calculated for a subset of chosen days. A second advantage is that it ensures that there is a large number of days for which both modeled and monitored data exist. This provides a large data base which can be used to evaluate performance of the model.

A variation of this approach, which is acceptable for applications of photochemical grid models, is to choose a single, recent "representative" year and model every day in that year. A State should consider two factors in deciding whether a particular year is a "representative" one. First, the observed annual mean concentrations of $PM_{2.5}$ should be close to the 3-year observed design value at all, or most, monitoring sites. Second, the pattern of quarterly mean values should be similar to the pattern of quarterly mean concentrations averaged over 3 years. For example, if the 3-year mean quarterly $PM_{2.5}$ concentration for winter is higher than that for the other three seasons, this should also be the case for the selected "representative" year. Highest priority should be given to ensuring that these two criteria are met at monitoring sites which exceed the annual NAAQS for $PM_{2.5}$.

The disadvantage of modeling every day is the burden it places on central processing (CPU) time and data storage needs. States will likely want to explore several emission control strategies as well as perform diagnostic/sensitivity tests to estimate the extent to which conclusions drawn from the modeling are affected by a variety of factors. Thus, the CPU and data storage requirements could be large, especially using regional photochemical grid models. This disadvantage may diminish over time as computer hardware and related factors improve. In the meantime however, it may be necessary to use simplified versions of photochemical grid models if one chooses to model every day. This introduces a potential tradeoff between the detail with which physicochemical phenomena are treated vs. possible errors estimating RRF values introduced by simulating a subset of days which is not representative of a year. Thus, we describe two additional approaches for States needing to apply regional photochemical grid models to address the annual NAAQS.

Approach 2. Choose at least a minimum number of days to model from each calendar quarter. The obvious question posed by this approach is, what is a "minimum number of days"? We have reviewed past applications of the REMSAD model to address this question. The REMSAD model has been applied to the contiguous United States for an entire year using meteorological inputs available for 1990. Thus far, we have investigated two emissions scenarios: (1) emissions which are reflective of 1996 estimates, and (2) emissions which have been projected to 2010. These latter emissions reflect Clean Air Act related-measures for VOC, NOx and SO_2 including reductions in NOx expected if the "NOx SIP call" is implemented, the acid rain-related reductions and imposition of maximum available control technology (MACT--mostly reflecting reductions in VOC).

We used the modeled results to calculate quarterly mean concentrations for each component with both emission scenarios. The four quarterly component-specific mean concentrations corresponding to current emissions were averaged to get current annual mean estimates for each component. An identical procedure was followed to get estimates of the annual mean component concentrations corresponding to the 2010 emissions. The relative reduction factor (RRF) for each component is the ratio of the annual mean predicted for 2010 divided by the predicted annual mean with 1996 emissions. We assumed these component-specific RRF's represent "ground truth"—RRF values you would find if every day were modeled.

We next re-derived component-specific RRF's as described above, but with one difference. Only a limited number of modeled days were used to derive the RRF. For example, we chose 4 days at random from each quarter to calculate the component-specific RRF values. Typically, this produces estimated component-specific RRF's which differ somewhat from what is obtained when every day is considered. We then replaced the 4 days/quarter chosen previously and randomly redrew 4 days from each quarter. Once again, resulting RRF values were calculated using the limited sample. We repeated this process 200 times. This gives a distribution of component-specific RRF values which is attributable to not considering results from every modeled day. We then calculated the mean (μ) and standard deviation (σ) of the RRF values obtained with the 200 limited samples, as well as the coefficient of variation (σ/μ) for this

distribution.

We can use the coefficient variation to get a sense of how many days need to be modeled to get RRF values which agree closely with what would be obtained if all days were modeled. This insight is possible by considering what happens to the coefficient variation when all (approximately 91) days are modeled in each quarter. If every day were modeled, the coefficient of variation becomes zero. This follows, because the exact same days are used to calculate the RRF in each of the 200 trials----there is no variability in the resulting RRF value ($\sigma = 0$). As the number of sampled days included in each trial gets large, the variability in the estimate approaches zero. Further, there would likely be little difference in the coefficient of variation as the sample size increases.

Figures 12.1 (a) - (c) show the coefficient of variation in the RRF calculated for NO_3 and SO_4 in each of 64 surface grid cells in and near Atlanta, Cincinnati and the Grand Canyon, respectively. The coefficient of variation (the ordinate in the graphs) is expressed as a percentage of the mean calculated RRF. The whisker diagrams in each of the figures indicate that, for any given sample size, there is some cell to cell variability in the coefficient of variation. The thick line in each plot represents typical values for the coefficient of variation. For example, in Figure 12.1 (a), we see that a typical coefficient of variation in the RRF for NO_3 in Atlanta is about 7% of the estimated RRF value when the RRF is derived using 8 randomly chosen modeled days per quarter. Notice in the figure how there is a relatively large increase in the typical coefficient variation (to about 11%) when the RRF calculations are based on 4 modeled days/quarter. In contrast, differences in the coefficient of variation become progressively smaller as the sample size is increased. For example, the typical coefficient of variation obtained with a sample size of 16/quarter (~ 5%) is slowly reduced to about 3% with a sample size of 32 modeled days/quarter.

Note that our results suggest that the coefficient of variation of the RRF for SO_4 is less than that for NO_3. For example, in Figure 12.1 (a) with a sample size of 8 modeled days/quarter, the coefficient of variation of the RRF for SO_4 is about 2%, whereas the value related to the RRF for NO_3 is about 7%. We believe this difference is due in part to much lower NO_3 than SO_4 concentrations which the model predicts in the vicinity of Atlanta. This introduces increased "noise" in the model predictions for NO_3. This hypothesis is supported by the generally higher coefficients of variation of RRF's for NO_3 and SO_4 which we calculate in the vicinity of the Grand Canyon (Figure 12.1 (c)) compared to those estimated in Figures 12.1 (a) and (b) for Atlanta and Cincinnati.

Figure 12.1 (a).

Figure 12.1 (b).

Figure 12.1 (c).

Choice of a limited number of days/quarter to model may sometimes be dictated by the component of $PM_{2.5}$ which has the highest coefficient of variation in its RRF value with a limited sample size. Due to "noise" in model predictions, this is often likely to be a component for which very low current concentrations are estimated. Why do we say a lower acceptable limit on number of days modeled may only "sometimes" be dictated by the component with the highest coefficient of variation? If the observations for this component are also very low relative to other components, variability/uncertainty in the component's RRF value may be unimportant (i.e., you gain little by reducing the uncertainty in the RRF for this component by modeling many days). Recall that in the attainment and reasonable progress tests, we recommend that component-specific RRF's be multiplied by currently monitored component concentrations. If the observations are also very low, uncertainty in the RRF (due to limited modeling) will likely have little effect on the outcome of the test.

Looking at the information we have generated to date, we believe that modeling about 15-20 days/quarter should provide an acceptable approximation of RRF values related to the annual NAAQS. This is based on the relatively low values for the coefficient of variation in the RRF seen in the figures for randomly chosen sample sizes of 16 days/quarter (i.e., typically only about 5% for the NO_3 RRF, even at the Grand Canyon) and the relatively small reduction in the uncertainty which occurs when the RRF estimates are based on more modeled days. Smaller sample sizes may be used on a case by case basis, if justified (e.g., by choosing days in something other than a random manner).

Approach 3. Classify observed air quality data into groups defined by differences in meteorological conditions. Model at least three days from each group. This approach is to divide days within each calendar quarter into major meteorological categories. Modeling should include at least three days from each identified category. Quarterly mean concentrations are then computed for each major component of $PM_{2.5}$ by weighting predictions obtained for each category by the frequency with which that category is observed in the calendar quarter over a 3-year period, preferably corresponding to the current measured design value for $PM_{2.5}$. Quarterly means obtained in this manner are then averaged to derive estimated annual means. Annual means are calculated using current emissions and then with future estimated emissions reflecting growth and controls. The component-specific RRF values are the ratios of annual mean concentrations predicted for each component with future emissions to those predicted with current emissions.

The biggest impediment to using approach 3 is developing a sound rationale for dividing days into a finite number of meteorological classes and then being able to assign each day to the appropriate class. If the model application is for a mesoscale or smaller problem, air quality observations may be used to help identify distinctive classes. The CART procedure, illustrated by Dueul and Douglas, (1998), is an example of such a methodology. Even though these analysts were working with a regional (synoptic scale) air quality model, they were able to use this approach because they were interested in applying model results to a limited number of receptor locations (i.e., Shenandoah and Great Smokies National Parks).

Using air quality observations as a means for identifying meteorological classes becomes impractical if one is applying a regional scale model to simultaneously derive solutions for problems meeting the annual NAAQS or another air quality goal in numerous locations. Cohn, et al., (2000) have developed a means for classifying days based solely on observed geographic patterns of wind velocity observed aloft at 700 mb. Methods such as this one, which use synoptic scale observations to identify meteorological classes, may be most practical for regional scale model applications needing to consider many receptor sites.

Recommendations. The preferred approach for choosing days to model in applications addressing the annual NAAQS for $PM_{2.5}$ is to model every day for a 3-year period which is as recent as feasible. For applications requiring use of a photochemical grid model, use of a single, representative year is an acceptable alternative.

If the preferred approach is not practical, States may use one of two alternatives for choosing days to model.

1. Model all 4 calendar quarters, considering at least 15 days/quarter. Fewer days/quarter may be acceptable if justified on a case by case basis.

2. Identify major classes of meteorological conditions occurring in each calendar quarter. Model at least three days from each class in each quarter. Obtain quarterly mean concentration estimates by weighing each class according to its relative frequency of occurrence.

12.3 Additional Recommendations For The 24-Hour $PM_{2.5}$ NAAQS

The preferred approach for choosing days to model primary components of $PM_{2.5}$ is to model every day over a consecutive 3-year period, preferably corresponding to the period in which the current design value is observed. This approach should generally be followed for hot spot analyses. We also recommend it for use in grid models used to support the modeled attainment test. If a nonattainment area violates both the annual and 24-hour NAAQS and our recommended approach for selecting days for modeling to address the annual NAAQS is followed (i.e., model every day for a year), a modeled data base suitable for use with the 24-hour NAAQS may also exist. If it is not practical to model every day in a year to derive RRF values for components of $PM_{2.5}$ in the modeled attainment test, States may consider the suggestions outlined below.

When a photochemical grid model is used, an alternative approach for choosing days to model for applications related to the 24-hour NAAQS is to model every day where there is an observed daily concentration of $PM_{2.5}$ which exceeds 65 $\mu g/m^3$. There could be a substantial amount of uncertainty associated with predictions on any single day. Thus, the modeled attainment test is most likely to be reliable when relative reduction factors reflect composite responses from many days. Therefore, we recommend modeling as many days as feasible where

observed $PM_{2.5}$ is greater than 65 $\mu g/m^3$. States should make note of the quarter in which each selected day occurs. As with the annual NAAQS, the preferred approach is to develop relative reduction factors which are season (i.e., quarter) specific. Emissions, meteorological conditions and thus composition of the mixture comprising $PM_{2.5}$ could vary substantially from season to season.

We have noted that it is desirable to base our conclusions on a composite response of the model(s) over many days. However, from limited monitoring data available as of 2000, there is not likely to be many days with observed concentrations greater than 65 $\mu g/m^3$. If this results in a sample size of smaller than about 10 days per quarter, days with $PM_{2.5}$ observations > ~55 $\mu g/m^3$ and speciated measurements should also be considered. If, after doing this, there is still a small number of days in a single quarter, it may be combined with the most appropriate other quarter. If the sample of high days is limited in 2 or more quarters, a State may develop relative reduction factors which reflect a combined sample of days from the entire year.

We expect relative reduction factors for components of $PM_{2.5}$ to behave similarly to those for ozone. That is, the larger the number of modeled days used to derive the RRF, the greater the likelihood we will obtain a representative estimate for the RRF. Hogrefe, et al., (2000) have noted that RRF values for ozone become relatively stable if derived using estimates obtained on 10 or more days[25]. Therefore, in applications related to the 24-hour NAAQS, States should base the RRF on modeled results from 10 or more days with monitored observations $\geq$ 55 $\mu g/m^3$. If there are fewer than 10 such days available, use as many days as are available.

Recommendations. In performing hot spot analyses, States should model every day in a recent 3-year period. This is also the preferred procedure for deriving RRF values for components of $PM_{2.5}$ considered in the modeled attainment test. An acceptable alternative is to model everyday in a representative year. If it is not practical to use either of these approaches to derive RRF values, States may consider the alternative identified below.

States should base RRF values obtained with photochemical grid models on at least 10 days/quarter with monitored concentrations of $PM_{2.5} \geq 55$ $\mu g/m^3$. If there is fewer than 10 such days/quarter available, quarters may be pooled to obtain RRF values based on approximately 10 or more days observing $PM_{2.5}$ concentrations ≥ 55 $\mu g/m^3$. Preference should be given to pooling quarters with as similar emissions data and meteorological conditions as feasible.

12.4 Additional Recommendations For The Regional Haze Goals

The goals for regional haze focus on the 20% of days with best and worst visibility.

---

[25]We intend to perform model sensitivity tests to confirm or refine this estimate for PM-related applications. Results will be reflected in the final version of the guidance.

Sampling in Class I areas is expected to occur once every three days. Thus, each year there will be about 24 "worst" days and 24 "best" days to choose from. Since the base period against which reasonable progress is to be gauged is 5 years long, there could be as many as 120 "best" and "worst" days to choose among for modeling.

It is likely that numerous Class I areas will be considered simultaneously in a modeled reasonable progress demonstration. Thus, the preferred approach for regional haze-related model applications is to simulate an entire, representative year (i.e., one whose mean derived b_{ext} values for "20% worst" and "20% best" days approximates mean values for b_{ext} averaged over 5 years for the best and worst days). States can then base the RRF values on the 24 best and worst days in each Class I area for that year. If it is not feasible to model an entire year, we offer some alternatives in the following paragraphs.

Days with "worst" visibility. Earlier, we showed that RRF values based on a random sample of 15 days/quarter was likely to provide a good estimate of the RRF which would be obtained if one were to model every day in a year. This implies that modeling about 1/6 of a large population of possible days produces a mean RRF value which is very close to what one would get if one were to model every day. Modeling 1/6 of "worst" visibility days in a year would mean modeling about 4 days with observed "worst" visibility. However, based on information presented for ozone by Hogrefe, et al., (2000), our sense is that the reliability of RRF estimates based on samples this small may not be reliable. Thus, as of late in 2000, we feel that a minimum number of days to model is probably better described by the results obtained by Hogrefe, et al. (2000) for ozone. Therefore, we recommend that each Class I area considered in a regional modeling analysis have RRF values based on model results from at least 10 monitored "worst" visibility days. States should consider seasonal patterns in observed "worst" visibility in choosing days to model.

Since regional scale models will likely be used for regional haze-related applications, strategies for several Class I areas will likely be developed from the same model results. Thus, having a criterion that the RRF for each Class I area be based on at least 10 modeled days drawn from a population of "worst" days with monitored data will result in more than 10 "worst" days being modeled for some Class I areas. Thus, meeting this criterion ensures that the RRF for most Class I areas will be based on model results from more than 10 of their worst visibility days.

Days with best visibility. The criterion for choosing days to model to meet the goal for "best" visibility days is the same as that for "worst" days, except that the minimum number of 10 days should be drawn from the population of days with observed "best" rather than "worst" visibility.

Recommendations. The preferred approach for regional haze-related modeling is to simulate a full, representative year. RRF values should be based on model results averaged over the 20% worst and 20% best days determined (using monitored PM data) for each Class I area in the chosen year. If it is not feasible to model a full year, States should derive RRF values for each Class I area based on modeling

results from at least 10 days with observed "worst" ("best") visibility.

12.5 Some Other Considerations

Choose periods which have already been modeled. That is, of course, provided that past model performance evaluation for such an episode was successful in showing that the model worked well in replicating observations. Given that criteria described in Sections 12.1, 12.2, 12.3 and 12.4 are met approximately as well by such episodes as they are by other candidate episodes, a State could likely save a substantial amount of work in evaluating model performance.

Choose periods which are drawn from the years upon which the current design value is based (i.e. the "base period" for regional haze-related applications). Fewer emission estimates and fewer air quality model simulations are needed if the period used to evaluate model performance, and the "current period", used in the recommended modeled attainment test, are one in the same.

It is desirable to include weekend days among those chosen. Weekend days reflect a different mix of emissions than occurs on weekdays. Thus, for increased confidence that a control strategy is effective it needs to tested on weekends as well as on weekdays. If emissions and spatial patterns of high $PM_{2.5}$ or its components do differ on weekends vs. weekdays, including weekend days in the choice of episodes may provide a potential for evaluating accuracy of a model's response to changes in emissions. As we note in Section 16.0, such evaluations are highly desirable.

If a State chooses to model several nonattainment areas simultaneously, choose periods which meet the other criteria in as many of these nonattainment (or Class I) areas as possible. A State or group of States may decide to apply a regional model or a nested regional model to demonstrate attainment in several nonattainment areas or reasonable progress in several Class I areas at once. Time and resources needed for this effort could be reduced by choosing periods to model which meet the other criteria in several nonattainment or Class I areas which are modeled.

<u>Recommendations</u>. The following are additional, practical criteria for choosing which days to model for an attainment or reasonable progress test. States may identify, document and present the rationale for practical criteria in addition to these if they choose.

1. Give preference to previously modeled periods.

2. Give preference to periods occurring during the years corresponding to the current design value (base period) used in the modeled attainment (reasonable progress) test.

3. Include weekends among the selected days.

4. If applying a regional model, choose periods to model which meet other, previously identified selection criteria in as many nonattainment (or Class I) areas as possible.

13.0 What Should I Consider When Selecting A Modeling Domain And Its Horizontal/Vertical Resolution?

A modeling domain identifies the geographic bounds of the area which is modeled. In a grid model, horizontal resolution is related to the horizontal dimensions of individual grid cells. Vertical resolution is determined by the number of grid cells (i.e., layers) considered in the vertical direction.

Appropriate domain size and resolution differ for applications related to the NAAQS for $PM_{2.5}$ vs. regional haze. Further, different recommendations are appropriate when using Lagrangian models, such as Gaussian models, to assess effects of reducing emissions of primary particulate matter. In Section 13.1, we address the issues of domain size/grid resolution as they relate to model applications for the NAAQS for $PM_{2.5}$. We begin by discussing appropriate domain size and resolution for grid model applications. We then discuss Lagrangian model applications which may be used for developing relative reduction factors for primary particulate matter or for hot spot analyses. We conclude Section 13.1 by discussing what States should do for analyses in which both grid and Lagrangian models are used. In Section 13.2, we turn our attention to model applications addressing reasonable progress reducing regional haze.

13.1 How Do I Decide What Domain Size And Resolution Are Appropriate To Address The Two $PM_{2.5}$ NAAQS?

13.1.1 How Do I Decide What Domain Size To Use?

Grid model applications. Choice of domain size should be heavily influenced by a State's conceptual description of its nonattainment problem as well as by the remedies it wishes to consider. For example, if a review of ambient measurements suggests that primary particulate matter (IP, primary OC, EC) constitutes enough of measured $PM_{2.5}$ at sites violating the NAAQS so that control measures for these components may reduce them sufficiently to meet the NAAQS, a State may choose to use an urban scale domain. Using an urban scale domain introduces an issue about what to assume for boundary conditions (or "background" in Lagrangian applications). We discuss this further in Section 14.0. Generally, an urban scale analysis is most appropriate for nonattainment areas where monitored data suggest transport of particulate matter or its precursors into the nonattainment area is well below the level of the NAAQS or is believed unlikely to change. If a State's attainment strategy is likely to depend on efforts to reduce secondary components of $PM_{2.5}$ (i.e., SO_4, NO_3 , secondary OC), the model application should use a regional domain. Size of the regional domain will be influenced by the configuration of sources and by wind patterns observed or predicted during periods in which the 24-hour or annual NAAQS is frequently exceeded by monitored observations.

What do we mean by "urban scale" and "regional" domains? An urban scale domain is one having horizontal dimensions less than ~ 300 km on a side. Assuming the nonattainment area is located near the center of the domain, the domain should be large enough to ensure that emissions occurring shortly before sunrise in its center are still within the domain at the end of

the same calendar day. If recirculation of the nonattainment area's previous day's emissions is believed to contribute to an observed problem, the urban scale domain should be large enough to characterize this. If recirculation encompasses distances larger than about 300 km, an urban scale model is probably not sufficient to address an area's problem.

A regional domain is one having horizontal dimensions typically exceeding 1000 km on a side. Data base management problems may make it infeasible to use the same horizontal grid cell size in urban scale and regional models. Using nested regional models addresses this problem. A nested regional model is one whose domain typically exceeds 1000 km on a side. However only a portion of that domain (e.g., < 300 km on a side) has grid cells with a size similar to that recommended for urban scale models. States should ordinarily include all monitoring sites considered in the modeled attainment test within the area covered by a grid with size of individual cells comparable to that recommended for urban scale modeling.

Lagrangian model applications. The more pertinent way of expressing this issue is, "how far downwind should I make concentration estimates using a Lagrangian model"? The ability of the most commonly used form of Lagrangian model (i.e., a Gaussian model) to estimate concentrations at distances greater than about 50 km has not been demonstrated. Therefore, use of Gaussian models should be limited to urban scale applications.

A limitation of 50 km introduces a potential inconsistency since, as we just noted, urban scale domains may sometimes be larger than this. There are at least three ways of addressing this discrepancy which may be considered by those implementing a modeling protocol. First, perform the modeling anyway, recognizing that calculated impacts from remote sources will be much less than those closer by. Further, effects of any error will likely be diminished by using the model in a relative sense to adjust *observed* concentrations of primary components. A second alternative is similar to the first, except use a Lagrangian puff model, like CALPUFF (Strimaitis, et al., 1998) to consider effects of large, individual sources located more than 50 km away from the monitor. A third alternative is to limit Gaussian modeling to within 50 km of each monitor and assume that a part of the measured primary particulate matter is irreducible. The irreducible portion may be estimated using available monitored speciated data at background sites. If those implementing the modeling/analysis protocol conclude that none of these three alternatives is acceptable, States may use a numerical grid model (with the chemistry "turned off") to estimate effects of a control strategy on RRF values for primary components of $PM_{2.5}$.

Hot spot analyses. Gaussian or other Lagrangian approaches for estimating 24-hour concentrations at hot spots are likely to be most reliable at distances less than about 15 km. In addition, plumes from individual or small groups of sources are likely to be "smeared out" at further distances. Hence, little is likely to be gained, beyond treatment in a grid model, using a hot spot or plume-in-grid analysis of individual sources of *primary* particulate matter further than ~15 km from receptors of interest. Thus, hot spot analyses generally should be limited to distances within about 15 km of large sources which are of potential concern.

Recommendations. Size of a modeling domain depends on the importance of

measured primary vs. secondary components of measured $PM_{2.5}$ and whether a State's attainment strategy depends on secondary components being reduced. If attainment depends on reducing secondary components, States should use a regional scale modeling domain (> ~ 1000 km on a side).

Urban scale domains (< ~ 300 km on a side) are most suitable if a State's control strategy emphasizes reduction in primary components of $PM_{2.5}$. Maximum domain size may be restricted if Gaussian models are used to compute relative reduction factors for primary components of particulate matter. We present three alternatives for dealing with major sources of primary particulate matter located within an urban scale domain, but more than 50 km from a particular monitoring site. Those implementing the modeling/analysis protocol should consider which, if any, of these alternatives is appropriate for their application.

Hot spot analyses and use of plume in grid algorithms to estimate effects of individual sources on 24-hour concentrations of *primary* particulate matter should be limited to within about 15 km of the source.

13.1.2 What Horizontal Grid Cell Size Is Necessary?

(NOTE TO REVIEWERS: Several of the recommendations in this subsection are based on past experience modeling ozone (e.g., 8-hour and 1-hour daily maxima) rather than particulate matter. We are planning sensitivity tests using models which address $PM_{2.5}$ and its components to confirm or refine our current recommendations regarding horizontal grid cell size for PM-related applications. We solicit information from reviewers which is related to the effect of grid cell size on estimated relative reduction factors for components of particulate matter.)

This discussion, as well as the discussion in Section 13.1.3, applies solely to use of numerical grid models. If a Lagrangian model is used, estimates at the monitor locations should be used to calculate RRF values for primary particulate matter. Centerline concentrations estimated at breathing height (~ 2 m) should be used in hot spot analyses.

In grid model applications, there is some evidence which suggests that size of a horizontal grid cell may be larger if we are concerned with longer averaging times. Analyses supporting this hypothesis have been addressed toward the problem of predicting daily maximum ozone concentrations. However, as we describe shortly, their findings should also be pertinent for particulate matter. Sensitivity tests performed by Kumar, et al. (1994) in the South Coast Air Basin compare hourly ozone predictions obtained with 5 km vs. 10 km vs. 20 km grid cells. Results indicate that use of finer grid cells tends to accentuate highest hourly ozone predictions and increase localized effects of NOx titration during a given hour. However, statistical comparisons with observed hourly ozone data in this heavily monitored area appear comparable with the 5 and 20 km grid cells in this study. Comparisons between hourly ozone predictions obtained with 4 km vs. 12 km grid cells have also been made in an Atlanta study (Haney and

Douglas (1996)). As in Los Angeles, use of smaller (i.e., 4 km) grid cells leads to higher domain wide maximum hourly ozone concentrations. However, when reviewing concentrations at specific sites, Haney and Douglas (1996) found that, for some hours, concentrations obtained with the 12 km grid cells were higher than those obtained with the 4 km cells. Since signs of the differences obtained with 4 km vs. 12 km grid cells vary for different hours, this may suggest that 8-hour daily maximum ozone predictions are less sensitive to the selected grid cell size than 1-hour daily maxima. Recent sensitivity tests comparing relative reduction factors in predicted 8-hour daily maxima near 272 sites in the eastern United States indicate generally small unbiased differences ($\leq .04$, in 95% of the comparisons) using a grid with 12 km vs. 4 km grid cells (LADCO (1999)).

For particulate matter, the averaging time specified in the NAAQS and used for the measurements is 24-hours. Since there appears to be little difference in RRF estimates for ozone with 12 km vs. 4 km grid cells when one goes to 8-hour averaging times, we would expect this finding to hold for secondary components of particulate matter which are computed for even longer averaging times. But what about primary components of particulate matter? Sharpness of spatial gradients in primary particulate matter should be much reduced over 24-hour averaging times. This follows from variability in meteorological conditions as well as in emissions. We would expect lack of atmospheric chemistry to have two counterveiling effects on sensitivity of RRF values for primary particulate matter to changing grid cell size. Because there is no chemistry, we expect spatial gradients in hourly concentrations to be much higher for primary particulate matter than for secondary particulate matter. This follows, because it takes some time (less than 24 hours) for the chemistry to occur. However, we would also expect these differences between primary and secondary components to diminish as averaging times increase. Unlike with secondary pollutants, exact positioning a plume relative to other sources has no effect on concentrations within that plume. Hence we would expect gradients of primary components to flatten out more rapidly as a function of averaging time. We do not know whether the second effect is sufficiently large so that similar conclusions regarding maximum adviseable grid cell size can be applied for primary and secondary components averaged over 24-hours.

The bottom line of the preceding discussion is that we feel comfortable recommending that States may use grid cell sizes as large as 12 km for urban scale applications addressing secondary components of particulate matter. We are less sure about an acceptable upper limit for cell size in applications addressing primary components. We believe it is prudent to assume that cells as small as 4-5 km are needed. Those implementing the modeling/analysis protocol may wish to perform a diagnostic test using a grid model without chemistry to see whether estimated RRF's for primary components are affected if one increases grid cell size to 12 km.

We expect that modeling analyses for nonattainment areas will model such areas using grid cell sizes of 12 km or less. If a regional scale model is applied, most of the domain will likely cover rural/remote areas or locations which are not out of compliance with the NAAQS. Because of the effects of 24-hour averaging times and the lack of close proximity of these locations to a nonattainment area, grid cells as large as 36 km may be used.

Recommendations. **For urban scale applications or in nested portions of regional grids focused on nonattainment areas, States may use grid cells as large as 12 km on a side to estimate RRF's for secondary particulate matter. RRF's for primary particulate matter should be estimated using cells which are 5 km or less on a side. Exceptions may be considered on a case by case basis.**

States may use grid cells as large as 36 km on a side in portions of a regional grid which do not include a nonattainment area.

13.1.3 How Many Layers Should I Consider?

(NOTE TO REVIEWERS: Several of the recommendations in this subsection are based on past experience modeling ozone (e.g., 8-hour and 1-hour daily maxima) rather than particulate matter. We are planning sensitivity tests using models which address $PM_{2.5}$ and its components to confirm or refine our current recommendations regarding number and placement of layers for PM-related applications. We solicit information from reviewers which is related to the effect of number and placement of layers on estimated relative reduction factors for components of particulate matter.)

As described in Section 14.0, the preferred means for generating meteorological data fields for input to numerical grid models and Lagrangian models is to use a dynamic meteorological model with four dimensional data assimilation (FDDA). Such models often consider as many as 20-30 layers. To minimize a number of assumptions needed to interface meteorological and air quality models, it is better to use identical vertical resolution in the air quality and meteorological models. However, application of air quality models with so many layers may not be feasible nor cost effective. In this Section we identify factors affecting number of layers chosen for use in an air quality model, as well as the placement of these layers.

Diagnostic analyses to estimate an appropriate number of layers to use in an air quality model have focused on the problem of predicting ozone accurately. However, with some exceptions (which we will note), we believe the guidance for ozone related applications applies for modeling related to the NAAQS for $PM_{2.5}$ and regional haze. Accuracy of predicted ozone concentrations depends, in part, on how accurately a model is able to characterize dilution of precursors and ozone. This, in turn, depends on how precisely the model can estimate maximum afternoon mixing heights (i.e., the planetary boundary layer). Precision of mixing height estimates is affected by the thickness of the model's layers aloft which are near the anticipated mixing height (Dolwick, et al., (1999)). Because maximum mixing heights may vary on different days and it is necessary to simulate numerous days and locations, model predictions can be influenced by the number of layers considered by the model. Dolwick, et al., (1999) have shown that predictions are not sensitive to the number of layers considered *above* the planetary boundary. Thus, States may assume as few as one layer above the highest conceivable maximum afternoon mixing height with the rest of the layers occurring within the planetary boundary layer. Since wet chemistry occurring within clouds may have an important effect on secondary particulate matter, States should be sure that the "highest conceivable maximum mixing height"

is high enough to include strato-cumulous clouds, often present during summer or late spring.

Earlier, we noted that it may often be necessary to consider a large modeling domain when addressing $PM_{2.5}$ nonattainment and (especially) regional haze-related problems. There undoubtedly will be geographical differences in the maximum mixing height in a regional modeling domain. This poses a problem in specifying an appropriate number and placement of layers for use in a regional model application. Air quality models available as of late 2000 assume a fixed number and placement of layers throughout a modeling domain. This means that a less than optimal specification may be necessary in parts of the domain. We suggest that States give the greatest priority to choosing appropriate layers for areas which are the principal focus of a modeled demonstration. For applications related to the NAAQS, this means the major nonattainment area(s) being addressed. In cases where there are numerous areas which are the focus of an analysis (e.g., as in reasonable progress demonstrations covering many Class I areas), specify a set of layers which appears most appropriate for a majority of the areas.

Placement of layers within the planetary boundary layer is also an important issue. For practical reasons, it is best to have an air quality model's layer placement coincide with layers considered in the meteorological model used to generate meteorological inputs. So the placement issue really is, which ones of the boundaries between the meteorological model's layers should I match with the boundaries between layers used in the air quality model? Modeled concentrations used to compute component-specific RRF values are based on 24-hour estimates, and nigttime periods may often be characterized by significant wind shear near the ground. Further, nearly all emissions are likely to occur within ~300 m of the ground. Thus, we suggest that States give greatest attention to matching layers in the meteorological model within the first 300 m or so of the ground. The layer adjacent to the ground should generally be $\leq$ 50 m deep. In general, layers below the mixing height should not be too thick, or large, unrealistic step increases in mixing may occur. States should try to avoid using layers within the planetary boundary layer thicker than about 200-300 meters.

Based on recent sensitivity studies performed for ozone (Dolwick, et al. (1999) and LADCO (1999)), it appears as though as few as 7-9 layers (including one above the planetary boundary layer) may suffice in a modeling study if care is taken in specifying placement of these layers. Because the domain needs to extend high enough to include cloud tops and it is desirable to have a shallow layers near the surface and avoid having thick layers further aloft, we raise this minimum number slightly to 9-11 layers. Prior to modeling, we recommend that States review available meteorological measurements aloft to get a sense of where the maximum afternoon mixing height is likely to be on days which might be modeled. This information should be used to avoid having layers below the planetary boundary layer which are so thick that they could lead to large, discrete changes in mixing within very short periods. As we noted previously, air quality models with which we are familiar all require that the number of layers considered in coarse and fine portions of a nested regional grid be identical.

Recommendations. An air quality model may consider fewer layers than commonly considered in a meteorological model. However, boundaries between layers used in

an air quality model should coincide with selected boundaries in the meteorological model.

Care should be taken to configure an air quality model's layers so that greatest vertical detail is provided in the lowest 300 meters or so of the atmosphere. The surface layer considered in the model should generally be no more than 50 meters deep. States should also try to define the maximum afternoon mixing height precisely. Except in a few instances where the the maximum mixing height is very large, States should try to specify a model's layer configuration so that no layer beneath the mixing height is more than about 300 meters thick. The goal is to avoid large, stepwise increases in mixing over very short time periods.

The maximum height of the domain should be great enough to include tops of strato-cumulous clouds. The minimum number of layers chosen depends on the meteorological conditions and characteristics of the area to be simulated. To meet the preceding criteria, it will most likely be necessary to use at least 9-11 layers within the planetary boundary layer and 1-2 layers above it.

13.1.4 Should I Consider Plume In Grid Modeling?

Imbedding a plume-in-grid model within a grid model application temporarily "isolates" emissions of one or more large sources from those emanating from many nearby smaller sources. The reason for doing this to avoid having to assume that emissions from the large source(s) disperse instantaneously within a relatively large grid cell. As of late 2000, we are unaware of any air quality grid model which has a plume-in-grid capability that has been tested successfully for use with particulate matter. However, plume-in-grid modeling offers a potentially cost-effective alternative to using very fine grid cells for treating emissions from major sources in the immediate vicinity . It is likely that one or more models with plume-in-grid capabilities will have been developed and tested for particulate matter by the time States need to develop their SIP revisions to meet goals for $PM_{2.5}$ and regional haze.

As noted in Section 13.1.1, we think it is unlikely that a plume-in-grid calculation will change a relative reduction factor estimated for primary particulate matter, unless there is a major point source of primary particulate emissions within about 15 km of a monitor. To reiterate, the reason for this is the likelihood that the plume will meander during a day such that it gets "smeared out" over a 24-hour period.

Plume-in-grid treatment for sources of SO_2 and/or NOx may be justified over a longer distance, however. This follows, because chemistry required for the formation of ozone and secondary particulate matter is affected by the speed with which a major source's NOx or SO_2 plume mixes with the surrounding environment. For example, in their review/analysis of plume data collected during an intensive field study near Nashville, TN in 1995, Gillani, et al. (1998) show that it takes anywhere from about 30 - 40 km for a typical large plume and as much as 100 km from an extremely large source to become sufficiently diluted so that its chemistry is not

affected by limited dilution. Thus, estimated production of secondary particulate matter both within and outside of a plume could be affected by a model's failure to properly consider limited dilution of a major source's plume. Even though these limits occur over timeframes which are considerably less than the 24 hour period which serves as the basis for our $PM_{2.5}$ NAAQS and visibility calculations, we cannot rule out the possibility that failure to consider this delay in mixing from nearby sources could affect relative reduction factors which are estimated for SO_4, NO_3 or secondary OC.

Largest sources of NOx or SO_2 are unlikely to occur in the middle of urban nonattainment areas. That is, they are more likely to occur in the "regional portion" of a regional grid (i.e., with 36 km cells). More moderately sized sources could conceivably occur within an urban domain (i.e., with 12 km grid cells). Therefore, a reasonable rule of thumb for plume in grid applications is to limit them to sources which occur within ~3 grid cells of a monitoring site.

<u>Recommendations.</u> Plume in grid calculations for sources of primary particulate matter are unlikely to be necessary. States may wish to perform a diagnostic test including a plume in grid calculation if a major source of primary particulate matter is located within 15 km of a monitor recording a violation of a NAAQS for $PM_{2.5}$ and the control strategy is focused on reducing primary particulate matter.

States should consider performing a diagnostic test to determine whether relative reduction factors for SO_4, NO_3 or secondary OC are affected by applying a plume in grid algorithm to major point sources of SO_2 or NOx located within ~3 grid cells of a monitor.

13.1.5 What Else Should I Consider In Choosing Finely And Coarsely Resolved Portions Of Nested Regional Models?

Coarse Grid Domain. Size of a coarse grid domain should be consistent with the chemical/physical lifetimes of pollutants to be modeled. It should also reflect the purpose for which regional modeling is undertaken. For example, if a regional analysis is performed to assess effects of a regional strategy simultaneously for a number of nonattainment areas, the domain needs to be larger than if only a limited number of nearby areas were the focus of the study.

Lifetimes vary for ozone, secondary particulate matter and their precursors. Lifetime for NOx (i.e., NO + NO_2) may be less than a day. Regional analyses performed in the U.S. to date suggest that lifetimes for sulfates and nitrates are two days or less (Dennis, 1994). Sources of VOC are believed to be ubiquitous, due to natural emissions. Many of these natural emissions are relatively reactive, so that multi-day transport of stable species of VOC or radical products resulting from oxidation of more reactive species may not be a critical factor for selecting size of a domain for modeling or secondary particulate matter. Transport of primary OC, such as results from vegetative burning, may be important over several days. Lifetime for ozone is notoriously difficult to estimate due to the recycling of this compound with free radicals, concentrations of

oxidized species of nitrogen and emissions of fresh NOx and VOC precursors which occur in transit. Given information about the lifetime of nitrates however, it is probably safe to assume a lifetime for ozone which is on the order of 2-3 days. The foregoing information suggests that, ideally, the size of a regional modeling domain should be large enough so that emissions occurring two days prior to the beginning of daylight on a modeled day of interest are included within the domain. Thus, we suggest States focus on their receptor areas of interest, perform some screening analyses with trajectory models to ensure that major source areas within two days' travel time are included in the domain.

Fine Grid Domain. Size of the fine grid domain should be influenced by several factors: (1) proximity of receptor sites to major sources of $PM_{2.5}$ and its precursors; (2) presence of topographical features which appear to affect observed air quality, and (3) desire to limit resource intensive efforts needed to use numerical models on a fine scale. The last factor is an important concern for use of nested regional models. Size of a fine grid domain could be smaller than that recommended for an urban scale analysis. This follows, since the coarse domain is available to estimate impacts of sources located at intermediate distances from the receptor area, whereas this information is not available for an isolated urban scale analysis. The issue of how far to extend a fine scale grid is one which may need to be resolved on a case by case basis. We recommend that States examine the issue using diagnostic sensitivity tests (see Section 16.0). For consistency with the modeled attainment tests, we recommend that the fine grid should initially extend 15 km beyond any receptor of interest.

> **<u>Recommendations</u>. Size of a coarse grid should be large enough to include potentially important sources located two days' travel time from receptor sites of interest. Applications which need to consider numerous receptor sites located some distance apart therefore need to use larger domains than do applications focusing on receptors in close proximity to one another. Extent of a fine grid depends on the number of receptor sites. States should perform diagnostic analyses to ascertain how far a finely resolved grid needs to extend. As a starting assumption, we recommend extending the finely resolved grid sufficiently so that it extends at least 15 km beyond all monitoring sites considered in the modeled attainment tests.**

13.2 How Do I Decide What Domain Size And Resolution Are Appropriate To Address The Reasonable Progress Goals For Reducing Regional Haze?

The reasonable progress test focuses on Class I areas. Most of these are in remote or rural locations. Regional haze is not likely to be dominated by local sources. Further, light extinction is more likely to be dominated by secondary particulate matter, due to the latter's efficiency of scattering light. All these attributes indicate that it will be necessary to model a regional scale domain for regional haze related applications. Thus, a domain at least 1000 km on a side, in which pollutants emitted at least two days prior to a day of interest remain within the grid is needed. Regional modeling may be performed by or for regional planning organizations who wish to consider effects of a strategy on many Class I areas simultaneously. Thus, for applications in the contiguous United States, it may be most efficient to consider two

superdomains—one covering the western (mostly low humidity) half of the nation, and a second covering the eastern (mostly high humidity) half.

Because of the remoteness of Class I areas, grid cell sizes up to 36 km on a side should suffice for regional haze-related modeling. States may perform diagnostic tests using plume-in-grid analyses, as described in Section 13.1.4. Guidance on vertical resolution presented in Section 13.1.3 is also applicable for regional haze-related applications.

<u>Recommendations</u>. A regional domain is needed for model applications related to regional haze. If an analysis is intended to address effects of a strategy on many Class I areas simultaneously, using two superdomains covering the contiguous United States may be appropriate.

Horizontal grid cell sizes up to 36 km on a side may be used. States may use a plume in grid algorithm, if they choose, to assess effects of major point sources of SO_2 and/or NOx located within ~3 (36 km) grid cells of a monitor in a Class I area.

Modeling should consider at least 9-11 layers beneath a mixing height associated with the highest strato-cumulous clouds and 1-2 layers above this. Layer boundaries should coincide with boundaries used in meteorological models used to generate input to the air quality model. Greatest detail is needed within ~300 m of the ground, and the surface layer should be no more than ~50 m deep. No layer should be more than about 300 m thick.

14.0 How Do I Produce Meteorological And Air Quality Inputs Needed By An Air Quality Model?

After episodes are selected for modeling, corresponding meteorological inputs need to be generated for use in an air quality model. Although the resulting inputs remain constant, they can affect outcomes of a number of the modeling outputs we have identified in Section 5.1 as potentially useful for weight of evidence determinations. They may also potentially affect relative reduction factors used in the attainment and reasonable progress tests. In contrast to meteorological data, air quality inputs may change between times corresponding to "current" and "future" emissions used in the modeled tests. This presents a potential problem which needs to be addressed.

In this Section, we describe two approaches for generating meteorological inputs to numerical air quality grid models, and identify advantages/disadvantages associated with each. We note that using dynamic meteorological models with output "nudged" by observations is usually the preferred approach for generating needed meteorological data. For some applications, use of these models for horizontal grid cells smaller than 12 km may present practical problems. We identify ways to diminish these, if they occur. It is important to quality assure meteorological inputs prior to their being used in an air quality model. We next discuss how these inputs can be evaluated. We conclude our discussion of meteorological inputs by recommending procedures for deriving them for use in Lagrangian (e.g., Gaussian) models. Meteorological needs of Lagrangian models depend on how these models are to be used. We first discuss these needs when the models are used to compute relative reduction factors for primary particulate matter. Then we address needs for meteorological inputs when Lagrangian models are used for hot spot analyses. We next identify the role of air quality inputs as initial and boundary conditions for an air quality simulation, and note ways to reduce limitations to the simulation resulting from sparseness of these data. We conclude Section 14.0 by recommending procedures for estimating background concentrations when Lagrangian models are used.

14.1 What Approaches Exist To Generate Meteorological Input To Grid Models?

Two approaches have been widely used to generate meteorological data needed in gridded air quality models. The first of these (diagnostic models) relies primarily on observed data and introduces some additional constraints on wind flow due to terrain features. Observed surface temperatures and sounding data are used to develop other information needed to characterize mixing.

Most frequently used diagnostic wind models are described by Douglas, et al. (1990) and by Scire, et al., (1998). The main advantage of diagnostic models is that they are relatively easy and inexpensive to apply. Further, they make maximum use of wind observations. There are several disadvantages, however. First, there are seldom enough observations to adequately define a windfield, particularly aloft. Much of the input to the air quality model is derived through interpolation or subjective methods. Further, mixing heights need to be derived through external estimates. In addition, current (2000) generation air quality models feature a fixed

vertical cell structure. Thus, it is necessary to somehow estimate vertical exchange rates between these cells. Because of the sparseness of observations in many areas, we do not encourage use of diagnostic models for generating inputs to regional scale air quality model applications. Finally, meteorological estimates derived with a diagnostic model are not necessarily physically consistent with one another. In the atmosphere, there is a physical dependency existing between temperature, pressure and windfields. This interdependency is not extensively accounted for in diagnostic models, and the extent to which it is considered depends on the expertise of those applying the model. Nevertheless, if ambient concentrations of a pollutant (e.g., as for one or more components of primary $PM_{2.5}$) are believed to be primarily affected by winds and urban scale source/receptor orientation, the disadvantages are not serious enough to preclude use of diagnostic models.

The second approach for generating needed meteorological data is to use dynamic meteorological models with four dimensional data assimilation (FDDA). These models attempt to characterize theoretical relationships between meteorological variables and topographical/terrain characteristics. Use is made of relatively sparse observations aloft to help steer (i.e., "nudge") solutions so that they do not diverge from observed meteorological fields. Wind observations aloft are typically used for this purpose. See Seaman (1997) for a further summary of the attributes of dynamic meteorological models. The MM5 (Grell, et al., 1994) and Seaman and Stauffer, 1996) and the RAMS (Pielke, et al., 1992 and Lyons, et al., 1995) models are among those which have been most widely used with numerical air quality models which have the capability of considering precipitation events (i.e., potentially important for making PM predictions). The major advantage of dynamic meteorological models is that they provide a way of characterizing meteorological conditions consistent with theory, terrain and each other at times and locations where observations do not exist. Further, their vertical structure is consistent with that of current (2000) generation air quality models. Disadvantages have been large required computer resources and considerable expertise needed to apply the approach. Recent advances in computer technology have resulted in increased use of dynamic meteorological models for air pollution applications. The MM5 model is used as the default approach with the CMAQ model in MODELS3. States need to consider compatibility between candidate meteorological models and the air quality model(s) chosen for use. We believe that use of dynamic meteorological models with FDDA is generally the preferable approach for generating meteorological inputs to air quality models for secondary particulate matter and ozone.

Although improvements in computers have made increased use of dynamic meteorological models possible, we have found that data storage requirements and CPU time increase dramatically as the horizontal grid cell size required of the meteorological model becomes finer. For example, the CPU time needed to generate meteorological data resolved to 12 x 12 km grid cells is considerably greater than the expected factor of "9" increase in that needed to process meteorology for a domain with 36 x 36 km grid cells. This suggests that States may need to limit the spatial extent of areas and the number of episodes for which dynamic meteorological models are used to process meteorological data for grids with horizontal resolution <12 km. Generally, a finely resolved meteorological field needs to extend about 3 grid cells beyond the bounds of the fine scale grid used to make air quality predictions. For example,

if 4 km grid cells were used in the fine portion of a nested regional air quality model, meteorological fields at this detail would need to extend 12 km beyond the bounds of the 4 km grid used for air quality predictions.

<u>Recommendations</u>. States should ordinarily use a peer reviewed dynamic meteorological model with four dimensional data assimilation as the means for generating meteorological inputs to gridded air quality models. "Peer reviewed" means the models basic formulation has been published in the peer reviewed literature and subsequent changes have undergone third-party beta testing and scientific review. Peer reviewed diagnostic models may be used on a case by case basis. Grid cell size used in dynamic models should be chosen considering factors discussed in Section 14.0.

14.2 How Do I Deal With Data Management And Computer-related Constraints When Applying Dynamic Meteorological Models?

States should ordinarily use dynamic meteorological models resolved to the same level as desired for making air quality predictions. Occasionally, this may not be feasible, or may lead to poor performance of the dynamic model. In this Section, we identify possible means for reducing one or both of these problems. The methods we discuss may increase the risk of discontinuities at the bounds of a finely resolved grid. This may result when one matches up surface (or, more properly, 10 m) observations used in diagnostic models with layer average estimates provided by dynamic models. One means for diminishing potential mass conservation problems might by to derive layer average estimates by applying an appropriate power law or similarity theory to the meteorological observations. These should be checked and corrected to the extent possible before proceeding.

The first approach is to use available results from dynamic models on the next greatest coarse scale (i.e., 36 km for a desired 12 km estimate, 12 km for a desired 4 km estimate) to interpolate more finely resolved fields. An objective approach like bilinear interpolation could be used (U.S. EPA, 1991). This approach would be particularly useful if the major reason for desiring finely resolved meteorological estimates is related to a need to resolve emission estimates more finely. For example, in the case of $PM_{2.5}$, fine grid cells may be needed to most accurately characterize gradients in emissions of primary particulate matter.

A second approach for circumventing major resource requirements needed to apply dynamic models for finely resolved grids considers topographic information (e.g., presence of land/water interfaces) and measured meteorological data to refine coarser fields generated by a dynamic model. This second approach may be preferred if the major reason for desiring finely resolved meteorological inputs has to do with perceived importance of mesoscale features which cannot be adequately considered through an objective interpolation procedure. In essence, the second approach is to apply a diagnostic wind model to the wind field generated by the more coarsely resolved dynamic model.

Finally, consequences of using coarse grid cells (e.g., 12 km when 4 km might be more desirable) can be reduced by specifying a land use for each cell that corresponds to usage near the major portion of emissions within a cell. This approach is most applicable at land/water interfaces. By assuming the cell is entirely "land", vertical dispersion of fresh emissions is likely to be better characterized. This might also result in a better characterization of subsequent transport of coastal emissions over adjacent large bodies of water.

<u>Recommendations</u>. Prohibitive computer-related constraints associated with applying a dynamic meteorological model to derive a finely resolved (4-12 km) set of meteorological data can be addressed in one of two ways.

1. Interpolate more coarsely resolved data using objective analysis.

2. Apply a diagnostic wind model using "observations" generated by the dynamic meteorological model for a coarser grid. Assume other variables remain the same as for the coarser grid.

Consequences of using coarser than desired grid cells may be reduced by assigning a land use factor for each surface cell which corresponds to the location of most emissions within the cell (e.g., at cells including an interface between land and a large body of water).

14.3 How Do I Quality Assure Results Generated By A Meteorological Model?

There are several ways to evaluate performance of a meteorological model. Although it is desirable to evaluate meteorological inputs *before* air quality predictions are made, some of the means available for evaluating the meteorological model's predictions must wait until the air quality model is run. Important meteorological outputs warranting scrutiny include wind velocity patterns, mixing heights (e.g., estimated by noting the vertical layer at which vertical diffusivity (K_v) is suppressed), temperature, pressure, water vapor and cloud cover. Methods for evaluating output from a meteorological model include comparison with selected upper air measurements, derivation of trajectories, use of computer graphics, use of non-reactive tracers, comparing results obtained with different models, use of dimensionless parameters, comparing spatial patterns of observed and predicted components of $PM_{2.5}$ and use of process analysis. Each of these is briefly described in the following paragraphs.

Comparison with upper air observations. This can be done by excluding selected upper air observations from use in four dimensional data assimilation (FDDA) so that they can be used to assess model performance. Wind velocity, temperature, pressure and water vapor are important variables to compare. If aloft measurements are available at more than one altitude, they can provide a means for evaluating how well a model characterizes vertical exchange in the lowest few layers. Generally, routine data bases (e.g., widely separated soundings taken twice per day) are needed to support FDDA. Collection of continuous vertical meteorological profiles

at a site is becoming more common with increasing use of radar profilers. Resulting data should be helpful for checking performance of a dynamic model. Nevertheless, the data base is insufficient to rely solely on these data to evaluate model performance. In Section 6.0, we noted that it is desirable to increase measurements aloft. One reason for doing this is to provide better means for evaluating performance of meteorological models.

Derivation of trajectories. A State could select several locations in the grid and use trajectory models such as HY-SPLIT (NOAA, 1999) to derive back- or forward-trajectories from the hourly wind fields generated by a meteorological model. If surface trajectories were limited to daylight hours, the computed trajectory could be compared with observed surface air quality observations. Of course, this means of evaluation only works if air quality observations are available on an hourly, or at most, 3-hourly basis. Thus, it is most appropriate to use observed ozone (or some other hourly observations) to help evaluate performance of the chosen meteorological model. If the timing of high ozone observed along the path of the trajectories is consistent with expectations, given the configuration of sources, this is an indicator that the meteorological model is performing adequately. A State could also derive daytime surface trajectories using *observed* wind data. These trajectories could also be compared with air quality patterns. Comparing the two sets of trajectories with observed air quality patterns makes it possible to assess whether the meteorological model increases the skill with which ozone plumes are oriented.

Use of computer graphics. Examining wind vectors for apparent discontinuities is possible using graphics. It is also possible to construct difference diagrams between observed and predicted temperatures and winds. Locations where agreement is poor may suggest areas needing more finely resolved estimates. Geographical orientation between areas of poor agreement and locations of major sources or observed poor air quality may be plotted to judge potential significance of any disagreement.

Simulation of inert tracers. This approach is to assume a uniform concentration field (e.g., 10 ppb) of an inert tracer in an air quality model with grid cells identical to the horizontal and vertical size of the cells used in the meteorological model (this may be feasible, since it is unnecessary to consider atmospheric chemistry, deposition or fresh emissions). Identical, constant boundary conditions should also be assumed. In theory, the concentration field should remain uniform, and there should be no systematic drift in the mass of material remaining within the grid. Predicted concentrations of the tracer can then be examined to see whether there are major discontinuities in the concentration field or problems with mass balance. If there are, this may suggest a problem with the meteorological model or with the ability of the air quality model to consider divergence/convergence predicted with the meteorological model.

Compare results obtained with different models. This approach is to compare results from two different models for a subset of days being considered. For example, MM5 and RAMS results could be compared to note differences in predicted surface temperatures as well as wind velocities at the surface and aloft. Reasons for major differences would then need to be diagnosed.

Compare estimated divergence or dimensionless parameters with expected ranges. Calculations can be made in selected portions of the grid to see whether they appear reasonable.

Compare spatial patterns of air quality predicted with a grid model with observed patterns on the days of interest. If the predictions are systematically skewed from the observations, this could suggest a problem with the meteorological outputs generated by the meteorological model.

Compare meteorological outputs from the meteorological model vs. those generated from the air quality model. This comparison entails looking at a few key meteorological variables (e.g., winds, temperature profiles). The purpose of this comparison is to ascertain whether the source of a meteorological "problem" is the meteorological model itself or the data processing used to convert the meteorological model's outputs for use in the air quality model.

Use process analysis. Process analysis applies to the output generated by an air quality model. It is described by Jeffries (1997) and by Lo and Jeffries (1997). Its use with air quality models is noted in Section 16.0. Process analysis determines the relative importance of different chemical or physical factors as contributors to predicted air quality concentrations. If process analysis suggests that a variable influenced by meteorological inputs, such as vertical exchange (i.e., vertical diffusivity), plays a large, unanticipated role leading to a high or a low prediction of one or more components of $PM_{2.5}$, this might warrant a closer examination of what led to such a prediction.

<u>Recommendations</u>. To the extent possible, States should quality assure results from meteorological models prior to using them in the intended air quality model. States should select a mix of approaches for evaluating meteorological inputs to an air quality model on a case by case basis. Candidate approaches include:

1. **comparison with upper air measurements "held back" from use in FDDA;**

2. **comparison of calculated trajectories with observed air quality patterns;**

3. **use of computer graphics to discern spatial discrepancies;**

4. **simulation of inert tracers to identify discontinuities or mass balance problems;**

5. **comparing results obtained with different meteorological models;**

6. **calculating and comparing divergence and/or dimensionless parameters and comparing these with expected ranges;**

7. **comparing spatial air quality patterns obtained with a grid model vs. observed patterns;**

8. comparing outputs from a meteorological model for key variables vs. estimates for the same variables used in the air quality model, and

9. using process analysis to flag contributions made to unexpected concentrations of $PM_{2.5}$ components by meteorological factors.

14.4 What Are Some Past Applications Of Dynamic Meteorological Models?

Table 14.1 lists some recent air quality modeling applications using the two most widely available dynamic meteorological models. Choice of a meteorological model may be influenced by compatibility with a chosen air quality model, as well as by past experience of those applying the air quality model. The listing in Table 14.1 is not comprehensive. Inclusion on the list does not necessarily imply an endorsement for a specific application. Exclusion does not necessarily imply that an approach is inappropriate for a specific application. States should consider using methods such as those in Section 14.3 to determine whether the output generated by a meteorological model is adequate for use in a specific application.

Table 14.1. Some Past Applications Of Dynamic Meteorological Models

Meteorological Model	References Describing Model Performance	Sponsors (Applications)
MM5	**Seaman, et al., (1995), (1996b)** **Tesche and McNally, (1993a), (1993b)**	**CARB (San Joaquin Valley, South Coast Air Basin, CA)**
RAMS	**Tesche and McNally, (1993c), (1993d), (1993e), (1993f)**	**LADCO (eastern U.S., with emphasis on Lake Michigan States)**

14.5 How Do I Generate Meteorological Input For Lagrangian Models?

States should review Section 9.3 in 40CFR Part 51, Appendix W, which discusses choice of meteorological data to use in Lagrangian model applications and recommends preprocessor programs to convert these data into formats required by the air quality models. Appendix W suggests using one of two data bases: (a) at least one year of meteorological data collected at an individual source to be modeled, or (b) 5 years of data collected at nearby National Weather Service or other valid monitoring sites. Because of the nature of the NAAQS for $PM_{2.5}$, and the recommended modeled attainment and hot spot tests, we suggest using at least at least 3 years of meteorological data. Preferably, the meteorological data should cover the 3-year period upon which the modeled attainment test's "current" design value is based. If a Lagrangian model is used to help estimate relative reduction factors in the modeled test for reasonable progress

reducing regional haze, meteorological data for a 5-year period is needed. Preferably, these data should cover the 5-year base period specified in the regional haze rule.

<u>Recommendations</u>. States should review information presented in Section 9.3 of Appendix W to 40 CFR Part 51 prior to selecting meteorological data for use in Lagrangian (e.g., Gaussian) models. For applications related to the NAAQS for $PM_{2.5}$ at least 3 years of data are needed. For applications related to assessing reasonable progress reducing regional haze, 5 years of data are preferred.

14.6 How Do I Address An Air Quality Model's Need For Air Quality Inputs?

Grid model applications. Air quality inputs are needed in gridded air quality models for two purposes: to specify initial conditions, and to specify boundary conditions. There is no satisfactory way to specify initial conditions in every grid cell. Thus, we recommend beginning a simulation at least 2 days prior to a period of interest for urban scale applications, and 3 days prior to periods of interest for regional applications to diminish importance of arbitrary assumptions about initial conditions.

Boundary conditions can be specified in several ways. One way is to nest the area of interest within a much larger domain. This approach is exemplified by using nested regional models, as described previously. The need to diminish importance of boundary conditions is why we recommended in Section 13.0 that States use a large regional domain, with the upwind bound 2 or more days' travel time from the area(s) which is the focus of an analysis. If it is not practical to use a nested regional modeling approach, a second approach is to use a large single domain in an urban scale analysis. The domain should be approximately symmetrical about the major local sources affecting local monitoring sites of interest , and should be large enough so that emissions occurring in the center of the domain just before sunrise remain within the domain until the end of the same calendar day. If recirculation is thought to be part of the problem, the domain size would need to be extended to be able to consider it. Use of a large, single domain requires one to make use of monitored data and interpolation to estimate boundary conditions. This approach begs the question about what to assume for *future* boundary conditions. It works best where boundary conditions are low and are expected to remain so.

Lagrangian applications. If a State uses a Lagrangian (e.g., Gaussian) air quality model to estimate primary components of particulate matter, assumptions need to be made about "background" air quality concentrations. Estimates made with the Lagrangian model should get superimposed on this background. Definition of "background" depends on the scope of the Lagrangian model application. If the model is applied to estimate RRF values for primary components of $PM_{2.5}$, a State will need to estimate effects of multiple sources of primary particulate matter and, as discussed in Section 13.1.1, the model will need to be applied over an urban area. In cases like this, "background" should be derived from regional modeled estimates made outside the urban area or from monitored observations at remote/rural sites outside of the urban area (i.e., we are using the model itself to estimate effects of nearby sources). To estimate background, a State should use only the observations or regional modeled estimates for the same

components being modeled with the Lagrangian model. If background is estimated using monitored data, future background concentrations should remain unchanged.

For hot spot analyses, Lagrangian model applications will often be limited to one or a very small number of sources. Thus, "background" will need to consider effects of other sources of primary particulate matter located within the urban area but which are not modeled. The best procedure is to exclude the "hot spot" source(s) from the inventory in a multi-source model application to estimate a "background" concentration at the site(s) of the source(s) subject to the hot spot analysis. If this is not practical, States may use a spatially averaged set of gridded surface estimates or spatially averaged monitored concentrations obtained within 50 km of the source. If the distance from the source to the bounds of the urban nonattainment area is less than 50 km, averages within this smaller distance may be used instead. This procedure diminishes the effect of "double counting" source(s) for which the hot spot analysis is performed. If monitored data are used to estimate background, the default assumption for future background concentrations is that it remains unchanged.

<u>Recommendations</u>. Simulations should begin at least two days prior to the period of interest for gridded urban applications and three days for gridded regional applications. Use of nested regional models is the preferred approach for addressing boundary conditions for grid model applications. Where such an approach is not feasible, States should consider a single domain large enough to ensure that emissions occurring in the center of the domain just before sunrise remain within the domain until the end of the same calendar day or that next-day recirculation (if important) can be considered.

"Background" concentrations are needed for Lagrangian model applications. If a Lagrangian model is used in the modeled attainment test, "background" should be estimated using 24-hr air quality observations or regional model estimates obtained just outside the domain considered with the multi-source Lagrangian (e.g., Gaussian) model. For hot spot analyses, States should exclude the "hot spot" source from the inventory used in a multi-source model and use the multi-source model to calculate "background" at the site of a hot spot. Alternatively, "background" for hot spot analyses may be estimated by spatially averaging 24-hour gridded model estimates or air quality observations at all locations within the urban area or within 50 km of the source, whichever is less.

(This page is left blank intentionally)

15.0 How Do I Produce Emission Inputs Needed For An Air Quality Model?

(NOTE to reviewers: As we describe in this Section, there are a number of ongoing efforts to improve emissions models and related inventories for PM and its precursors. The status of emissions models to deal with $PM_{2.5}$ is still very much in a state of flux. We anticipate that much current uncertainty inherent in ongoing emission model development and documentation will be resolved prior to the time the guidance herein is finalized.)

In this section, we first summarize emissions information needed to obtain the modeled estimates used to assess model performance and to apply the modeled tests we recommend in Sections 3.0 and 4.0 for attainment of the NAAQS and reasonable progress reducing regional haze. We then discuss how to obtain the needed emissions estimates.

15.1 What Emissions Information Is Needed?

Inventories are needed for as many as three time periods: (1) a year which corresponds closely to the air quality data base used to evaluate model performance (the "performance-related inventory"); (2) a year having emissions corresponding to the air quality observations (i.e., the design value) used in the modeled attainment test or in the modeled test for reasonable progress (i.e., the 2000-2004 base period), called the "current inventory", and (3) a year which is two years prior to the deadline for meeting the NAAQS or initial reasonable progress goal ("future inventory"). If the aerometric data base used to evaluate model performance is drawn from a period corresponding to that represented by the "current inventory", the emissions data base used to derive the "performance-related inventory" and "current inventory" is one in the same.

The specificity with which emissions estimates need to be made depends on whether a State is evaluating model performance or applying one of the modeled tests for attainment or reasonable progress. For model performance evaluations, States should estimate day-specific inventories to the extent possible. We also recommend that this be done when applying the modeled attainment test for the 24-hour NAAQS. In contrast, month-specific average inventory estimates are adequate for use in the modeled annual attainment and reasonable progress tests. If a State believes that there is not likely to be much variability in emissions from month to month within a calendar quarter, quarterly average inventory estimates suffice. Air quality models require hourly emissions estimates for each location (e.g., grid cell) considered. Hourly, location-specific emissions estimates may be obtained using emissions models, as described in Section 15.4.

To provide emissions information needed for performance evaluation and the modeled tests, States need to estimate emissions for the following in their inventories: SO_2, NOx, NH_3, primary organic particulate matter, primary inorganic particulate matter (speciated), VOC (speciated) and CO. For NAAQS-related applications, it is only necessary to estimate emissions of primary organic and inorganic particulate matter having $\leq$ 2.5 μm aerodynamic diameters. For visibility-related applications, it is also necessary to estimate emissions of primary particulate matter with aerodynamic diameters $\leq$ 10 μm but > 2.5 μm (i.e., "coarse" particulate matter).

Emissions need to be segregated into those emanating from numerous small stationary (i.e., "area") sources, roadway and non-road emissions from mobile sources, emissions from large sources which need to be treated individually ("point" sources) and emissions from natural, biogenic sources. Each of these types of emissions require certain associated pieces of information to be considered adequately in models.

Point sources. States should identify emissions according to the source's SCC process(es). Geographical location of the source, stack parameters affecting estimating plume height and the operating schedule should also be estimated. To estimate a future inventory, States should make note of which SIC category is appropriate for each point source, current source activity levels and emission factors. Future emissions are the product of future estimates for the emission factors and activity levels.

Area and non-road mobile sources. SCC process level emissions, typical operating schedules, activity levels, an identified surrogate for apportioning emissions geographically and current emissions factors are needed for each area source category and non-road mobile sources.

Mobile sources. States need to estimate a level of vehicle activity (such as vehicle miles traveled, "VMT"), develop a means for geographically apportioning this activity and be able to estimate vehicle speeds.

Biogenic sources. As we describe in Section 15.4, States may use default estimates for land use and biogenic emission factors to estimate emissions from biogenic sources.

The most reliable emissions estimates for a nonattainment area or for the vicinity of a Class I area of particular interest are likely to be obtained if a State is able to provide area-specific estimates of activity levels and/or emissions factors based on area-specific measurements or data, rather than relying on national or even Statewide data bases/procedures. Deriving emissions estimates based on locally specific information and data is often referred to as the "bottom-up" approach for estimating emissions. We recommend this approach whenever it is feasible. Examples include estimating VMT using locally specific transportation demand models, using Metropolitan Planning Organization data to estimate activity levels for specific area source categories and using continuous emissions monitoring (CEM) data to estimate point source emissions.

> **<u>Recommendations.</u> States need to develop inventories for three time periods: a year corresponding to the data base used to evaluate model performance, a year corresponding to "current air quality" used in the modeled tests identified in Sections 3.0 and 4.0 and a future year occurring two years before the identified deadline for meeting an air quality goal. We recommend day-specific emissions estimates for performance evaluations and for the 24-hr NAAQS. Monthly average emissions estimates, and sometimes quarterly averages, suffice for modeled tests related to the annual NAAQS and visibility.**

States should develop inventories for SO_2, NOx, NH_3, primary organic and inorganic particulate matter $\leq$ 2.5 µm and for "coarse" particulate matter (visibility-related applications), VOC and CO. Speciated emissions estimates are needed for primary particulate matter and for VOC.

Emissions should be distinguished according to whether they are from individual "point" sources, stationary area source categories, off-road mobile sources, onroad vehicle sources and biogenic sources.

Emission estimates are a product of emission factors times activity levels. We recommend that States use locally available data, and measurements to estimate emission factors and activity levels whenever feasible.

As noted in Sections 11.0, 12.0 and 13.0, we believe that many model applications will need to cover large geographical areas over extended periods of time. For much of a modeling domain it may not be feasible to use a "bottom-up" approach to estimate emissions. In Sections 15.2 - 15.6, we provide recommendations for using national data bases and guidance for estimating emissions. Deriving local emissions from national (or Statewide) data bases and guidance is often called the "top-down" approach. We identify information which needs to be compiled on a State- and countywide basis. We then identify emissions models which can be used to convert countywide estimates to the inputs needed by air quality models. We next describe several approaches useful for quality assuring estimates obtained as the first two steps proceed. We then present a short discussion of emission projection methods. We conclude Section 15.0 by noting some emissions-related assumptions which are appropriate when using Lagrangian models for hot spot analyses.

While our discussions focus on air quality model needs, there is more extensive guidance available on emissions inventory requirements and preparation (U.S. EPA, 1999d). The U.S. EPA also maintains an internet website which serves as a good starting point for States and others seeking information about estimating emissions related to $PM_{2.5}$ or regional haze-related problems (see ***http://www.epa.gov/ttn/chief/eiip/pm25inventory/pm25home.htm***). Additional guidance has been developed to prepare emission inventories. Much of this guidance has resulted from a joint State/local agency and U.S. EPA effort called the Emission Inventory Improvement Program (EIIP). A series of ten documents has been issued as a result of the EIIP. The most pertinent of these are listed below:

- Volume I: Introduction and Use of EIIP Guidance for Emissions Inventory Development (U.S. EPA, 1997b)
- Volume II: Point Sources Preferred and Alternative Methods (U.S. EPA, 1997c)
- Volume III: Area Sources Preferred and Alternative Methods (U.S. EPA, 1997d)
- Volume IV: Mobile Sources preferred and Alternative Methods (U.S. EPA, 1997e)

- Volume V: Biogenics Sources Preferred and Alternative Methods (U.S. EPA, 1997a)
- Volume VI: Quality Assurance Procedures (U.S. EPA, 1997f)
- Volume VII: Data Management Procedures (U.S. EPA, 1997g)
- Volume IX: Particulate Emissions (U.S. EPA, 1999e)
- Volume X: Emission Projections (U.S. EPA, 1999f)

In addition, guidance exists which describes the National Emission Trends inventory methodology (U.S. EPA, 1998a). The EIIP documents are available electronically through the U.S. EPA Internet website at ***http://www.epa.gov/ttn/chief/eiip/techrep.htm***. States should consult these documents as they prepare their emission inventories.

15.2 What Countywide Emission Estimates Are Needed To Support Air Quality Models?

Statewide and countywide emissions need to be divided into 4 broad categories: stationary point source emissions, stationary area source emissions, mobile emissions for on-road and off-road sources and biogenic/geogenic emissions. Point sources should be classified by SCC and have associated location information (e.g., latitude/longitude coordinates) as well as diurnal and weekly operating schedules. Area source emissions should be classified by SCC and reported by county. Surrogate factors, used to spatially allocate emissions from the source category within an air quality model grid superimposed over the county, should be identified for each area source category. Defaults for surrogates are available in current emissions models. Examples of surrogate factors might be such things as population or employment by census tract, land use, etc. If information exists concerning weekly and diurnal emission patterns for different area source categories, this information should also accompany the state- and countywide area source emission estimates. On-road and off-road mobile source emissions should be estimated using the most current version of the U.S. EPA MOBILE model (or, in California, the current version of EMFAC) in concert with activity (i.e., vehicle miles traveled (VMT)) estimates. The mobile source emission estimates should be accompanied by recommended surrogates for spatially disaggregating the mobile emissions and by diurnal and weekly activity patterns so that gridded, hourly estimates can be obtained for mobile emission estimates in subsequent steps. We recommend States distinguish weekend vs. weekday activity levels for mobile and stationary area sources. Estimates for biogenic emissions can be made using the most current BEIS emissions model (Geron, et al., 1994, or updates approved by the U.S. EPA). A State should report biogenic emissions on a county basis. Information regarding spatial pattern of land use is needed within each county if a State wishes to distribute biogenic emissions within a county in a non-uniform manner.

For model applications addressing the $PM_{2.5}$ NAAQS, emission estimates for each source category should include countywide $PM_{2.5}$, NH_3, SO_2, NOx, VOC and CO estimates for each month of the year. Regional haze-related applications require the same information plus countywide estimates for emissions of PM_{10}. The $PM_{2.5}$ and VOC estimates should be accompanied by a recommended speciation profile for each source category. We recommend

that States rely on local measurements to the maximum extent possible for the speciation profile estimates. However, default information for PM and VOC species profiles is available in U.S. EPA (1993), if needed. These data and updates can be obtained electronically through the U.S. EPA's Internet website at ***www.epa.gov/ttn/chief/software.html#speciate.*** Finally, information regarding PM size profiles may be needed for major source categories, since some previously archived PM emissions may reflect estimates for PM_{10} rather than $PM_{2.5}$.

Recommendations. States should be familiar with guidance in U.S. EPA (1999d) and with U.S. EPA Emission Inventory Improvement Program guidance describing appropriate procedures for estimating Statewide and countywide emissions needed to support SIP revisions related to $PM_{2.5}$ and regional haze. Air quality models require emission estimates from point, area, mobile and biogenic sources. In order to convert this information for use in air quality models, $PM_{2.5}$ and VOC species profiles, rationale for suballocating emissions within a county and for assuming diurnal and weekday vs. weekend variability in emissions are needed for each point source and for each major area source category, as well as for mobile sources. Default assumptions for spatial/temporal emission allocations are available in emissions models, as are speciation profiles. However, assumptions which are more appropriate for a specific area should be substituted for these, whenever possible.

For applications related to the NAAQS for $PM_{2.5}$, emission estimates for $PM_{2.5}$, NH_3, SO_2, NOx, VOC and CO are needed for each month of the year to support possible use of regional model applications performed throughout the year. Needs are similar for regional haze-related applications, except emission estimates for PM_{10} are also required.

15.3 Can I Use the National Emissions Trends Inventory As A Starting Point?

If there are no previously available modeled inventories to serve as a starting point, we recommend that States derive an inventory suitable for use with models starting from the National Emissions Trends inventory (NET) (U.S. EPA , 1998a). As of late 2000, the most recent NET inventory reflects statewide, annual emission estimates for 1996. The U.S. EPA undertakes a major effort to update the NET estimates every three years. We plan to have a NET inventory reflecting nominal 1999 emissions available in 2001. Since the 1999 estimates reflect a greater effort to compile information regarding NH_3 and primary $PM_{2.5}$ emissions, we recommend that it be used as the basis for initial estimates when needed. Such nominal NET inventories are updated continually based on new information provided by the various States. Thus, it is important that those implementing the modeling protocol make consistent assumptions about which version of the 1999 NET inventory is being used and that the version used is included in the documentation described in Section 7.0. Assuming the lag time between the nominal year of a NET inventory and the availability of that inventory remains approximately the same, we anticipate that a NET inventory reflecting 2002 emissions should be available for use early in 2004.

If the NET inventory is used, it should be for the same year as the current inventory used by the State. This may require some minor adjustments reflecting growth and effects of known national/regional control programs. Statewide emissions, by county, are in the NET inventory and are available electronically through the U.S. EPA Internet website at ***www.epa.gov/ttn/chief/net/index.html***. If a State is performing a regional or nested regional modeling analysis, the NET inventory can often serve to provide countywide estimates for locations far removed from the area which is the focus of the modeled attainment or reasonable progress demonstration. Closer in (within ~ 300 km all major source categories, ~1000 km for large point sources), States should quality assure and improve emission estimates as necessary. Highest priority should be given to quality assuring and improving assumptions underlying the NET inventory in areas where monitored data suggest violations of a NAAQS for $PM_{2.5}$ are likely. For more remote emissions, the NET inventory may be used as is, at a State's discretion.

<u>Recommendations</u>. States may use the National Emissions Trends (NET) inventory to provide initial emission estimates for $PM_{2.5}$ and its precursors. States should focus efforts to improve the NET estimates on all sources near (e.g., ~ 300 km) the nonattainment or Class I area which is the focus of a modeling analysis. NET estimates of emissions from large point sources of $PM_{2.5}$ or its precursors should undergo similar scrutiny if located within ~1000 km of the area which is the focus of the analysis. We recommend that the NET inventory reflecting 1999 emissions be used as a starting point for making emissions estimates. If its availability is timely, 2002 NET inventory estimates should replace those compiled for 1999 when applying modeled tests for attainment of the $PM_{2.5}$ NAAQS or reasonable progress reducing regional haze.

15.4 How Do I Convert Countywide Inventory Information Into Data Used In Air Quality Models?

Air quality models predicting $PM_{2.5}$ or its components require hourly emission estimates which are either day-specific or applicable for types of days (e.g., week days, weekend days). Estimates are needed for primary organic and inorganic $PM_{2.5}$, NH_3, SO_2, NOx, VOC and CO for each cell of a grid superimposed over the area modeled. Typically, there are thousands of grid cells in a model application. To utilize atmospheric chemistry in the air quality simulation model and to perform computations for previously identified major components of $PM_{2.5}$, $PM_{2.5}$ and VOC emissions also need to have their component chemical species identified. We recommend that source specific, local information be used for this purpose whenever possible. The U.S. EPA maintains the SPECIATE data base. SPECIATE can be used when more source-specific information is lacking. It may be accessed electronically at ***www.epa.gov/ttn/chief/software.html#speciate***. In addition, it is a good idea to gather information on size fractionated profiles for particulate matter so that past estimates of PM_{10} emissions can be more readily converted to estimates of $PM_{2.5}$ emissions. Finally, emission factors for some sources are dependent on meteorological conditions such as temperature. Thus, meteorological conditions need to be known to estimate day specific emissions. Emissions models should be used to account for the numerous and diverse factors which need to be

considered to derive emissions inputs to air quality models.

As of late 2000, documentation for three comprehensive emissions models is available or is expected to be so shortly: EPS2.5 (Systems Applications International, 2000), EMS2000 (LADCO, 2000) and SMOKE (MCNC, 2000a). These comprehensive models treat stationary point and area sources and incorporate information developed in models for onroad and offroad mobile emissions and models for biogenic emissions.

Comprehensive emissions models. Two comprehensive emissions models have been used to convert estimated emissions from stationary, mobile and biogenic sources for use in air quality simulation models for PM-related applications.. The first of these is EPS2.5. A description and user's guide for this model is expected to be available in 2001 (Systems Applications International, 2000). EPS2.5 is an adaptation of an earlier model (EPS2.0) which focuses on emissions of VOC, NOx and CO (Causley, et al., (1990), U.S. EPA, (1993a)). EPS2.5 is being used by the U.S. EPA in regional applications of the REMSAD air quality model.

EMS2000 is a second emissions model which will convert available emissions data into emission estimates for $PM_{2.5}$ and its precursors in the form needed by air quality models. EMS2000 is a result of a joint effort between LADCO and Alpine Geophysics, Inc. It is expected to be documented and available by 2001 (LADCO, 2000). EMS2000 is an adaptation of the EMS95 model developed by Alpine Geophysics, Inc., (1995). An adaptation of EMS95 is currently being used in a regional air quality model to investigate causes of regional haze as part of the Southern Appalachian Mountain Initiative (SAMI) (see ***http://environmental.gatech.edu/SAMI/downloads.htm***).

A third comprehensive emissions model, Version 2 of the Sparse Matrix Operator Kernel Emissions (SMOKE), has had limited use to date (MCNC, 2000a). SMOKE is similar theoretically to EMS95. However, it is computationally more efficient, reducing time and memory required to formulate individual control strategies simulated in an air quality model.

Anthropogenic emissions from mobile sources. As of late 2000, MOBILE5B is the most recent available model to estimate mobile emission factors for ozone precursors from a vehicle fleet representative of any specified year (U.S. EPA, 1996). Part 5 of MOBILE5B specifically deals with emissions of primary particulate matter from or associated with mobile sources. For all categories except road dust, States may use scaling factors provided in the documentation to estimate $PM_{2.5}$ emissions from the PM_{10} estimates provided by the model. Size fractionated profiles for resuspended dust, or some other category which those implementing the modeling/analysis protocol believe is similar to road dust, will need to be used in concert with Part 5 emission factors to estimate $PM_{2.5}$ emission factors associated with road dust. To estimate mobile emissions, one must combine information generated with the MOBILE model (or, in California, the EMFAC model) with vehicle activity estimates, such as vehicle miles traveled (VMT).

The U.S. EPA's Office of Transportation and Air Quality (OTAQ) (formerly the Office of Mobile Sources) is developing the MOBILE6 model for highway vehicles as well as a NONROAD model to improve estimates for off-highway vehicles. These two models are expected to be available early in 2001. Although changes in MOBILE6 and NONROAD should generally improve estimates for on- and off-road emissions, neither set of improvements focuses specifically on emissions of primary components of $PM_{2.5}$. Emissions of particulate matter are expected to be addressed more specificallly in the next update to the MOBILE model. We do not expect this to be available before 2002. Prior to the availability updated treatment of emission factors for primary particulate matter from mobile sources, it will be necessary to use Part 5 in concert with the most current available versions of the MOBILE and NONROAD models and external information about size fractionated profiles for road dust.

States may download, as well as track the status of the U.S. EPA's mobile emissions models by accessing the following internet address: ***http://www.epa.gov/otaq/models.htm***. As noted previously, emission factors estimated with MOBILE models need to be combined with activity levels (e.g., vehicle miles traveled) to estimate emission levels which have been suitably disaggregated spatially and temporally for use as inputs in air quality models. Methods for estimating activity levels are included in U.S. EPA, (1997e).

Biogenic Emissions. We recommend the Biogenic Emissions Inventory System (BEIS) model as the default approach for estimating biogenic sources of precursors for $PM_{2.5}$. Version 2 of BEIS (BEIS2) has been the most widely used procedure for estimating biogenic emissions (Geron, et al. 1994 and U.S. EPA, 1997a). This model requires a mix of land uses to be specified for each county, as well as hourly temperature information. If a State believes the average land use mix characterized for a county is inappropriate for certain gridded locations within a county, this may be overridden for the grid cells in question on a case by case basis. The model makes use of stored information regarding geographic distribution of plant species, as well as the provided land use and temperature information, to generate gridded biogenic emissions.

As of late 2000, an updated version of BEIS (BEIS3) is being tested by the U.S. EPA. BEIS3 includes estimates for mono-terpene compounds, needed for visibility- and $PM_{2.5}$-related model applications. Much of the information used as the basis for BEIS3 is currently available in a NARSTO critical review paper by Guenther, et al., (2000) and in Geron, et al., (2000). Much of the documentation for BEIS2 (Geron, et al., 1994) remains relevant. In addition, the Texas Natural Resources Conservation Commission has developed a biogenics emissions model (GLOBEIS) based on the Guenther, et al., (2000) work (TNRCC, 2000). Both BEIS3 and GLOBEIS consider emissions of mono-terpenes. States should apply the most recent available version of BEIS (i.e., BEIS3 or an equivalent model considering mono-terpene emissions) in concert with land use data to support their modeled demonstrations of attainment or reasonable progress. The BELD3 land cover data base, with 1 km resolution and vegetation speciation, has recently become available and may be used to provide default land use estimates for use with the BEIS models when county average figures are not believed to be adequate. BELD3 is described on the following U.S. EPA website: ***http://www.epa.gov/asmdnerl/biogen.html***.

Table 15.1 summarizes available emissions models used in recent air quality model applications, and identifies some example applications. Those which include estimates for primary particulate matter are indicated. There is clearly much work ongoing to improve existing emissions models and to make them more explicit for problems related to $PM_{2.5}$ and regional haze. The U.S. EPA has recently established a website which serves as a clearinghouse for new developments in emissions modeling and related issues. It may be accessed at ***http://www.epa.gov/ttn/chief/emch/***. In addition, a website at ***http://www.epa.gov/otaq/models.htm*** is a good source of current information on emissions modeling related to mobile sources. We recommend that States routinely check these sites for new developments.

Recommendations. States should use emissions models to convert emission inventory estimates into emissions inputs required by air quality models. Emission models require additional inputs concerning chemical speciation, size fractionation, spatial and temporal disaggregation. Generally, States should use the most current available version of an emissions model. Because there is much current activity underway to adapt/improve current emissions models to address ambient problems related to fine particulate matter, States should check the websites identified in this Section for latest developments. Choice of models also depends on compatibility with the chosen air quality model and the application at hand, as well as past experience of those implementing the modeling/analysis protocol. States should quality assure outputs from emissions models prior to making air quality estimates.

15.5 What Should I Do To Quality Assure Emissions Estimates?

The most efficient means to quality assure (QA) emission estimates is to apply QA during the initial emissions estimation process. The previously mentioned EIIP quality assurance document, U.S. EPA (1997f), contains a number of QA procedures that should be used to develop the basic countywide emission inventory. Once the basic emission inventory is ready for modeling, there are three additional quality assurance techniques that may be appropriate. The first is to compare emission estimates and estimating methodologies with those reported by other States. States can use results from such comparisons to see whether their estimates are unusual in any way and, if so, perhaps explain why. This focuses attention on portions of the inventory which appear to differ from estimates made for other locations, so that a State can confirm whether or not its initial estimates are appropriate. The NET inventory provided by the U.S. EPA may be useful for this approach.

Displaying emissions estimates graphically is also a useful means for quality assuring them. Emissions models identified in Section 15.4 can produce graphic displays useful for quality assurance. For example, a tile plot of emissions made for a grid superimposed over the area to be modeled is an effective means for identifying misplaced sources and for assuring oneself that spatial patterns of emissions are consistent with where sources are believed to be. Other graphical displays include pie charts and time series plots. Pie charts are useful for assessing whether distribution of emissions among source types or categories is consistent with

Table 15.1. Some Emissions Models And Example Applications

Emissions Model	References	Sponsors (Applications)
EMS95 (ozone precursors)	Alpine Geophysics, (1995)	LADCO (eastern half of the U.S. for ozone and soon for PM), U.S. EPA, OAQPS (eastern half of the U.S., ozone), NY DEC (eastern half of the U.S., ozone). SAMI (southeastern/mid-Atlantic States, ozone and PM)
EMS2000 (precursors plus primary PM)	LADCO, (2000)	
EPS2 (ozone precursors)	U.S. EPA (1993a)	U.S. EPA, Region IV (Gulf States, ozone), U.S. EPA, OAQPS (contiguous 48 States, ozone) U.S. EPA, OAQPS (contiguous 48 States, PM)
EPS2.5 (precursors plus primary PM)	Systems Applications International, (2000)	
SMOKE (precursors plus primary PM)	MCNC, (2000a)	NC DEM (Charlotte, most of NC and parts of surrounding States, ozone)
MOBILE or EMFAC emission factor models, which must be used in concert with activity estimates to obtain emissions estimates (precursors plus primary PM)	U.S. EPA, (1997e) *http://www.epa.gov/otaq/models.htm*	MOBILE: Many sponsors (throughout the U.S. outside of California) EMFAC: CARB (California)
BEIS2 (ozone precursors)	Geron, et al., (1994), U.S. EPA, (1997e)	U.S. EPA OAQPS (contiguous 48 States, ozone, PM)
BEIS3 (precursors plus primary PM)	Guenther, et al., (2000) Geron, et al., (2000)	
GLOBEIS (precursors plus primary PM)	TNRCC, (2000) Guenther, et al., (2000)	TNRCC (Texas and portions of nearby States)

expectations. Time series displays enable a State to look at estimated diurnal patterns in emissions to see whether these appear logical. They enable comparisons to be made for weekends vs. weekdays to see whether estimated differences appear reasonable.

Comparison with monitored speciated data may become an increasingly important means for quality assuring emissions estimates. Availability of speciated $PM_{2.5}$ and VOC data in ambient networks makes it possible to use monitored observations to apply source attribution approaches (i.e., "receptor models"). A finding suggesting that air quality observations are the product of a mix of emissions which differs greatly from that inferred from the inventory can point the way toward parts of the inventory which may need further scrutiny. Receptor models and their uses have been summarized by Seigneur, et al. (1997). Use of ambient data from the PAMS network to quality assure VOC emissions estimates is described in U.S. EPA (1996a) and in U.S. EPA (2001).

Recommendations. Quality assurance of emissions estimates is an essential part of the modeling process, and should be performed continually. States should consider the following approaches to quality assurance: primary QA emphasis during the initial development of the basic emission inventory, comparison with available emissions estimates and methods performed by others, computer graphics depicting emissions model estimates, and comparison with speciated air quality data.

15.6 How Do I Estimate Emissions For Future Years?

15.6.1 How Far In The Future Should I Project Emissions?

Emissions projections are needed for sources within a modeling domain to determine if a nonattainment area will meet air quality goals by a required date. For applications related to the NAAQS for $PM_{2.5}$, States should estimate future emissions for at least one future date--two years before the date of required attainment. For example, if the required attainment date were 2012, the emission projections should be made to 2010. This earlier date is chosen, because the form of the NAAQS requires that the standard be attained over a 3-year period which ends at the "deadline". As discussed in Section 6.0, when a required attainment date is distant (say, 5 or more years) from the date required for a SIP submittal, we recommend that a State project emissions to an intermediate period as well in order to support a subsequent mid-course review.

For reasonable progress-related applications, we recommend that States retain the convention of projecting emissions to a time which is two years prior to the deadline. If this convention is followed, the projection year would be in the middle of the 5-year period which will later be used to see if air quality observations are consistent with the reasonable progress goals. These projections are most likely to be representative of emissions occurring during this 5-year period. Thus, if the first set of goals for reasonable progress must be realized by 2018, and data from 2013-2017 are to be reviewed to see if this goal is met, emission projections should be made to 2015. The regional haze rule requires a mid-course review of progress to be made at 5-year intervals (U.S. EPA, 1999). Although the rule does not require modeling to meet

this obligation, past model results provide one means for assessing whether air quality observed during 2009-2013 is consistent with expected reasonable progress by that time. Thus, States may find it useful to project emissions to two years prior to the period used for the mid-course review (i.e., 2011) so that air quality projections can be made which are consistent with the selected plan to realize reasonable progress reducing regional haze.

15.6.2 What Emissions Should I Assume In Future Years—Actual Or Allowable?

By "allowable emissions", we mean the maximum amount of emissions from a source which is consistent with emission limitations contained in an applicable operating or construction permit for that source. It could happen that an allowable emissions limit exceeds the maximum amount of emissions possible from a source. If this occurs, the maximum physical limit may be used instead, provided the allowable limit is subsequently reduced (as part of the SIP revision) to be consistent with the physical limits. We define future "actual emissions"as the best estimate of future emissions for a source or source category. Often, future actual emissions may be obtained by multiplying current typical month-specific emissions from a source or source category times an annual growth rate in the activity level for the source compounded by the number of years between the current and future period. However, this need not be the case. For example, if a projection exceeds an allowable emission limit, it should be capped at that limit. In this case, projected "actual" and "allowable" future emissions for a source are one in the same.

In Table 15.2, we provide some general guidance regarding when it is most appropriate to use projected "actual" emissions and when it is most appropriate to use projected "allowable" emissions. However, exceptions to this general guidance are likely to be necessary from time to time. Typically there are a wide range of sources and source categories which will emit $PM_{2.5}$ or its precursors, and no individual source will make an important contribution to total $PM_{2.5}$ concentrations. In such cases, emissions projections can be based upon actual emissions estimates. This method is normally used in ozone SIP planning. However, there may also be situations where a relatively small number of large stationary sources of $PM_{2.5}$ are important contributors to the total $PM_{2.5}$ concentration. This would more likely involve the 24-hour NAAQS than the annual average, and occur in an urban-scale, or topographically constrained modeling domain. In such instances to ensure NAAQS protection, maximum allowable emission rates should be used in preparing future year projections. This method has been used in SIP planning for largely inert pollutants such as SO_2, lead and PM_{10}. We recommend that those implementing the modeling protocol discuss the circumstances under which it is necessary to use "allowable" rather than "actual" projected emissions with the appropriate U.S. EPA regional office(s) and reach agreement on this issue prior to simulating prospective control strategies.

The entries in Table 15.2 contain several words which are subject to varying interpretations. For example, a "major" source is likely to differ for different nonattainment areas. That is, in some areas there may be a substantial number of large sources which do not particularly stand out from one another. In another area there may be a small number of large sources whose emissions are very much greater than those from other sources. In Table 15.2, our meaning of "major" is more applicable to the large sources in the latter case. "Small number" is

another subjective term in Table 15.2. Obviously, the larger the number of sources we are talking about, the less likely it is that they are all operating at allowable limits simultaneously. We do not have a good basis for setting an upper limit in general guidance. Normally, the number of stationary sources projected at allowable rates would be a small fraction of the total number contained in the modeling domain.

Table 15.2. Default Suggestions For Using Projected "Actual" And "Allowable" Emissions

Type Of Modeled Test	When To Use "Allowable" Projected Emissions	When To Use "Actual" Projected Emissions
Attainment Test For Annual NAAQS For $PM_{2.5}$	Rare instances of major $PM_{2.5}$ contribution from a small number of extremely large sources	Typically for all sources
Test For Reasonable Progress Toward Meeting Goals For Regional Haze	-----	For all sources
Attainment Test For 24-Hour NAAQS For $PM_{2.5}$	For major sources whose activity levels are likely to be high during periods with observed exceedances	For all remaining sources
Hot Spot Analysis (Related To 24-Hour NAAQS)	For a source subject to the hot spot review plus a small number of nearby sources which may be operating at allowable limits simultaneously with the source under review	For background estimates resulting from all other sources
Hot Spot Analysis (Related To Annual NAAQS)	Rare instances of major $PM_{2.5}$ contribution from a small number of extremely large sources	Typically for all sources

15.6.3. Additional Guidance For Making Projections

A State's needs for inputs to an air quality model are a factor in making projections. Models require source locations (coordinates) as input. Thus, a projection approach that makes its computations at this level is preferred. A less desirable alternative is to assume that all growth and retirement occurs at existing facilities and that there is no variation in growth or control within each source category. Information detailing the different types of projections that might

be required of a State or local air pollution control agency can be found in the EPA publication "Procedures For Preparing Emissions Projections" (U.S. EPA, 1991a). In addition to the necessary types of projections, methods for projecting changes in future air pollution generating activities, quantifying the effects of current and future controls, and combining effects of growth and control are addressed in this document. Although last published in 1991, much of this guidance for estimating future year emissions is still valid. There have been updates to some of the information provided in the 1991 guidance (BEA projection phase-out and EGAS and MOBILE model revisions, etc.) and therefore States should review additional documentation concerning emissions projections in U.S. EPA (1999f).

States may find it useful to examine techniques that have been applied in other areas where control strategy planning has been performed. In U.S. EPA (1991a), examples of emission projection preparation are recorded in a form suitable for input to a grid-based photochemical model. In the simplest sense, this approach relied on developing a growth factor and a control factor for each major source category.

<u>Recommendations.</u> Emission estimates should be projected to two years earlier than an established deadline for attaining the NAAQS for $PM_{2.5}$ and for meeting the goals for regional haze. States should consider making an additional projection to an intermediate year so as to facilitate a subsequent mid-course review.

States should generally use projected actual emissions in applications related to the annual NAAQS and visibility goals. In applications related to the 24-hour NAAQS, actual projected emissions should also generally be used. However, a small number of the largest stationary sources of particulate matter or its precursors may need to be projected at allowable emission rates. Projected allowable emissions should also be used for a source subject to a hot spot analysis to assess whether the 24-hour NAAQS is likely to be met near a source of primary particulate matter emissions. States may also need to assume allowable emissions for some nearby major sources in a hot spot analysis related to the 24-hour NAAQS. If estimated allowable emissions exceed what is physically possible, they may be reduced to physical limitations or below, provided the allowable limit is reduced accordingly.

States should review guidance on emission projections issued by the U.S. EPA in 1991 and in updates (U.S. EPA, 1999f). States should quality assure their emission projections using several methods designed to validate the spatial and temporal allocations, as well as any speciation that may be calculated. States should review past emission projection efforts as part of any subsequent review performed for the reasons identified in Section 6.0.

15.7 What Emissions-Related Issues Do I Need To Consider When Using Lagrangian Models?

Information presented in Sections 15.1 - 15.6 is generally applicable for using Lagrangian

as well as grid models. However, use of models in hotspot analyses differs from what is done in the modeled attainment or reasonable progress tests. In hotspot analyses models are used in an absolute sense. Thus, a State should undertake considerable effort to review emission factors and activity levels for individual sources subject to a hotspot analysis. Prior to performing a hot spot analysis, a State should assure itself that the current emission factor(s) and activity levels used for a source's emissions are as specific as possible for the source in question. Second, States should compare model predictions for similar sources which are near monitoring sites with speciated measurements made at the monitor to provide additional assurance that the emission factors are appropriate. When there is some doubt, care should be taken to ensure that any applicable allowable emission limit is physically possible for the source in question. If it is not, the limit should be reduced so that a more realistic analysis can be performed.

> **<u>Recommendations</u>. Because a hot spot analysis requires using absolute model predictions, States should make a substantial effort to ensure that emission factors used for the source(s) subject to the review are source-specific and have been quality assured.**

(This page is left blank intentionally)

16.0 How Do I Assess Model Performance And Make Use Of Diagnostic Analyses?

States should evaluate performance of an air quality model in their specific application prior to using model predictions to support an attainment or reasonable progress demonstration. Performance of an air quality model can be evaluated in two ways: (1) how well is the model able to replicate observed concentrations of components of $PM_{2.5}$ and total observed mass of $PM_{2.5}$, and (2) how accurately does the model characterize sensitivity of changes in component concentrations to changes in emissions? The modeled attainment and reasonable progress tests recommended in Sections 3.1 and 3.2 and in Section 4.2 use models to predict sensitivity of components of particulate matter to controls and then apply resulting relative reduction factors to *observed* (rather than modeled) concentrations. Thus, while both types of performance test are important, the second type is the most important. Unfortunately, it is also more difficult to do. As noted in Section 12.1, model performance can best be assessed using extensive data bases, such as those obtained in major field studies. However, we recognize that such data may not be available for every model application which needs to be done. At a minimum, a State should supplement its performance tests with available data with a review of results from performance tests for the model(s) which were completed elsewhere where extensive data bases were used.

Diagnostic analyses are potentially useful for several reasons. First, these analyses can be used to better understand why the air quality model predicts what it does. This yields further insight into whether or not the predictions are plausible. Second, diagnostic analyses provide information which helps prioritize efforts to improve/refine model inputs. Third, diagnostic tests can provide insight into which control strategies may be the most effective for meeting the goals for $PM_{2.5}$ and regional haze. Fourth, diagnostic analyses can be used to assess how "robust" a control strategy is. That is, do I reach the same conclusion regarding adequacy of a strategy when using a variety of assumptions regarding current conditions?

In this section, we first identify methods which may be useful for evaluating model performance. We then discuss each of these methods in greater detail. We next note that there is no single method which offers a panacea for evaluating model performance. We recommend that performance be assessed by considering a variety of methods, much as is done in a weight of evidence determination. We then identify methods for performing diagnostic analyses. We conclude by identifying several potentially useful diagnostic tests which States may consider at various stages of the modeling analysis. Our discussion of diagnostic tests is limited to procedures which may help explain reasons underlying a model's poor (good) performance and may perhaps point the way toward improving performance. In Section 17.0, we identify several tests which may be useful for helping States to choose a control strategy to meet air quality goals.

16.1 How Can I Evaluate Performance Of An Air Quality Model?

As noted above, model performance can be assessed in one of two broad ways: (1) how accurately does the model predict observed concentrations?, and (2) how accurately does the model predict *responses* of predicted air quality to changes in inputs? An example of the latter type of assessment is, "how accurately does the model predict relative reduction factors (RRF)?"

Using terminology found in Seigneur, et al. (2000a), the first type of evaluation is an "operational" evaluation and the second is a "diagnostic" evaluation.

Given existing data bases, nearly all analyses have been operational performance evaluations. The underlying rationale is that if we are able to correctly characterize changes in concentrations accompanying a variety of meteorological conditions, this gives us some confidence that we can correctly characterize future concentrations under similar conditions. Computer graphics, PM component metrics, $PM_{2.5}$ metrics and observational models are all potentially useful for evaluating a model's ability to predict observed air quality.

Diagnostic model performance assessments can be made in several ways. One way is by looking at predicted differences on weekends vs. week days, provided reliable emissions estimates are available for both, and differences in weekend/week day emissions are substantial. A second way is to examine predicted and observed ratios of "indicator species". If observed ratios of indicator species are very high or very low, they provide a sense of whether further production of secondary particulate matter at the monitored location is limited by availability of NOx, SO_2, NH_3 or VOC. Agreement between paired observed and predicted high (low) ratios suggests a model may correctly predict sensitivity of secondary particulate matter at the monitored locations to emission control strategies. Thus, use of indicator species methods shows potential for evaluating model performance in a way which is most closely related to how models will be used in attainment demonstrations. We recommend that greater advantage be taken of these methods in the initial demonstration and in subsequent reviews. A third way for assessing a model's performance in predicting sensitivity of $PM_{2.5}$, its components and the visibility extinction coefficient (b_{ext}) to changes in emissions is to compare projections after the fact with observed trends. One reason States should retain data files and output generated in simulating the control strategy selected for a SIP is to facilitate retrospective analyses. As explained in Section 6.0, these analyses provide potentially useful means for diagnosing why a strategy did or did not work as expected. They also provide an important opportunity to evaluate model performance in a way which is closely related to how models are used to support attainment and reasonable progress demonstrations.

States can assess model performance using graphics, metrics for components of $PM_{2.5}$, predicted total mass of $PM_{2.5}$, corroborative analyses with observational models, weekend/weekday comparisons, ratios of indicator species and retrospective analyses with observed air quality and emission trends. These methods are described in the following subsections.

16.1.1 How Can I Use Graphics To Make A "Big Picture" Assessment Of Model Performance?

States should refer to guidance in U.S EPA (1991) regarding use of graphics to evaluate model performance. Graphs plot predictions and observations. The 1991 guidance describes the following graphical displays: time series plots, tile plots, scatter plots and quantile-quantile (Q-Q) plots. Each of these graphics can also be used to display *differences* between predictions and

their paired observations. Graphics are useful means for understanding *how* predictions and observations differ. For example, time series plots tell whether there is any particular time of day or day(s) of the week when the model performs poorly. Utility of time series presentations is greatly enhanced if some monitored data for $PM_{2.5}$ and at least some components/species are available continuously or for sampling periods of 6 hours or less. Tile plots reveal geographic locations where the model performs poorly. Information from tile plots and time series may provide clues about where to focus quality assurance efforts for model inputs. Scatter plots and Q-Q plots show whether there is any part of the distribution of observations for which the model performs poorly. These plots are also useful for helping to interpret calculations of bias between observations and predictions. For example, they could show large differences between observations and predictions which just happen to balance, producing low estimated aggegrated bias. Since the NAAQS for $PM_{2.5}$ and the regional haze goals will likely require modeling different times of year, season-specific graphic displays may prove helpful for evaluating and diagnosing model performance.

16.1.2 What Metrics And Criteria Should I Consider For $PM_{2.5}$ Components And Related Gaseous Pollutants?

Pollutants/Pollutant Categories of Interest. Because $PM_{2.5}$ is a mixture, a meaningful performance evaluation should include an assessment of how well the model is able to predict individual components of particulate matter. In addition, performance metrics for certain gaseous pollutants are useful for at least two reasons. First, their presence may affect response of components of secondary particulate matter to strategies to reduce precursors. Second, the model's performance in predicting certain gaseous species may provide diagnostic clues which help explain poor performance predicting secondary components of PM. Components of particulate matter of interest include:

- SO_4 and/or S
- NH_4
- NO_3
- mass associated with SO_4
- mass associated with NO_3
- EC
- OC
- IP (mass of inorganic primary particulate matter)
- mass of individual constituents of IP.

Gaseous pollutants of interest include:

- ozone
- HNO_3
- NO_2
- PAN
- NH_3

- NOy
- SO_2
- CO
- H_2O_2.

A few words are needed about some of the less obvious entries listed above. For purposes of model performance evaluation, comparing predicted and observed SO_4 ion, sulfur species and NO_3 ion accomplishes the same purpose as comparing predicted and observed mass associated with sulfate and nitrate. Comparing predicted and observed ammonium (NH_4) ion provides a means for checking whether assumptions made about the mass assumed to be associated with measured sulfate ion or sulfur is correct. When considered in concert with the predictions of sulfur (S) or sulfate ion, it may also help provide insight into whether the model is accurately predicting if mass associated with sulfate and/or nitrate is limited by availability of ammonia. Making continuous measurements of some key components of particulate matter, like sulfates, would permit hourly comparisons to be made between observations and predictions. This would be valuable in assessing whether the model is capturing observed diurnal patterns correctly. Mass of individual constituents of IP is of use for determining whether the modeling is predicting an appropriate mix of sources which contribute to IP.

Presence of hydrogen peroxide (H_2O_2), ozone and related OH radical concentrations can affect both gas- and liquid phase oxidation of SO_2 to particulate sulfate, as well as conversion of NO_2 to HNO_3 (and subsequently to particulate nitrate). Further, ozone measurements are relatively easy to make and are available on an hourly basis. Thus, a model's performance in predicting observed ozone has substantial diagnostic power, not available in 24-hour samples of PM and its components. Comparison of observed and predicted concentrations of HNO_3, PAN and NO_2 may provide an explanation for poor agreement between observed and predicted concentrations of particulate nitrate. That is, it could provide insight into whether poor agreement arises from the model's mischaracterizing volatility (i.e., gaseous species too high, nitrate too low or vice versa) or mischaracterizing deposition, emissions or meteorological processes (i.e., disagreements of gaseous and particulate species are in the same direction). Comparisons between observed and predicted concentrations of ammonia (NH_3) may provide similar insights into whether apparent disagreements between predicted and observed nitrate concentrations are attributable to mischaracterizing volatility of the nitrates on the filter or to some more fundamental cause.

Metrics. Recall that relative reduction factors for components of particulate matter are estimated by taking the ratio of mean spatially averaged predicted concentrations near a monitor obtained with future emissions to that obtained with current emissions. Thus, a key question to address is, how well does the model predict current spatially averaged concentrations near a monitor which are averaged over the modeled days with corresponding monitored observations? For performance at a monitoring site, j, this metric is given by Equation (16.1).

$$A_{ij} = [(1/n) \Sigma_{\textit{all modeled days w. monitored data}} \textit{ABS} [(C'_p - C_o) / C_o]_{i,t}]_j \qquad (16.1)$$

where "ABS" denotes an absolute (unsigned) value of the difference between observations and predictions,

C'_p is the spatially averaged 24-hour prediction for component i of particulate matter on day t;

C_o is the monitored 24-hour observation of component i on day t;

n is the number of modeled days with monitored data at site j.

This metric should be applied at each monitoring site with the needed data. For applications related to the NAAQS for $PM_{2.5}$, greatest concern for good performance is at sites where observations indicate the standard is not being met.

Other useful insights may result from metrics which are also frequently used in evaluating a model's ability to predict ozone concentrations. These include fractional bias between means (i.e., averaged over several days) and standard deviations associated with predictions vs. observations, normalized bias, normalized gross error and unpaired comparisons between observed and predicted peak concentrations. Formulas for estimating the first four of these metrics at an individual monitoring site, j, are shown in Equations (16.2) -(16.5).

Fractional bias

$$F_{mj} = [2 \times (m_p - m_o)/(m_p + m_o)]_j \quad (16.2)$$

and

$$F_{sj} = [2 \times (\sigma_p - \sigma_o)/(\sigma_p + \sigma_o)]_j \quad (16.3)$$

where

F_{mj} = fractional bias of means at site j

F_{sj} = fractional bias of standard deviations at site j

m_p, m_o = the mean predicted and observed concentration averaged over several days (hours)

σ_p, σ_o = the standard deviation in predicted and observed concentrations computed for several days (hours).

Normalized bias

$$B_j = [1/n \, G_{\text{all modeled days with monitored data}}(C_p - C_o)/C_o]_j \quad (16.4)$$

Normalized Gross Error

$$G_j = [1/n \, G_{\text{all modeled days with monitored data}} \, ABS \, [(C_p - C_o)]/C_o]_j \quad (16.5)$$

where

B_j = normalized bias estimated at site j

G_j = normalized gross error estimated at site j

C_p = predicted concentration at site j on a given day (or hour)

C_o = observed concentration at site j on a given day (or hour)

n = the number of modeled days (hours) for which observations are available.

Note that units of time associated with C_p, C_o, etc. can be days (i.e., usually for particulate matter and its species) or hours (i.e., usually for species with continuous measurements, like ozone). Also note that the preceding metrics may not be that meaningful if the number of modeled days with monitored data is limited at a site. There is no hard and fast rule about how many comparisons are needed for the preceding metrics to be meaningful. However, States should be cautious about interpretations which are made with sample sizes of less than ~ 15 or so.

If a State is able to model only a relatively few days with monitored observations, there are "aggregated versions" of the preceding metrics which may be used to evaluate model performance. By this, we mean predictions and observations occurring at different sites are pooled into a single data base. Formulas which are very similar to those shown in Equations (16.2) - (16.5) may then be applied. Aggregated statistics have been widely used to estimate model performance in predicting observed ozone. In fact, performance criteria mentioned in past U.S. EPA modeling guidance are specific for several of these aggregated metrics (U.S. EPA, 1991, U.S. EPA, 1996b, U.S. EPA, 1999b). Equations (16.6) - (16.9) reflect some widely used aggregated metrics.

Aggregated Fractional Bias in Means

$$F_m = 1/N \, [G_{\text{all monitoring sites}} \, (F_{mj}) \quad (16.6)$$

Aggregated Fractional Bias in Standard Deviations

$$F_s = 1/N \, [G_{\text{all monitoring sites}} \, (F_{sj}) \quad (16.7)$$

where

N = the number of monitoring sites modeled, and the other terms are defined previously.

Aggregated Normalized Bias

$$B = 1/N_{TOT} \left(G_{\text{all monitoring sites}} \left[G_{\text{all modeled days with monitored data}}(C_p - C_o)/C_o\right]_j\right) \quad (16.8)$$

Aggregated Normalized Gross Error

$$G = 1/N_{TOT} \left(G_{\text{all monitoring sites}} \left[G_{\text{all modeled days with monitored data}} ABS \left[(C_p - C_o)\right]/C_o\right]_j\right) \quad (16.9)$$

where

N_{TOT} = the number of days (or hours) with paired predictions and observations at all monitoring sites which are modeled. The other terms have been defined previously.

Performance Goals. It is difficult to establish generally applicable numerical performance goals for the previously identified metrics. This is, in part, a result of how we recommend models be used in attainment and reasonable progress demonstrations. For a given component of particulate matter, Equation (16.10) reflects our recommended methodology.

$$(C_p)_{future} = [(\text{Modeled Conc.})_{future} / (\text{Modeled Conc.})_{current}] [\text{Current Obs.Conc.}] \quad (16.10)$$

Where $(C_p)_{future}$ is the estimated future concentration for the component in question.

If the current observed concentration of a component is small relative to observed concentrations of other components, it is not particularly important how closely the model replicates its observations—since the observed concentration is small, a poor prediction can have little effect on the outcome of the modeled attainment test.

With some modification, the same conclusion reached in the preceding paragraph applies to the modeled test for reasonable progress as well. Here however, the product of a component's observed concentration and that component's light scattering/absorption efficiency needs to be small relative to those for other components for model performance predicting the component's concentration to be unimportant. Recall that component-specific light scattering/absorption efficiencies are represented by the numerical coefficients on the right hand side of Equation (4.1).

The other difficulty in establishing meaningful, generally applicable performance goals results from the model's use in a relative sense.[26] For example, if a component is overpredicted

[26]Performance goals for models used in hot spot analyses, where absolute predictions are used, will be discussed shortly.

with current emissions, a similar bias will probably occur with future emissions since the projections assume constant meteorological conditions. Hence, the effects of the bias may be reduced. Because the model is used in a relative sense, all things being equal, poor performance may have its greatest effect if the model predicts low concentrations of a component with current emissions. This could lead to low concentrations in the denominator of the component-specific relative reduction factor and could potentially magnify the effects of any error accompanying future estimates.

Based on the preceding discussion, States should be most concerned over model performance if there are discrepancies between observations and predictions for components constituting a major portion of the observed mass of $PM_{2.5}$ or accounting for a major portion of b_{ext} derived from observed components of particulate matter. In addition, large underpredictions of these major components are likely to be of greatest concern.

An important additional goal which States should consider when evaluating model performance is to be able to predict concentrations of components (e.g., SO_4, NO_3, OC) as well as ozone in *relative proportions* which approximate those observed. This follows, since the response of secondary components to a control strategy may depend on how much of the other components are present. Thus, for example, if observations suggest sulfate concentrations are about 3 times nitrate concentrations, good model performance should result in predictions which approximate this 3:1 ratio.

Finally we come to the most difficult issue of all—how good should a State expect performance of a model to be? Frankly, there is little basis for making recommendations at present (January 2001). Seigneur (2000b) summarizes information based on very few sets of comparisons. Some of the sets are based on data for as little as one day. Table 16.1 summarizes the information presented by Seigneur (2000b). Note that the numbers presented in the table suggest performances which are not as good as typically expected for ozone models. For example, performance criteria for aggregated gross error and aggregated normalized bias for ozone models are $\leq$ 30% gross error and $\leq$ ABS [15%] normalized bias. Poorer performance for PM models should not be surprising. Since processes involved with ozone formation are also instrumental in leading to secondary particulate matter, we believe that good performance predicting ozone is a prerequisite for good performance predicting secondary particulate matter. However, there are additional physicochemical pheonomena affecting particulate matter. It would be hard to achieve performance better than what is expected for ozone models. Thus, States should regard previously identified performance goals for ozone models as upper bounds for what one might expect for models of particulate matter and its components. We list these goals following Table 16.1.

Table 16.1. Limited Observed Performance Statistics For Particulate Matter Models (Dec. 2000)

Pollutant	Aggregated Gross Error*	Aggregated Normalized Bias**
$PM_{2.5}$	~ 30 - 50%	~ +10% (only one set of comparisons)
Sulfate	~ 30 - 50%	~ -20 to - 30%
Nitrate	~ 20 - 70%	~ -15 to + 50%
EC	~ 15 - 60%	None Available
OC	~ 40 - 50%	~ +38% (only one set of comparisons)

*** Based on 3 sets of comparisons (different models & locations)**

****Unless otherwise noted, based on 2 sets of comparisons (different models & locations). Note also that the statistics for aggregated gross error and aggregated normalized bias were not obtained with the same models nor with the same data base. Thus, performances shown for the two metrics in the table are not associated with each other.**

Metric or Criterion	Ozone-related Recommendations
Difference in Averaged Predictions Near Each Monitor (A) (Eq. (16.1))	≤ 0.15
Unpaired 24-hr peaks	$\pm$ 20% difference
Aggregated Normalized Bias & Normalized Bias	≤ 0.15 (absolute value)
Aggregated Normalized Gross Error & Normalized Gross Error	≤ 0.30

Performance goals for predicting relative importance of components of PM differ, according to whether a component comprises a major portion of measured PM. For major components (i.e., those observed to comprise at least 30% of measured $PM_{2.5}$), we propose that the relative proportion predicted for each component averaged over modeled days with monitored data agrees within about 20% of the averaged observed proportion. For example, if

one observed that on average over a 20-day period measured mass associated with sulfates were 50% of measured mass associated with (sulfates + nitrates + organic carbon + elemental carbon), this performance goal is that the model should predict that sulfate comprised 40-60% of the modeled sum of (sulfates + nitrates + organic carbon + elemental carbon) on the same days as those with the observations.

For minor observed components of PM, we suggest a goal that the observed and modeled absolute proportion of each minor component agree within 5%. For example, if , on average, mass associated with nitrates was observed to comprise 15% of the mass associated with (sulfates + nitrates + organic carbon + elemental carbon), for the model to meet this goal, it would need to predict that, on average, nitrates comprised 10-20% of the predicted mass of (sulfates + nitrates + organic carbon + elemental carbon).

There may be some inconsistencies between the two preceding performance goals for "minor" components which are almost "major" or vice versa. If this occurs, we suggest using the least stringent of the two goals.

Although we have identified several performance goals, we believe that it is preferable for those implementing the protocol to decide what constitutes acceptable performance on a case by case basis using diagnostic analyses like those described in Section 16.2.

Performance evaluation of hot spot models. Hot spot analyses superimpose predictions of primary components of $PM_{2.5}$ on assumed background levels for secondary components plus primary components not attributed to the modeled source(s). Thus, performance evaluations should focus on comparisons between predicted and observed concentrations of IP, EC and primary portions of OC. One reason for performing a hot spot analysis, of course, is that there is no monitor near a major source or major group of sources. Thus, in the absence of monitored data, the performance evaluation should consist of two parts: (1) assurance that the model used has a history of good performance, and (2) assurance that the emissions and meteorological data used as inputs in the analysis at hand are reasonable.

States should review Chapter 10 in 40 CFR Part 51, Appendix W concerning evaluation of Lagrangian (e.g., Gaussian) models. In particular, available results obtained for metrics described in U.S. EPA (1984) and in U.S. EPA (1992) should be reviewed. Generally, models identified as "preferred models" in Appendix W have a history of adequate performance. In conducting this review, States should focus on results obtained in applications which are as similar as possible to the one being considered. Efforts to quality assure emissions and meteorological estimates used as input in the hot spot analysis should include the following. First, States should compare emission factors and activity levels assumed for the modeled source(s) to those made elsewhere for similar source categories to ensure that the estimates are reasonable. Engineering limits on activity levels should also be observed. Meteorological data used to support the analysis should be reviewed to ensure that the data used do not reflect physically inconsistent combinations of meteorological inputs.

16.1.3 What Metrics And Criteria Should I Consider For Predicted $PM_{2.5}$ Concentrations And Light Extinction?

Since total measured mass of $PM_{2.5}$ is the entity which is specified in the NAAQS, States should review how well a model is able to reproduce these observations. We recommend using the following metrics: difference in mean spatially averaged predictions near a monitor and observations at the monitor (A), fractional bias, normalized bias, normalized gross error, aggregated fractional bias, aggregated normalized bias, aggregated normalized gross error and comparison of unpaired peak predictions and observations. These metrics and related performance goals are described in Section 16.1.2. Whenever possible, States should make sure that performance evaluations related to total $PM_{2.5}$ are accompanied by information about how well the model is able to replicate observed concentrations of components. Failure to do so increases the possibility of (a) "getting the right answer for the wrong reasons", or (b) observing poor performance with few clues as to why.

Light extinction (i.e., b_{ext}) is a function of concentrations of components of particulate matter plus relative humidity. However, as explained in Sections 4.1 and 4.2, our recommended test for reasonable progress uses climatological mean estimates for relative humidity rather than day-specific values. Therefore, the metrics and performance criteria for components of $PM_{2.5}$, as well as similar metrics for coarse particulate matter (PM_{10} - $PM_{2.5}$) are the best way to evaluate model performance predicting light extinction.

Evaluation of a model's ability to reproduce observed day-specific values for b_{ext} may have some diagnostic value, however. For example, if the model reproduces observed concentrations of particulate matter well, but doesn't do well replicating day-specific values for b_{ext}, this suggests that a poor job may be done predicting relative humidity. This could then lead a State to focus on the predictions of humidity obtained with the meteorological model as a possible source of error. Use of tile diagrams showing differences in predicted and observed values for relative humidity and time series plots depicting diurnal differences may be useful to assess sources of error.

16.1.4 How Can I Use Corroborative Analyses With Observational Models To Help Evaluate Air Quality Model Performance?

Recently, techniques have been developed to embed procedures within the code of an air quality model which enable users to assess contributions of specific source categories or of specific geographic regions to predicted air quality at specified sites (ENVIRON, 1997, Yarwood and Morris, 1997, Yarwood, et al., 1997a, Yang, et al., 1997, 1997a). These source attribution procedures characterize what the air quality model *says* are the effects of targeted areas or sources on predicted air quality. Provided speciated $PM_{2.5}$ data are available at a site, source attributions estimated with these approaches can be compared with those obtained using other models which rely directly on observed air quality data.

The chemical mass balance model (U.S. EPA, 1990, Seigneur, et al., 1997) is probably

the most directly applicable observational approach for this purpose, since it can focus on the same day(s) considered with the air quality model. Cautions raised previously about representativeness of the monitored data continue to apply. Available multi-variate statistical models (see, for example, Paatero, et al., 1994 and Henry, 2000) may provide a more qualitative means for assessing an air quality model's performance. Multi-variate statistical models work by examining temporal variability in monitored precursor species at a single site or spatial variability on one or a few occasions at many sites. A qualitative comparison is possible if one can contrast observations on days when winds suggest a source contribution is unlikely vs. days when a contribution is likely, or at locations where a source category is important vs. those where it isn't. If the observational approach suggests a major change in a source category contribution, and the air quality model also suggests that category is important or unimportant under similar wind conditions, the observational model lends credence to the air quality model's predictions.

16.1.5 What Data Bases Reflecting Changes In Emissions Are Available To Evaluate Model Performance?

Activity levels and patterns, leading to $PM_{2.5}$ and precursor emissions from mobile, area and some point sources, may differ on weekends vs. week days. If these differences are substantial, simulating weekend as well as week days could provide a means for evaluating how accurately a model predicts the effects of changing emissions. Similar comparisons could be made between seasons (e.g., does the model accurately reflect observed differences in component concentrations in summer vs. winter?). We describe weekend/week day comparisons in the following paragraphs. Where noted, parts of the discussion are also applicable for seasonal comparisons.

Weekend/week day information could be used in one of two ways. The first way is to compare mean predicted and mean observed $PM_{2.5}$ or, preferably, components of $PM_{2.5}$ at each monitoring site for weekends vs. week days. If there are a sufficient number of monitors available, it is also desirable to make these comparisons for categories of monitors, grouped according to whether they represent "downwind", "center city" or "upwind" conditions. Metrics, described in Section 16.1.2, as well as other tests, could be applied first for week days and then for weekend days. If the performance is adequate for both weekends and weekdays, this suggests that the model is accurately characterizing composite effects of different meteorological conditions and different emissions. Seasonal comparisons could also be used for this purpose.

A second way for using weekend/week day information is to first screen the available data to identify weekend days and week days for which meteorological conditions are "similar". For example, for urban analyses, wind orientation, daily maximum surface temperature, presence of precipitation and maximum mixing height might be considered for this purpose. Alternatively, a CART analysis might be performed to identify classes of "similar" days. If similar sets of meteorological conditions are identified for weekends and week days, *changes* in mean observed 24-hr $PM_{2.5}$ or component concentrations can be compared with *changes* in mean predictions made at each monitor site, as well as for groups of sites characterized as "upwind", "center city" and "downwind". Metrics like those described in Section 16.1.2 may be used for

this purpose. If predicted changes generally provide an unbiased estimate of observed changes, this suggests that the model characterizes effects of changing emissions accurately. Since meteorological conditions are likely to differ in different seasons, it probably will not be feasible to use seasonal comparisons in this manner.

We need to mention several caveats regarding weekend/weekday comparisons. First, changes in emissions between week days and weekends may be small compared to uncertainties associated with the weekend and week day estimates. Second, dividing a (already small) sample into weekends and week days may mean that conclusions are based on few comparisons. Third, identifying "similar" meteorological conditions may be somewhat subjective. Finally, the changes between weekend and week day emissions may be considerably less than the changes needed to meet the NAAQS in some areas. Since the relationship between precursor and $PM_{2.5}$ concentrations may be nonlinear in some cases, the weekend/weekday comparisons may not be definitive evaluations. Although several of these caveats also apply to seasonal comparisons, they may do so to a lesser extent. Despite these reservations, weekend/week day (and seasonal) comparisons provide two of a relatively few means for evaluating a model's ability to accurately predict changes in $PM_{2.5}$ or component concentrations. States should include these comparisons in their efforts to evaluate model performance, whenever feasible.

16.1.6 How Do I Use Ratios of Indicator Species To Evaluate Model Performance?

A performance evaluation which includes comparisons between modeled and observed ratios of indicator species carries with it a large potential advantage. Such a comparison may reveal whether the model is predicting sensitivity of secondary components of $PM_{2.5}$ to changes in SO_2, NH_3, VOC and/or NOx controls correctly. That is, when the model predicts ratios within a certain range, predicted secondary components of $PM_{2.5}$ are sensitive to changes in a particular precursor. Within another range of ratios, predictions are sensitive to changes in another precursor. If a model predicts *observed* ratios of indicator species such that observed and predicted ratios fall within the same range of ratios, this provides some confidence that the predicted *change* in particulate matter may be accurate.

As of late 2000, nearly all applications of indicator species approaches have addressed ozone-related problems. Sillman (1995), Sillman (1998) and Lu and Chang (1998) provide good descriptions of the method, identify key ratios and illustrate application of the approach. Even though the preceding ratios are oriented toward ozone, they can provide insight into why modeled concentrations of secondary particulate matter are sensitive to changes in VOC, or NOx emissions (Pun and Seigneur, 1999).

Ansari and Pandis (1998) have developed an indicator ratio of species and applied it to several combinations of secondary particulate matter present under different environmental conditions. They use this ratio to predict how mass of particulate matter will respond to reductions in sulfate, nitrate and ammonia. Blanchard, et al. (2000) have also examined how indicator species might be used to assess whether particulate nitrate concentrations are limited by NOx or by ammonia emissions using mechanisms which incorporate reactions dealing with

secondary particulate matter. These authors identify two ratios of indicator species which appear potentially useful for identifying limiting precursors for secondary nitrate particulate matter: (1) the ratio of particulate ammonium plus gas-phase ammonia over the sum of nitric acid plus particulate nitrate plus particulate sulfate, and (2) the ratio of particulate to total nitrate. It is likely that additional indicator species approaches will be identified as the user community gains more experience with chemical mechanisms incorporating secondary particulate formation and more speciated particulate and gas phase ambient measurements become available.

Strength of the indicator species approach for assessing model performance depends on an assumption that the model is accurately characterizing the relationships between indicator species and secondary components of $PM_{2.5}$. A second precaution is that there may be a range of observed ratios for which the preferred direction of control is not clear. When this occurs, agreement between predictions and observations does not necessarily imply that the model correctly predicts sensitivity of secondary particulate matter to changes in precursors. Third, this method requires more measurements than are commonly made. In some cases, it may be difficult to achieve the required precision with routine monitoring. Finally, much of the work done to date with indicator species has focused on peak hourly concentrations of ozone. Applicability of the approach to secondary particulate matter averaged over 24-hour sampling times has not yet been extensively tested. Despite these precautions, the approach of comparing predicted and observed ratios of indicator species provides a means of assessing a model's ability to accurately characterize sensitivity of predicted secondary components of $PM_{2.5}$ to changes in precursors. States should use the method to help evaluate model performance, whenever feasible.

16.1.7 Are Retrospective Analyses Useful For Evaluating Model Performance?

Retrospective analyses compare past model air quality projections with observed trends in air quality and estimated trends in emissions. The approach is a direct assessment of what we are most interested in---does the model accurately predict changes in air quality? However, it is not as straightforward as it seems. Often, input estimates and assumptions used in past studies are ambiguous and the emissions trends are qualitative. Also, the past studies generally assume constant meteorology, which does not happen. One of the purposes of the reporting requirements described in Section 7.0 is to make it possible for others to replicate modeled analyses at future dates.

In Section 6.0, we noted that a retrospective analysis is an important means for diagnosing why an air quality goal has or has not been attained. Such an analysis provides assurance that improved air quality results from changes in emissions rather than meteorology and/or can identify reasons why satisfactory progress is not being observed. Retrospective analyses will have an ancillary benefit of providing an additional means for evaluating model performance. In order to ensure some planning for subsequent retrospective analyses and to promote some uniformity in the methods used for these analyses, they are probably best performed as part of a subsequent review rather than as supporting evidence in the initial SIP revision.

16.1.8 All Of These Performance Tests Have Shortcomings, So What Do I Do?

There is no single definitive test for evaluating model performance. All tests have strengths and weaknesses. Credence given to model results is increased if a variety of tests is applied and the outcomes either support a conclusion that the model is working well or, at least, are ambiguous. Thus, one can think of a model performance evaluation as a "mini-weight of evidence analysis" focused on the issue of how much credence to give model results in an attainment or reasonable progress demonstration. Table 16.2 summarizes the tests and their corresponding objectives or goals described in this guidance.

Table 16.2. Summary Of Methods To Evaluate Performance Of Air Quality Models

Method	Test(s)	Goals/Objectives
Do a "Big Picture" Assessment Using Computer Graphics	**-tile plots of observations & predictions.**	**-complement metrics for components of PM**
	-tile plots of *differences* in observations & predictions	**-determine whether performance is worse for high vs. low observations**
	-scatterplots & Q-Q plots	
	-time series plots	**-look at spatial patterns of performance--is performance better downwind than upwind?**
		-focus diagnostic tests on certain times/locations

Table 16.2. Summary Of Methods To Evaluate Performance Of Air Quality Models (continued)

Method	Test(s)	Goals/Objectives
Compare Predicted and Observed Concentrations of Components of PM & Related Gaseous Species	**-mean observed and predicted component concentrations near each monitor**	**- agreement within about 15%, especially where observed violations of the NAAQS occur**
SO_4 or S **NH_4** **NO_3** **EC** **OC** **IP** **Indivdual constituents of IP**	**-normalized bias near each monitor**	**-within or near ~20% most monitors**
	-fractional bias near each monitor	**-within or near ~20% most monitors**
<u>Gaseous Species</u> **O_3** **HNO_3** **NO_2** **NH_3** **PAN** **NOy** **H_2O_2** **SO_2** **CO**	**-correlation coefficients, all data, temporally paired means, spatially paired means**	**-moderate to large positive correlations**
	-normalized bias, all monitors	**-within or near ~ 15%**
	-normalized gross error all monitors	**-within or near ~ 30%**
	-compare predicted & observed relative proportions of SO_4, NO_3 & secondary OC	**-agreement within ~20% (relative measure)–major components,** **-agreement within ~5% (absolute measure)–minor components**
	-compare predicted & observed ozone concentrations using metrics identified in U.S. EPA (1991, 1999e)	**-agreement within bounds identified in U.S. EPA (1991, 1999e)**
	-partition pooled data base into "upwind", "center city" & "downwind" sites. Repeat analyses	**-get a better idea of what parts of the distribution of predictions & observations agree or disagree & whether there is any obvious pattern to the model's performance**
	-scatterplots & Q-Q plots of metrics	

Table 16.2. Summary Of Methods To Evaluate Performance Of Air Quality Models (continued)

Method	Test(s)	Goals/Objectives
Compare $PM_{2.5}$ and b_{ext} Predictions and Observations	**-similar to metrics for components of PM** **-this method should be accompanied by evaluation of a model's ability to reproduce observed concentrations of components whenever possible**	**-provide means for assessing whether model reproduces observations of $PM_{2.5}$** **-provide means for assessing whether model predicts relative humidity sufficiently well** **-no specific goal for performance of model in predicting precursor concentrations is given.**
Compare with Observational Models	**-compare source attribution estimates with observational models.** **-CMB** **-multi-variate models**	**-source attribution & CMB identify similar source categories as being important contributors to observed precursor or PM component concentrations** **-day to day variability in air quality model's source attribution is consistent with observations and/or with output from mulit-variate models** **-these are qualitative comparisons**

Table 16.2. Summary Of Methods To Evaluate Performance Of Air Quality Models (continued)

Method	Test(s)	Goals/Objectives
Weekend/Week day or Seasonal Comparisons	**-compare previously identified PM component metrics on weekends vs. weekdays (or between seasons)** **-if data base permits, partition data base into meteorological classes. For each class compare *differences* in weekday vs. weekend (or seasonal) predictions with *differences* in weekday vs. weekend (or seasonal) observations.** **-pool data base to compute bias and gross error on weekends and week days or for the two seasons being compared.** **-if data base permits, partition pooled data base into "upwind", "center city" and "downwind" bins & perform the previously identified pooled tests.**	**-objective is to test model's ability to accurately reproduce effects of *changing* emissions** **-same performance goals for weekends and weekdays or for each season as mentioned for PM component metrics.**
Compare Predicted and Observed Ratios of Indicator Species	**compare predicted and observed ratios at each site.** **see Ansari and Pandis (1998) and Blanchard, et al., (2000) for specifics regarding potentially useful ratios**	**predictions & observations should fall into the same class (e.g., SO_2-limited cases, NH_3-limited cases, VOC-limited cases, NOx-limited cases, cases where it is too close to call)**

Table 16.2. Summary Of Methods To Evaluate Performance Of Air Quality Models (concluded)

Method	Test(s)	Goals/Objectives
Perform Retrospective Analyses to Compare Previously Made Projections With Subsequently Observed Trends	**-project PM and its components to a future (preferably sooner than attainment or reasonable progress date) year** **-retain files** **-update emission estimates at future year & note observed future concentrations of PM and its components** **-characterize future meteorological episodes & model in future**	**-recommended primarily for a subsequent review.** **-note agreement/disagreements between projected PM components & subsequent observations** **-perform diagnostic tests to determine whether disagreement is due to** **--differences in projected emissions vs. emissions estimated at a future date** **--differences in assumed meteorlogical conditions** **--a combination of different meteorological and emissions assumptions** **--one or more limitations in the model.**

Finally, we need to address the issue of adjusting model inputs to improve model performance. One of the reasons we recommend a variety of tests for model performance is to reduce the possibility of "getting the right answer for the wrong reason". We recognize however, that many of the inputs to models have associated (often unknown) uncertainties. It is acceptable to adjust inputs within reasonable bounds to improve performance, providing it does not result in poorer performance in any of the several measures of performance which we recommend in Sections 16.1.1 - 16.1.7. If such an adjustment is made, it should be documented and accompanied by an explanation as to why those implementing the protocol believe it is justified.

<u>Recommendations.</u> States should undertake a variety of performance tests. Results from a diverse set of tests should be documented and weighed to qualitatively assess model performance. Those implementing the modeling/analysis protocol should review available data bases and decide the most appropriate mix of performance tests relatively early in the modeling process. Provided suitable data bases are available, greatest weight should be given to tests performed for components of particulate matter and which assess model capabilities most closely related to how

the model is used in the modeled attainment and reasonable progress tests. A narrative describing overall assessment of model performance should be included among the material submitted to support a recommended SIP revision requiring a demonstration of attainment or reasonable progress.

16.2 How Can I Make Good Use Of Diagnostic Tests?

Diagnostic tests are used to explain model performance and to provide clues about how to improve reliability of predictions. These tests are performed using one of two broad approaches. The first approach consists of tests in which sensitivity of air quality predictions to perturbations in one or a combination of model inputs is examined. This is the more traditional of the two approaches and has a longer track record. When it is applied, States should recall how model outputs are used in the modeled attainment and reasonable progress tests recommended in Sections 3.0 and 4.0. That is, models are used in a relative sense to provide relative reduction factors. Relative reduction factors are obtained by taking the ratio of mean concentrations of PM components obtained with future vs. current emissions.

The second type of diagnostic test is one in which means for "tracking" the importance of various phenomena contributing to predicted air quality at a location are embedded within the code of an air quality model. This generally increases running time and should not be used unless a vendor or someone very familiar with the computer code has installed this capability and performed benchmark tests to ensure that the model, with and without the code revisions, yields identical results. The major advantage of this latter type of diagnostic test is that it reduces the number of model simulations needed to obtain insights about what is causing high or unexpected air quality predictions. Another advantage is that it can provide insights into what additional sensitivity tests should be performed. Thus, once the initial effort to develop this capability has been expended, a better understanding of why predictions are the way they are can be obtained relatively efficiently.

16.2.1 Use Of Sensitivity Tests

Outcomes of sensitivity tests are useful for several purposes. First, the tests can be used to see whether model performance is especially sensitive to a particular input or combination of inputs. Second, sensitivity tests may help prioritize additional data gathering efforts so that a better subsequent review/diagnosis can be performed at the time of a mid-course review or required attainment date. models used to simulate formation, transport and deposition of secondary particulate matter are resource intensive. Third, the tests can be used to assess the robustness of a control strategy. For example, States can consider effects of assumed boundary conditions and meteorological assumptions on predicted effectiveness of a control strategy. If the control strategy appears to work for a variety of assumptions, this increases confidence in the strategy. Sensitivity tests also provide a means for prioritizing use of resources in applying the model. For example, how sensitive are relative reduction factors to use of more layers or smaller grid cells? Is using 12 km (rather than 36 km) grid cells more important than simulating many days? Finally, sensitivity analyses could be useful for prioritizing control efforts or for noting

sensitivity of predictions to uncertainties in the current or future emission inventory. The first and second types of sensitivity tests are relating to diagnosing reasons for the outcome of model performance tests. They are discussed in this Section. The third and fourth types of tests are useful to help choose a strategy which will meet the air quality goal. They are discussed in Section 17.0.

Sensitivity tests should be applied throughout the modeling process, not just when model performance is being evaluated. Tests should be selected on a case by case basis by those implementing the modeling/analysis protocol. We present several tests which may be useful for setting up model runs and for evaluating performance. In Section 17.0, we identify additional tests which may prove useful for selecting strategies and for estimating the degree of uncertainty associated with model predictions and whether or not a particular strategy will be successful in helping a State to meet an air quality goal. a sequence of activities likely to be followed in applying an air quality model. The list of tests is intended for illustrative purposes. The identified tests are not mandatory, nor is the list a comprehensive one.

Model Setup

-Change boundary conditions (is domain size adequate? do I need to consider using a nested regional model?).

-Alter initial conditions (do I need to extend the ramp-up period I have selected?).

Performance Evaluation/Troubleshooting

-Alter grid cell size and/or number of layers considered (how are predicted component predictions affected?).

-Perturb specific inputs (e.g., mixing height, cloud cover, etc.) which might explain why certain processes are identified as important by process analysis (see Section 16.2.2) (are results affected by perturbations within reasonable bounds? what additional measurements should I try to make for a better subsequent review?).

16.2.2 Use of Process Analysis

Occasionally a review of a graphical display, like a tile diagram, may indicate a limited number of locations or incidents which bear further investigation. Diagnostic tests may be used to perform focused analyses on these sites or incidents. These tests entail a more detailed look at a time series of predictions and (if available) observations at or above a site, including chemical species, winds and mixing. The examinations can be done qualitatively. However, more quantification is possible using the second type of diagnostic test described at the beginning of this subsection. A procedure called “process analysis” is an example of the second type of diagnostic test. Process analysis has been used to assess relative importance of various model assumptions as well as simulated physical and chemical phenomena contributing to a predicted ozone concentration at a particular time and location (Jeffries, 1997, Jeffries, et al., 1996, Jang, et al., 1995, Lo and Jeffries, 1997). The procedure works by breaking down a modeled simulation into a sequence of physical and chemical processes leading to a predicted concentration at a given time and location. Since models used to simulate ozone and secondary

particulate matter are similar, process analysis should also be useful for addressing $PM_{2.5}$- and visibility-related applications.

Process analysis requires a substantial amount of expertise to be interpreted to full advantage. However, useful insights are also possible with less detailed analyses. The procedure focuses on selected grid cells. Process analysis then takes advantage of the fact that a numerical grid model addresses physical and chemical factors affecting ozone in a sequential manner. For example, a typical sequence followed in a model for each time step (e.g., 1 hour) might be (1) advection of $PM_{2.5}$ components and precursors present at the beginning of the time step, (2) $PM_{2.5}$ and precursor emissions added during the time step, (3) vertical diffusion of the advected material and fresh emissions, (4) estimated cloud cover and its effects on photolysis rates, (5) atmospheric chemistry involving advected and diffused material with fresh emissions, and (6) deposition of certain compounds. Process analysis examines incremental effects on changes in component and/or $PM_{2.5}$ predictions from hour to hour attributable to each of the processes described above. In this way, one gets a sense of how important each process is as a contributor to predicted air quality at a specified time and location.

If a focused diagnostic analysis, such as one obtained with process analysis, suggests a particular model prediction may be an artifact of a model assumption rather than a result of real chemical/physical atmospheric processes, States may wish to go back to the meteorological or emissions model to verify that the inputs and assumptions that have been used are correct. If a prediction is the result of an apparent artifact which cannot be resolved, States may discount that prediction in the attainment or reasonable progress demonstration.

Table 16.2 shows examples of diagnostic tests which may be useful during and before model performance evaluation. The table is intended to illustrate types of analyses. Choice of tests needs to be made on a case by case basis by those implementing the modeling/analysis protocol.

Table 16.3. Potentially Useful Diagnostic Tests At Various Stages Of Modeling

Stage of Modeling	Test(s) (Examples)	Purpose(s)
Model Setup	**-change boundary conditions**	**-is domain size sufficiently large?** **-do I need to use a nested regional model or will an urban scale model suffice?**
Performance Evaluation & Troubleshooting	**-alter specific (uncertain) inputs (e.g., mixing heights, cloud cover).** **-alter grid cell size or number of layers considered.**	**-is agreement with observations improved?** **-what is the effect on other performance tests (e.g., comparisons with results obtained with observational models, weekend/weekday differences, indicator species ratios)?** **-what priorities should I assign to various kinds of improved measurements?**
Focused performance analysis	**-process analysis**	**-do suspicious looking results make physical sense?** **-prioritize needs for additional sensitivity tests**

<u>Recommendations</u>. States should include diagnostic analyses throughout the modeling process used to help understand model performance and to help develop information which may help improve performance. Sensitivity of relative reduction factors to input perturbations should be a prime focus of the tests. Provided capabilities have been properly installed and tested, States may use versions of a model's code which contain capability for tracing importance different phenomena as contributors to predicted ozone concentrations at selected locations.

17.0 How Do I Evaluate Prospective Control Strategies?

One of the major purposes for a conceptual description, described in Section 9.0, is to help a State to focus on a relatively small set of strategies which hold promise for meeting air quality goals for $PM_{2.5}$ or regional haze. In this Section, we identify additional analyses which may help States to select among remaining strategies. These additional analyses address three issues in more or less the following order:

(1) Given the choices I have made for model setup (e.g., grid cell size, domain, days modeled, etc.), what appears to be an effective strategy?

(2) How "robust" is this strategy—does it appear to work for a variety of alternative, reasonable assumptions regarding model inputs or setup?

(3) What associated uncertainty exists regarding likelihood of success for any strategy I might choose?

17.1 What Appears To Be The Most Effective Strategy?

To be "effective" a strategy needs to show attainment of the air quality goal. This may be done using the recommended modeled attainment or reasonable progress tests, or through a weight of evidence analysis which includes these tests. In the case of the reasonable progress goal for days with good visibility, attainment of the goal may be shown through a weight of evidence analysis which does not include modeling provided concentrations of particulate matter on such days are already very low so that they are approach natural background.

The following analyses may be useful in helping to select among prospective strategies.

- If the conceptual description suggests it will be necessary to reduce secondary components of PM, simulate across-the-board reductions in emissions of SO_2, ammonia, VOC, NOx and combinations of these with a relatively non-resource intensive "engineering model". This should help decide what direction I should be thinking of for my control strategy to reduce secondary particulate matter.

- Compare effects of broad reductions in emissions from point vs. area vs. mobile sources

- Simulate across-the-board reductions in broad source categories (e.g., power generation, motor vehicles, etc.)

- Simulate reductions in boundary conditions in concert with reduced emissions in the area of primary interest to estimate the degree to which my strategy needs help from regional controls in order to succeed.

Although we have identified a prerequisite for a strategy to be effective (i.e., it meets the

air quality goal). A number of factors which fall outside the scope of this guidance (e.g., costs, political feasiblity, enforceability, etc.) are also important determinants of a strategy's effectiveness. The relative importance of these needs to be considered on a case by case basis.

17.2 How Robust Is The Strategy I've Tentatively Selected?

A State needs to assure itself that a tentatively selected strategy works under a variety of reasonable assumptions regarding meteorological and emissions inputs. Ideally, a State should also check to see that the success of a strategy does not hinge on having made a single, fortuitous set of decisions regarding model setup (e.g., domain size, number of atmospheric layers considered, etc.). If a State can show that a strategy works under a variety of conditions, the modeling results have greater credence than otherwise. We list some sensitivity tests which may be useful in helping to determine how "robust" a strategy may be. The list is not comprehensive, nor do we require that all these tests be performed.

- Simulate the selected strategy starting with a different current inventory, reflecting reasonable uncertainties in current emissions.

- Simulate the selected strategy, but with different (reasonable) growth projections.

- Consider different (reasonable) combinations of the first two tests.

- Perturb meteorological inputs, like mixing heights or cloud cover, which may be poorly characterized but which earlier analyses have suggested may be important in affecting current predictions. Take care that meteorological changes which are considered continue to reflect a physically consistent combination of meteorological inputs.

- Simulate the selected strategy using different horizontal grid cell sizes and/or a different number of atmospheric layers (i.e., grid cells in the vertical dimension).

17.3 How Certain Can I Be That A Strategy Will Work?

Recall that we are recommending that models be used in concert with measured air quality to estimate whether attainment of future air quality goals is likely. That is,

Future Air Quality = (model response) (current measured air quality).

Uncertainty in estimated values for future air quality arises from at least 3 sources: (1) incomplete knowledge about all of the input data affecting a model's response; (2) inherent limitations in the model itself, and (3) variability in measured air quality data.

Performing tests like the ones described in Section 17.2 provides a qualitative means for addressing uncertainty attributable to incomplete knowledge about input data. That is, the previously described sensitivity tests help identify strategies that are more likely to work under a

variety of plausible input combinations. In addition, corroboratory information obtained in weight of evidence assessments provides reassurance that a strategy will be successful. While these analyses do not quantify uncertainty in future predicted air quality per se, they can be used to reduce the uncertainty over whether a selected control strategy will be successful in meeting an air quality goal.

At the present time (late 2000), we do not see a way to quantify uncertainty due to inherent limitations in a model. However, there are several ways to reduce this uncertainty. First, use an air quality model(s) whose scientific basis is fully and satisfactorily explained in its accompanying documentation. As we have noted, States may find it necessary to use a simplified "engineering" version of a model. Uncertainty inherent in such results may be reduced if it has been shown that the engineering and more complete versions of a model produce similar results under the conditions which are of greatest interest for a particular application.

A means to quantify uncertainty in future air quality estimates attributable to variability in air quality observations may become available shortly. Rao and Hogrefe (2000) have developed an approach which enables one to gain insight into the distribution of future air quality predictions attributable to variability in currently observed air quality (i.e., such as an observed design value) at a given location. The procedure is to fit a theoretical statistical distribution to the tail of a set of daily observations at a monitoring site (e.g., over a 3-year period) and compute a design value consistent with the form of the NAAQS. The next step is to perform a bootstrapping operation several hundred times to obtain different sets of air quality data. For each instance, a design value is determined from the resulting data. This leads to a distribution of current design values which can be translated into a distribution of future air quality estimates using the RRF approach recommended in this guidance. Work thus far has focused on the 1-hour and 8-hour NAAQS for ozone. However, it may be possible to apply the methodology to PM-related applications.

> **<u>Recommendations</u>. States should use a previously developed conceptual description to identify a relatively narrow set of strategy choices for attaining an air quality goal. Choices may be narrowed further by performing several additional modeling analyses (e.g., across-the board reductions, etc.) requiring fewer resources than a full fledged analysis. Uncertainty regarding likelihood that a chosen strategy will be successful can be reduced by choosing a model based on sound science with good accompanying documentation, performing several sensitivity tests to ensure a chosen strategy works under a variety of conditions and by considering year to year variability in air quality observations at key monitoring sites considered in the demonstration.**

18.0 References Cited

Alpine Geophysics, Inc., (1995), *The Emissions Modeling System (EMS-95) User's Guide*, Alpine Geophysics, Inc., Boulder CO.

Ansari, A.S. and S.N. Pandis, (1998), "Response of Inorganic PM to Precursor Concentrations", *Environmental Science and Technology 32*, pp.2706-2714.

Blanchard, C.L., P.M. Roth, S.J. Tanenbaum, S.D. Ziman and J.H. Seinfeld, (2000), "The Use of Ambient Measurements to Identify Which Precursor Species Limit Aerosol Nitrate Formation", Accepted for publication in *J. Air & Waste Management Association.*

California Air Resources Board, (1993), *Guidance for Using Air Quality-Related Indicators in Reporting Progress in Attaining the State Ambient Air Quality Standards*, (September 1993).

California Air Resources Board, (1999), *California Regional $PM_{10}/PM_{2.5}$ Air Quality Study: Objectives and Associated Data Analysis and Modeling Approaches,* available at ***http://www.arb.ca.gov/airways/crpaqs/***

Causley, M.C., J.L. Fieber, M. Jimenez and L. Gardner, (1990), *User's Guide for the Urban Airshed Model, Volume IV: User's Manual for the Emisssions Preprocessor System*, EPA-450/4-90-007D, U.S. EPA, Research Triangle Park, NC 27711, (NTIS No.: PB91-131250).

Chow, J.C., J.G. Watson, Z. Lu, D.H. Lowenthal, C.A. Frazier, P.A. Solomon, R.H. Thuillier and K.Magliano, (1996), "Descriptive Analysis of $PM_{2.5}$ and PM_{10} at Regionally Representative Locations During SJVAQS/AUSPEX", *Atmospheric Environment 30,* (12), pp.2079-2112.

Cimorelli, A.J., S.G. Perry, A. Venkatram, J.C. Weil, R.J. Paine, R.B. Wilson, R.F. Lee and W.P. Peters, (1998), *AERMOD: Description of Model Formulation*, Prepared for U.S. EPA, Research Triangle Park, NC, (12/15/98 Draft Report), 113 pp.

Cohn, R.D., B.K. Eder, S.K. LeDuc and R.L. Dennis, (2000), "Development of an Aggregation and Episode Selection Scheme to Support the Models-3 Community Multiscale Air Quality Model", *J. Applied Meteorology*, in press.

Cox, W.M. and S. H. Chu, (1993), "Meteorologically Adjusted Ozone Trends in Urban Areas: A Probabilistic Approach", *Atmospheric Environment, 27B*, (4), pp.425-434.

Cox, W.M. and S.H. Chu, (1996), "Assessment of Interannual Variation in Urban Areas from a Climatological Perspective", *Atmospheric Environment, 30*, pp.2615-2625.

Dabdub, D., L. DeHaan, N. Kumar, F. Lurmann and J.H. Seinfeld, (1998), *Computationally Efficient Acid Deposition Model for California*, Draft Report 92-304, California Air Resources Board, Sacramento, CA, available on CARB website, ***http://www.arb.ca.gov/airways/Documents/reports/***.

Dennis, R.L., (1994), "Using the Regional Acid Deposition Model to Determine the Nitrogen Deposition Airshed of the Chesapeake Bay Watershed", *Ch.21, Atmospheric Deposition of Contaminants to the Great Lakes and Coastal Waters*, J.E. Baker (ed.), SETAC Press, Pensacola, FL.

Deuel, H.P. and S.G. Douglas, (1996), *Regional Ozone Patterns in the Eastern U.S. 1985-1995: Statistical Pattern Analysis and Classification and Regression Tree (CART) Analysis in Support of the Ozone Transport Assessment Group (OTAG) Modeling Effort*, Prepared for Southeast Modeling Center by Systems Applications International, Inc., SYSAPP-96/50, San Rafael, CA., (October 1996).

Deuel, H.P. and S.G.Douglas, (1998), *Episode Selection for the Integrated Analysis of Ozone, Visibility and Acid Deposition for the Southern Appalachian Mountains*, Draft Technical Report Systems Applications International, Inc. (SYSAPP-98/07), prepared for the Southern Appalachians Mountains Initiative.

Dolwick, P.D., C.Jang and B.Timin, (1999), "Effects of Vertical Grid Resolution on Photochemical Grid Modeling", *Paper 99-309*, Presented at 99th AWMA Meeting, St.Louis, MO, (June 1999).

Douglas, S.G., R.C. Kessler and E.L. Carr, (1990), *User's Guide for the Urban Airshed Model, Volume III: User's Manual for the Diagnostic Wind Model*, EPA-450/4-90-007C, U.S. EPA, Research Triangle Park, NC 27711, (NTIS No.: PB91-131243).

ENVIRON International Corporation, (1997), *User's Guide to the Comprehensive Air Quality Model with Extensions (CAMx)*, Novato, CA, April 17, 1997.

Eckhoff, P.A. and T.N. Braverman, (1995), *Addendum to the User's Guide to CAL3QHC Version 2.0 (CAL3QHCR User's Guide)*, U.S. EPA, OAQPS, Research Triangle Park, NC 27711, available on U.S. EPA website at ***http://www.epa.gov/ttn/scram/.***

Flaum, J.B., S.T. Rao and I.G. Zurbenko, (1996), "Moderating the Influence of Meteorological Conditions on Ambient Ozone Concentrations", *J.Air & Waste Management Association, 46,* pp.35-46.

Gao, D., W.R.Stockwell and J.B.Milford, (1996), "Global Uncertainty Analysis of a Regional Scale Gas-Phase Chemical Mechanism", *J.Geophys.Res. 101*, pp.9107-9119.

Geron, C., A.Guenther and T.Pierce, (1994), "An Improved Model for Estimating Emissions of Volatile Organic Compounds from Forests in the Eastern United States", *J.Geophys.Res. 99*, pp.12,773-12,791.

Geron, C., R. Rasmussen, R.R. Arnts and A. Guenther, (2000), "A Review and Synthesis of Monoterpene Speciation from Forests in the United States", *Atmospheric Environment 34*, pp.1761-1781.

Gillani, N.V., J.F. Meagher, R.J. Valente, R.E. Imhoff, R.L. Tanner and M. Luria, (1998), "Relative Production of Ozone and Nitrates in Urban and Rural Power Plant Plumes: 1. Composite Results Based on Data From 10 Field Measurement Days", *J. Geophysical Research 103*, No.D17, pp.22,593-22,615, (September 20, 1998).

Gray, H.A. and G.R. Cass, (1986), "Characteristics of Atmospheric Organic and Elemental Carbon Particle Concentrations in Los Angeles", *Environmental Science and Technology 20*, pp**.** 580-589**.**

Grell, G.A., J.Dudhia and D.R. Stauffer, (1994), *A Description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5)*, NCAR/TN-398+STR, 138 pp.

Guenther, A., C. Geron, T. Pierce, B. Lamb, P. Harley and R. Fall, (2000), "Natural Emissions of Non-Methane Volatile Organic Compounds, Carbon Monoxide and Oxides of Nitrogen from North America", *Atmospheric Environment 34*, pp.2205-2230. Also available from NARSTO website, ***http://www.cgenv.com/Narsto/***.

Haney, J.L. and S.G. Douglas, (1996), *Analysis of the Effects of Grid Resolution on UAM-V Simulation Results for the Atlanta Nonattainment Area*, Systems Applications International Technical Memorandum SYSAPP-96/68, prepared for Southeast Modeling Center participants, October 21, 1996.

Henry, R.C., C.W. Lewis and J.F. Collins, (1994), "Vehicle-related Hydrocarbon Source Compositions from Ambient Data: The GRACE/SAFER Method", *Environmental Science and Technology 28*, pp.823-832.

Henry, R.C., (1997), "History and Fundamentals of Multi-variate Air Quality Receptor Models", *Chemometrics and Intelligent Laboratory Systems 37*, pp.37-42.

Henry, R.C., (1997a), "Receptor Model Applied to Patterns in Space (RMAPS) Part I. Model Description", *J. Air & Waste Management Association 47*, p.216.

Henry, R.C. (1997b), "Receptor Model Applied to Patterns in Space (RMAPS) Part II. Apportionment of Airborne Particulate Sulfur from Project MOHAVE", *J. Air & Waste Management Association 47*, p.220.

Henry, R.C. (1997c), "Receptor Model Applied to Patterns in Space (RMAPS) Part III. Apportionment of Airborne Particulate Sulfur in Western Washington State", *J. Air & Waste Management Association 47*, p.226.

Henry, R.C., (2000), personal communication. User's manual and software may be requested by addressing requests electronically to ***rhenry@usc.edu.***

Hogrefe, C., S.T. Rao, I.G. Zurbenko and P.S. Porter, (2000), "Interpreting the Information in Ozone Observations and Model Predictions Relevant to Regulatory Policies in the Eastern United States", Accepted for publication in *Bulletin of the American Meteorological Society*.

Huang, S., K.A. Rahn and R. Arimoto, (1999), "Testing and Optimizing Two Factor-analysis Techniques on Aerosol at Narragansett, Rhode Island", *Atmospheric Environment 33*, pp.2169-2185.

Jang, J.C., H.E.Jeffries and S.Tonnesen, (1995), "Sensitivity of Ozone Model to Grid Resolution Part II: Detailed Process Analysis for Ozone Chemistry", *Atmospheric Environment 29* (21), p.3114.

Jeffries, H.E., T.Keating and Z.Wang, (1996), "Integrated Process Rate Analysis of the Impact of NOx Emission Height on UAM-modeled Peak Ozone Levels", Final Topical Report, Prepared for Gas Research Institute, Chicago, IL. Available at the following Internet address: ***"ftp://airsite.unc.edu/pdfs/ese_unc/jeffries/reports/gri"***, then download file "uncgri.pdf".

Jeffries, H.E., (1997), "Use of Integrated Process Rate Analyses to Perform Source Attribution for Primary and Secondary Pollutants in Eulerian Air Quality Models", Presented at U.S. EPA Source Attribution Workshop, Research Triangle Park, NC, July 16-18, 1997. Available at the following Internet websites, ***"ftp://airsite.unc.edu/pdfs/ese_unc/jeffries/ipradocs",*** then download file ***"sourcewksp.pdf"***, and ***"http://www.epa.gov/ttn/faca/stissu.html"*** and then download file, ***"Source Att WS-Proc.analysis/mass track--Jeffries"***.

Kumar, N., M.T.Odman and A.G. Russell, (1994), "Multiscale Air Quality Modeling: Application to Southern California", *J.Geophysical Research 99*, pp.5385-5397.

Kumar, N. and A.G. Russell, (1996), "Multiscale Air Quality Modeling of the Northeastern United States", *Atmospheric Environment, 30,* pp.1099-1116.

Kumar, N. and F.W. Lurmann, (1997), "Application of a Three-Dimensional Aerosol Model (UAM-AERO) to South Coast Air Basin", *Paper 97-WP93.01*, presented at 90th annual AWMA meeting, Toronto, Ontario, (June 1997).

LADCO, (1999), Personal communication, Michael Koerber.

LADCO, (2000), Personal communication, Mark Janssen.

Lo, C.S. and H.E. Jeffries, (1997), "A Quantitative Technique for Assessing Pollutant Source Location and Process Composition in Photochemical Grid Models", Presented at annual AWMA Meeting, Toronto, Ontario (1997). Also available at the following Internet address, ***"ftp://airsite.unc.edu/pdfs/ese_unc/jeffries/ipradocs"***, then download file ***"PCA_awma.pdf"***.

Lu, C.H. and J.S. Chang, (1998), "On the Indicator-Based Approach to Assess Ozone Sensitivities and Emissions Features", *J. Geophysical Research 103*, pp.3453-3462.

Lurmann, F.W. and N. Kumar, (1996), *Development of Chemical Transformation Algorithms for Annual PM_{10} Dispersion Models*, draft report, STI-95410-1602-DFR, prepared for the South Coast Air Quality Management District, Diamond Bar, CA.

Lurmann, F.W., A.S. Wexler, S.N. Pandis, S. Musarra, N. Kumar and J.H. Seinfeld, (1997), "Modeling Urban and Regional Aerosols: II. Application to California's South Coast Air Basin", *Atmospheric Environment 31*, pp.2695-2715.

Lyons, W.A., C.J. Tremback and R.A. Pielke, (1995), "Applications of the Regional Atmospheric Modeling System (RAMS) to Provide Input to Photochemical Grid Models for the Lake Michigan Ozone Study (LMOS)", *J.Applied Meteorology 34*, 1762-1786.

Malm, W.C., J.F. Sisler, D. Huffman, R.A. Eldred and T.A. Cahill, (1994), "Spatial and Seasonal Trends in Particle Concentration and Optical Extinction in the United States", *J.Geophysical Research 99*, NO.D1, pp.1347-1370.

Malm, W.C., M.L. Pitchford, M. Scruggs, J.F. Sisler, R. Ames, S. Copeland, K.A. Gebhart and D.E. Day, (2000), *Spatial and Seasonal Patterns and Temporal Variability of Haze and its Constituents in the United States, Report III*, ISSN: 0737-5352-47, Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO.

Malm, W.C., (2000a), Personal communication, 5/18/00 E-mail note to Marc Pitchford, Subject: "RE: EPA calculations".

MCNC, (2000), Internet website describing the MAQSIP air quality model, available at ***http://www.iceis.mcnc.org/products/maqsip/***.

MCNC, (2000a), Internet website describing the SMOKE emissions model. ***http://envpro.ncsc.org/products/smoke/***.

Meyer, E.L., K.W.Baldridge, S.Chu and W.M.Cox, (1997), "Choice of Episodes to Model: Considering Effects of Control Strategies on Ranked Severity of Prospective Episode Days", *Paper 97-MP112.01*, Presented at 97th Annual AWMA Meeting, Toronto, Ontario, (June 1997).

Milanchus, M.L., S.T. Rao, I.G. Zurbenko, (1997), "Discerning Changes in Anthropogenic Ozone in the Presence of Meteorological Fluctuations", *Environmental Management* (July 1997), pp.17-21.

Milanchus, M.L., S.T. Rao, I.G. Zurbenko, (1998), "Evaluating the Effectiveness of Ozone Management Efforts in the Presence of Meteorological Variability", *J. Air & Waste Management Association, 48*, pp.201-215.

NOAA, (1999), HY-SPLIT model, available at ***www.arl.noaa.gov/ss/models/hysplit.html***.

Odman, M.T. and C.L. Ingram, (1996), *Multiscale Air Quality Simulation Platform (MAQSIP): Source Code Documentation and Validation*, MCNC Technical Report ENV-96TR002-v1.0, 83pp.

Paatero, P. and U. Tapper, (1994), "Positive Matrix Factorization: A Non-negative Factor Model with Optimal Utilization of Error Estimates of Data Values", *Environmetrics 5*, pp.111-126.

Pai, P., K. Vijayaraghavan and C. Seigneur, (2000), "Particulate Matter Modeling in the Los Angeles Basin Using SAQM-AERO", *J.Air & Waste Management Assoc. 50*, pp.32-42.

Paine, R.J., R.F. Lee, R.W. Brode, R.B. Wilson, A.J. Cimorelli, S.G. Perry, J.C. Weil, A. Venkatram and W.D. Peters, (1998), *Model Evaluation Results for AERMOD*, Prepared for U.S. EPA, Research Triangle Park, NC, (12/17/98 Draft Document).

Pielke, R.A., et al., (1992), "A Comprehensive Meteorological Modeling System-RAMS", *Meteor.Atmos.Phys., 49* , pp.69-91.

Pitchford, M.L. and W.C. Malm, (1994), "Development and Applications of a Standard Visual Index", *Atmospheric Environment 28* (5), pp.1049-1054.

Poirot, R., (1998), *Air Mass History Pollution Climatology for Northeastern Forests and Parks*, available at ***http://capita.wustl.edu/NEARDAT/Reports/TechnicalReports/ForestSer_TrajProp/fstrjsum.htm***

Polissar, A.V., P.K. Hopke, P. Paatero, W.C. Malm and J.F. Sisler, (1998), "Atmospheric Aerosol Over Alaska 2. Elemental Composition and Sources", *J. Geophysical Research 103,* (Number D15), pp.19045-19057.

Pun, B. and C. Seigneur, (1998), *Conceptual Model of Particulate Matter Pollution in the California San Joaquin Valley*, Report CP045-1-98 prepared for Pacific Gas and Electric Co., (September 8, 1998), available at California Air Resources Board website, ***http://www.arb.ca.gov/airways/Documents/reports/sjvpmc~1.pdf***.

Pun, B.K. and C. Seigneur, (1999), "Understanding Particulate Matter Formation in the California San Joaquin Valley: Conceptual Model and Data Needs", *Atmospheric Environment 33,* pp.4865-4875..

Rao, S.T. and I.G. Zurbenko, (1994), "Detecting and Tracking Changes in Ozone Air Quality", *J. Air & Waste Management Association, 44,* pp.1089-1092.

Rao, S.T. and C. Hogrefe, (2000), "Demonstrating Attainment of the Air Quality Standards: Deterministic vs. Probabilistic Modeling", presentation made to U.S. EPA, Research Triangle Park, NC (November 16, 2000).

Reynolds, S.D., H.M.Michaels, P.M.Roth, T.W.Tesche, D.McNally, L.Gardner and G.Yarwood, (1996), "Alternative Base Cases in Photochemical Modeling: Their Construction, Role and Value", *Atmospheric Environment*, *30*, (12), pp.1977-1988.

Schauer, J.J., W.F. Rogge, L.M. Hildemann, M.A. Mazurek, G.R. Cass and B.R.T. Simoneit, (1996), "Source Apportionment of Airborne Particulate Matter Using Organic Compounds as Tracers", *Atmospheric Environment 30*, pp.3837-3855.

Schichtel, B.A. and R. Husar, (1995), *Regional Simulation of Atmospheric Pollutants with the Capita Monte Carlo Model*, Center for Air Pollution and Trend Analysis, Washington University, St.Louis, MO. Available at "***http://capita.wustl.edu/CAPITA/CapitaReports/MonteCarlo/MonteCarlo.html***".

Schichtel, B.A., (1999), *Local and Regional Contributions of Fine Particulate Mass to Urban Areas in the mid-Atlantic and Southwestern U.S.,* Center of Air Pollution Impact and Trend Analysis, Washington University, St.Louis, MO., Contract number 7D-0869-NAEX, available at "***http://capita.wustl.edu/CAPITA/CapitaReports/PMvsWnd.html***".

Scire, J.S., F.R. Francoise, M.E.Fernau and R.J.Yamartino, (1998), *A User's Guide for the CALMET Meteorological Model (Version 5.0)*, EarthTech., Inc., Concord, MA.

Seaman, N.L., et al., (1995), "A Multi-Scale Four-Dimensional Data Assimilation System Applied to the San Joaquin Valley During SARMAP. Part I: Modeling Design and Basic Performance Characteristics", *J.Applied Meteorology 34*, pp.1739-1761.

Seaman, N.L. and D.R. Stauffer, (1996), *SARMAP Meteorological Model. Final Report to San Joaquin Valleywide Air Pollution Study Agency*, Technical Support Division, California Air Resources Board, Sacramento, CA., 174 pp.

Seaman, N.L., et al., (1996b), "Application of the MM5-FDDA Meteorological Model to the Southern California SCAQS-1997 Domain: Preliminary Test Using the SCAQS August 1987 Case", *Ninth Joint Conference on Applications of Air Pollution Meteorology*, American Meteorological Society, Atlanta, GA., January 28-February 2, 1996.

Seaman, N.L., (1997), "Use of Model-Generated Windfields to Estimate Potential Relative Mass Contributions from Different Locations", Prepared for *U.S. EPA Source Attribution Workshop*, Research Triangle Park, NC, July 16-18, 1997. Available from EPA Internet Website, ***"http://www.epa.gov/ttn/faca/stissu.html"***, then download file ***"source attribution workshop materials".***

Seigneur, C., P. Pai, J. Louis, P. Hopke and D. Grosjean, (1997), *Review of Air Quality Models for Particulate Matter*, Document Number CP015-97-1b, Prepared for the American Petroleum Institute, Available by accessing "***http://www.api.org/ehs/Publications/4669.pdf***".

Seigneur, C., P. Pai, P.K. Hopke and D. Grosjean, (1999), "Modeling Atmospheric Particulate Matter", *Environmental Science and Technology/News,* Feb.1, 1999, pp.80-86.

Seigneur, C., (2000), Personal Communication, 7/10/00 note to Ned Meyer, U.S. EPA.

Seigneur, C., B. Pun, P. Pai, J.F. Louis, P. Solomon, C. Emery, R. Morris, M. Zahniser, D. Worsnop, P. Koutrakis, W. White and I. Tombach, (2000a), "Guidance for the Performance Evaluation of Three-dimensional Air Quality Modeling Systems for Particulate Matter and Visibility", *J. Air & Waste Management Assoc., 50*, pp.588-599.

Seigneur, C., (2000b), "Current Status of Air Quality Models for Particulate Matter", *NARSTO 2000—Tropospheric Aerosols: Science and Decisions in an International Community,* (October 23-26, 2000, Queretaro, Mexico.

Sillman, S., (1995), "The Use of NOy, H_2O_2 and HNO_3 as Indicators for O3-NOx-ROG Sensitivity in Urban Locations", *J.Geophysical Research 100*, pp.14175-14188.

Sillman, S. (1998), *Evaluating the Relation Between Ozone NOx and Hydrocarbons: The Method of Photochemical Indicators*, EPA/600R-98/022.

Sisler, J.F. (Editor), (1996), *Spatial and Seasonal Patterns and Long Term Variability of the Composition of the Haze in the United States: An Analysis of Data from the IMPROVE Network*, Cooperative Institute for Research in the Atmosphere Report ISSN: 0737-5352-32, Colarado State University, Fort Collins, CO.

Solomon, P.A., T. Fall, L. Salmon, G.R. Cass, H.A. Gray and A. Davison, (1989), "Chemical Characteristics of PM_{10} Aerosol Collected in the Los Angeles Area", *J.Air Pollution Control Association 39,* pp.154-163.

Stoeckenius, T.E., (1990), *Adjustment of Ozone Trends for Meteorological Variation*, Paper presented at the Air & Waste Management Asssociation Speciality Conference on Tropospheric Ozone and the Environment, Los Angeles, CA (March 19-22, 1990).

Strimaitis, D.G., J.S. Scire and J.C. Chang, (1998), "Evaluation of the CALPUFF Dispersion Model with Two Power Plant Data Sets", paper presented at the Tenth Joint Conference on the Application of Air Pollution Meteorology, Phoenix, AZ, (January 11-16, 1998), American Meteorological Society, Boston, MA.

Systems Applications International, (1998), *User's Guide to the Regulatory Modeling System for Aerosols and Deposition (REMSAD)*, SYSAPP98-96/42r2, prepared for the U.S. EPA under contract 68D30032. Available on U.S.EPA website at ***http://www.epa.gov/scram001/***.

Systems Applications International, (2000), Personal communication, ongoing work assignment with U.S. EPA, OAQPS, Tom Braverman, EPA Work Assignment Manager.

Tang, I., (1996), "Chemical and Size Effects of Hygroscopic Aerosols on Light Scattering Coefficients", *J.Geophysical Research 101* No.D14, pp.19,245-19,250.

Tesche, T.W. and D.E. McNally, (1993a), *Operational Evaluation of the SARMAP Meteorological Model (MM5) for Episode 1: 3-6 August 1990*, Prepared for the Valley Air Pollution Study Agency by Alpine Geophysics, Crested Butte, CO.

Tesche, T.W. and D.E. McNally, (1993b), *Operational Evaluation of the SARMAP Meteorological Model (MM5) for Episode 2: 27-29 July 1990*, Prepared for the Valley Air Pollution Study Agency by Alpine Geophysics, Crested Butte, CO.

Tesche, T.W. and D.E. McNally, (1993c), *Operational Evaluation of the CAL-RAMS Meteorological Model for LMOS Episode 1: 26-28 June, 1991*, prepared for the Lake Michigan Air Directors Consortium by Alpine Geophysics, Crested Butte, CO.

Tesche, T.W. and D.E. McNally, (1993d), *Operational Evaluation of the CAL-RAMS Meteorological Model for LMOS Episode 2: 17-19 July, 1991*, prepared for the Lake Michigan Air Directors Consortium by Alpine Geophysics, Crested Butte, CO.

Tesche, T.W. and D.E. McNally, (1993e), *Operational Evaluation of the CAL-RAMS Meteorological Model for LMOS Episode 3: 25-26 August, 1991*, prepared for the Lake Michigan Air Directors Consortium by Alpine Geophysics, Crested Butte, CO.

Tesche, T.W. and D.E. McNally, (1993f), *Operational Evaluation of the CAL-RAMS Meteorological Model for LMOS Episode 4: 20-21 June, 1991*, prepared for the Lake Michigan Air Directors Consortium by Alpine Geophysics, Crested Butte, CO.

TNRCC, (2000), Personal communication, E-mail note from Mark Estes, TNRCC to Ned Meyer, U.S. EPA, August 17, 2000.

Turpin, B. and H-J. Lim, (2000), "Species Contributions to $PM_{2.5}$ Mass Concentration: Revisiting Common Assumptions for Estimating Organic Mass", *Aerosol Science and Technology, 33*, in press.

U.S. EPA, (1990), *Receptor Model Technical Series, Volume III (1989 Revision): CMB7 User's Manual*, EPA-450/4-90-004.

U.S. EPA, (1991), *Guideline for Regulatory Application of the Urban Airshed Model*, EPA-450/4-91-013.

U.S. EPA, (1991a), *Procedures for Preparing Emissions Projections*, EPA-450/4-91-014.

U.S. EPA, (1993), *Volatile Organic Compound (VOC)/Particulate Matter (PM) Speciation Data System (SPECIATE), Version 1.5*, EPA/C-93-013.

U.S. EPA, (1993a), *User's Guide for the Urban Airshed Model, Volume IV: User's Manual for the Emissions Preprocessor System 2.0, Part A: Core FORTRAN System and Part B: Interface and Emission Display System*, EPA-450/4-90-007d(R), NTIS Accession No. PB93-122380.

U.S. EPA, (1996), Office of Mobile Sources, *User's Guide for MOBILE5*, EPA-AA-AQAB-94-01, updated as version 5b, see website, ***http://www.epa.gov/otaq/models.htm***.

U.S. EPA, (1996a), *Photochemical Air Monitoring System 1996 Data Analysis Results Report*, EPA-454/R-96-006.

U.S. EPA, (1996b), *Guidance on Use of Modeled Results to Demonstrate Attainment of the Ozone NAAQS,* EPA-454/B-95-007 (June 1996).

U.S. EPA, (1997), "NAAQS for Particulate Matter", *Federal Register 62* , p.38652, (July 18, 1997).

U. S. EPA, (1997a), *Biogenics Sources Preferred and Alternative Methods*, July 1997, EPA-454/R-97-004e.

U. S. EPA, (1997b), *EIIP Volume I, Introduction and Use of EIIP Guidance for Emission Inventory Development*, July 1997, EPA-454/R-97-004a**.**

U. S. EPA, (1997c), *Point Sources Preferred and Alternative Methods*, July 1997, EPA-454/R-97-004b**.**

U. S. EPA, (1997d), *Area Sources Preferred and Alternative Methods*, July 1997, EPA-454/R-97-004c**.**

U. S. EPA, (1997e), *Mobile Sources Preferred and Alternative Methods*, July 1997, EPA-454/R-97-004d.

U. S. EPA, (1997f), *Quality Assurance Procedures*, July 1997, EPA-454/R-97-004f.

U. S. EPA, (1997g), *Data Management Procedures*, July 1997, EPA-454/R-97-004g.

U.S. EPA, (1998), *EPA Third-Generation Air Quality Modeling System, Models-3 Volume 9b Users Manual*, EPA-600/R-98/069(b).

U.S. EPA, (1998a), *National Air Pollutant Emission Trends, Procedures Document 1900-1996*, EPA-454/R-98-008, see ***http://www.epa.gov/ttn/chief/ei/*** .

U.S. EPA, (1999), "Regional Haze Regulations: Final Rule", *Federal Register 64,* pp.35714-35774, (July 1, 1999).

U.S. EPA, (1999a), *Guideline on Data Handling Conventions for the PM NAAQS*, EPA-454/R-99-008, (April 1999).

U.S. EPA, (1999b), *Draft Guidance on the Use of Models and Other Analyses in Attainment Demonstrations for the 8-hour Ozone NAAQS*, EPA-454/R-99-004, (May 1999).

U.S. EPA, (1999c), *Science Algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) Modeling System*, EPA/600/R-99/030.

U.S. EPA, (1999d), *Emission Inventory Guidance For Implementation Of Ozone And Particulate Matter National Ambient Air Quality Standards (NAAQS) and Regional Haze Regulations*, EPA-454/R-99-006, (April 1999).

U.S. EPA, (1999e), *Particulate Emissions*, September 1999, EPA-454/R-97-004i.

U.S. EPA, (1999f), *Emission Projections*, December 1999, EPA-454/R-97-004j.

U.S. EPA, (2000), *Criteria Docuent for Particulate Matter (draft)*.

U.S. EPA (2000a), Efforts to obtain climatological mean humidity correction factors, in preparation, Tom Rosendahl, U.S. EPA project manager.

U.S. EPA, (2000b), Proposed revision to 40CFR, Part 51, Appendix W.

U.S. EPA, (2001), *$PM_{2.5}$ Data Analysis Workbook*, In preparation, Jim Szykman, U.S. EPA project manager.

Willis, R.D., (2000), *Workshop on UNMIX and PMF as Applied to $PM_{2.5}$,* EPA/600/A-00/048.

Wilson, W.E. and H.H.Suh, (1997), "Fine Particles and Coarse Particles: Concentration Relationships Relative to Epidemiologic Studies", *J.Air & Waste Management Association 47*, pp.1238-1249.

Yang, Y.J., W.R. Stockwell, J.B.Milford, (1995), "Uncertainties in Incremental Reactivities of Volatile Organic Compounds", *Environmental Science and Technology 29*, pp.1336-1345.

Yang, Y.J., J.G.Wilkinson and A.G.Russell, (1997), "Fast, Direct Sensitivity Analysis of Multidimensional Air Quality Models for Source Impact Quantification and Area-of-Influence Identification", Prepared for the U.S. EPA Source Attribution Workshop, Research Triangle Park, NC, July 16-18, 1997. Available from EPA Internet website, ***"http://www.epa.gov/ttn/faca/stissu.html"***, then download file "SAW-Direct Sensi Analys.--PowerP.--T.Russell".

Yang, Y.J., J.G.Wilkinson and A.G.Russell, (1997a), "Fast, Direct Sensitivity Analysis of Multidimensional Photochemical Models", *Environmental Science and Technology 31*, pp.2859-2868.

Yarwood, G. and R.Morris, (1997), "Description of the CAMx Source Attribution Algorithm", Prepared for U.S. EPA Source Attribution Workshop, Research Triangle Park, NC, July 16-18, 1997. Available from EPA Internet website, ***"http://www.epa.gov/ttn/faca/stissu.html"***, then download file "source attribution workshop materials".

Yarwood, G., G.Wilson, R.E.Morris, M.A.Yocke, (1997a), *User's Guide to the Ozone Tool: Ozone Source Apportionment Technology for UAM-IV*, Prepared for Mr. Thomas Chico, South Coast Air Quality Management District, Diamond Bar, CA, March 28, 1997.

Glossary

Attainment demonstration

A modeled attainment demonstration consists of two parts: an analysis estimating emission levels consistent with attainment of the NAAQS, and a list of measures that will lead to the desired emission levels once growth is accounted for. The first (analysis) part consists of a modeled attainment test. It may also include one or more hot spot analyses and a review of a diverse set of model outputs and emissions, air quality and meteorological data in a weight of evidence determination to assess whether attainment of the NAAQS is likely with the proposed control strategy.

Class I areas

Class I areas are geographical locations for which the Clean Air Act requires particular scrutiny to protect "air quality-related values". One of these values is visibility. Generally, Class I areas are locations of particular scenic beauty, locations of historical significance or locations which are to be preserved as wilderness. They include the nation's national parks, national monuments and wilderness areas. There are 156 Class I areas in the United States. These are identified on a map of the U.S. contained in the final rule on regional haze regulations (see Regional Haze Regulations; Final Rule, *Federal Register 64*, No.126, July 1, 1999, p.35716.).

Coarse mode particulate matter

Coarse mode particulate matter is emitted directly from sources as a result of physical activities like grinding or crushing or resuspension. Coarse mode particles range in size from 1.0 - 20 micrometers or more, with a relatively small fraction of the total mass of coarse mode particulate matter occurring in particles with aerodynamic diameters $\leq$ 2.5 micrometers.

Glossary (continued)

Coarse particulate matter	The measured mass of particles with aerodynamic diameters > 2.5 micrometers and $\leq$ 10 micrometers.
Components (of $PM_{2.5}$)	Major portions of measured mass of $PM_{2.5}$, grouped by common sets of chemical or physical attributes. Components of $PM_{2.5}$ considered in modeled demonstrations should include mass associated with sulfates (SO_4), nitrates (NO_3), organic carbon (OC), elemental carbon (EC), inorganic primary particulate matter (IP) and unattributed portions of measured mass (U).
Conceptual description	A qualitative way of characterizing the nature of an area's nonattainment or regional haze problem. Attributes included in the characterization include roles of primary and secondary particulate matter, relative importance of transport vs. local generation, and mix of major source categories which may be affecting measured concentrations of particulate matter.
Deciview	A measure of visibility derived by taking the logarithm of the ratio of the light extinction coefficient to the light extinction coefficient which results solely due to air (without particles). Deciview units are intended to serve as a linear measure of human perception of visibility degradation. Reasonable progress goals are expressed using deciviews.
Design values	A design value is a concentration defined consistently with the form of the NAAQS for $PM_{2.5}$. If the design value is $\leq$ the concentration defined in the NAAQS, the NAAQS is met. Thus, for the annual NAAQS, if the annual mean concentration, averaged over 3 consecutive years (i.e., the

	design value) is $\leq$ 15.0 $\mu g/m^3$ at a monitor, the NAAQS is met at that site.
Extinction coefficient	This is a measure of light extinction resulting from scattering and absorption of light by particles and gases.
Fine mode particulate matter	This is particulate matter emitted directly as a result of combustion or formed in the atmosphere as a result for reactions among precursors. Fine mode particulate matter is $\leq$ 2.5 micrometers in aerodynamic diameter.
Fine particulate matter	All particles $\leq$ 2.5 micrometers in aerodynamic diameter. This includes all fine mode particulate matter plus a small fraction of emitted coarse mode particulate matter.
Hot spot analysis	A modeling analysis performed for one or a limited number of sources of primary particulate matter. These analyses may be needed to supplement the modeled attainment test for the $PM_{2.5}$ NAAQS. Need for them is most likely for problems related to the 24-hour NAAQS. Analyses should focus on large sources or concentrations of sources situated in locations where there are no monitored concentrations of $PM_{2.5}$ and its components.
Mid-course review	An analysis of air quality and emissions data performed sometime between the submittal of a SIP revision and the date required for attainment or the date required for the current increment of reasonable progress reducing regional haze to be realized. Its purposes are to assess progress, compare observed progress with expectations at the time of the modeled attainment or reasonable progress demonstration and to

	identify possible means to improve progress, if necessary.
Modeled attainment test	The modeled attainment test uses monitored design values for $PM_{2.5}$ together with monitored PM species to estimate current concentrations for 6 major components of $PM_{2.5}$. Air quality models are then used to estimate relative reduction factors for each of the 6 components. A future concentration for each component is the product of its current concentration and calculated relative reduction factor. Future concentrations of each component are added to estimate the future concentration of $PM_{2.5}$. If the resulting future concentration of $PM_{2.5}$ is $\leq$ to the concentration specified in the NAAQS, the test is passed.
Modeled test for reasonable progress	This test first entails ranking current extinction coefficients derived from monitored data for components of $PM_{2.5}$ and coarse particulate matter. An empirical equation, based on a review of monitored data by the National Park Service, is used to derive the extinction coefficients which get ranked. Results are used to identify 20% of days with worst (best) current visibility, along with the mean extinction coefficient corresponding to the identified days with current worst (best) visibility. Air quality models are used to derive relative reduction factors for components of $PM_{2.5}$ and coarse particulate matter. Future concentrations are estimated by multiplying the current observed component concentrations times the corresponding relative reduction factors. The National Park Service equation is re-applied to estimate a future value for the mean worst (best) extinction coefficient. The estimated future worst (best) mean

	extinction coefficient (converted to deciviews) is compared to the current value (in deciviews) to see if the reasonable progress goals will be met by the simulated control strategy.
Modeling system	This is a group of models used to predict air qualilty concentrations. The group includes an emissions model to convert countywide emission information into gridded speciated emissions which vary diurnally and reflect environmental conditions. It also includes a meteorological model to provide gridded meteorological outputs and an air chemistry/deposition model which takes information provided by the emissions and meteorological models and uses it to develop gridded predictions of pollutant concentrations.
Relative reduction factor	The relative reduction factor (RRF) is a key part of the modeled tests for attainment and reasonable progress. It is calculated for each component of $PM_{2.5}$ and, for visibility-related applications, for coarse particulate matter as well. The RRF is a dimensionless ratio of estimated future concentration of a component divided by its current predicted concentration.
$PM_{2.5}$	This is the mass of all measured particulate matter having an aerodynamic diameter $\leq$ 2.5 micrometers. It consists of fine mode as well as some coarse mode particulate matter.
Primary particulate matter	This is all measured particulate matter which is either emitted directly as particulate matter or has been converted to particulate matter, through condensation, within a few dozen meters of its release. Generally, the following components of $PM_{2.5}$ consist of primary particles: IP, EC and portions of

	OC. Coarse particulate matter also consists of primary particulate matter.
Secondary particulate matter	This is all measured particulate matter which results from homogeneous or heterogeneous chemical reactions occurring in the atmosphere after release of gaseous precursors. Generally, the following components of $PM_{2.5}$ consist of secondary particles: SO_4, NO_3 and portions of OC.
Weight of evidence (WOE) determination	This is a set of diverse analyses used to judge whether an air quality goal is likely to be met. The credibility of each analysis is assessed and an oucome consistent with an hypothesis that the goal will be met is identified beforehand. If the set of oucomes, on balance, is consistent with meeting the goal, then the WOE can be used to show that a proposed strategy will be adequate. A weight of evidence determination is used primarily for NAAQS-related applications. However, it may also be needed to assess whether it is likely that the reasonable progress goal for days with current best visibility will be met. Weight of evidence generally includes results from the modeled attainment (or reasonable progress) test, hot spot analyses, as well as other analyses of air quality, emissions and modeled data.

(This page is left blank intentionally)