APPENDIX E

U.S. EPA Guidance on the Use of Models and Other Analyses

Guidance on the Use of Models and Other Analyses for Demonstrating Attainment of Air Quality Goals for Ozone, $PM_{2.5}$, and Regional Haze

EPA -454/B-07-002
April 2007

Guidance on the Use of Models and Other Analyses for Demonstrating Attainment of Air Quality Goals for Ozone, $PM_{2.5}$, and Regional Haze

U.S. Environmental Protection Agency
Office of Air Quality Planning and Standards
Air Quality Analysis Division
Air Quality Modeling Group
Research Triangle Park, North Carolina

ACKNOWLEDGMENTS

We would like to acknowledge contributions from members of an external review group and U.S. EPA Regional Office modeling staffs in providing detailed comments and suggestions regarding the final version of this guidance. In particular, we would like to thank staff members of the Lake Michigan Air Directors Consortium (LADCO), Carolina Environmental Program (CEP), South Coast Air Quality Management District (SCAQMD), California Air Resources Board (CARB), Texas Commission on Environmental Quality (TCEQ), North Carolina Division of Air Quality (NCDAQ), New York State Department of Environmental Conservation (NYSDEC), Georgia Environmental Protection Divsion (GAEPD) and Computer Sciences Corporation (CSC) for testing our ideas for a modeled attainment test and sharing the results with us.

We would also like to acknowledge the contributions and accomplishments of Ned Meyer. Ned wrote the original drafts of the ozone and $PM_{2.5}$ modeling guidance documents. He also developed the relative attainment tests and put his vision on paper. The final version of this guidance is shaped by Ned's words and thoughts.

ACKNOWLEDGMENTS iii

FOREWORD ix

1.0 Introduction 1
- **1.1 What Is The Purpose Of This Document?** 2
- **1.2 Does The Guidance In This Document Apply To Me?** 2
- **1.3 How Does The Perceived Nature Of Ozone, PM, and Regional Haze Affect My Attainment Demonstration?** 3
  - **1.3.1 Ozone, PM, and Regional Haze** 3
  - **1.3.2 How is Measuring and Modeling Particulate Matter Different (and Often More Complicated) than Ozone?** 5
- **1.4 What Topics Are Covered In This Guidance?** 11

Part I. How Do I Use Results Of Models And Other Analyses To Help Demonstrate Attainment? 14

2.0 What Is A Modeled Attainment Demonstration?--An Overview 15
- **2.1 What Is The Recommended Modeled Attainment Test for Ozone and $PM_{2.5}$?- An Overview** 15
- **2.2 What Is The Recommended Modeling Assessment For Regional Haze?- An Overview** 16
- **2.3 What Does A Recommended Supplemental Analysis/Weight Of Evidence Determination Consist Of? --An Overview** 17
- **2.4 Why Should A Model Be Used In A "Relative" Sense And Why May Corroboratory Analyses Be Used In A Weight Of Evidence Determination?** 18

3.0 What Is The Recommended Modeled Attainment Test? 20
- **3.1 Calculating site-specific baseline concentrations.** 21
- **3.2 Identifying surface grid cells near a monitoring site.** 24
- **3.3 Choosing model predictions to calculate a relative response factor $(RRF)_I$ near a monitor.** 26
- **3.4 Estimating design values at unmonitored locations: what is an unmonitored area analysis and why is it needed?** 29
  - **3.4.1 Why does the unmonitored area analysis need to use both ambient data and model output?** 29
  - **3.4.2 Implementation of Gradient adjusted Spatial Fields** 30
  - **3.4.3 Using the Results of the Unmonitored Area Analysis** 32
- **3.5 Which base year emissions inventory should be projected to the future for the purpose of calculating RRFs?** 33
- **3.6 Choosing a year to project future emissions.** 34

4.0 Ozone Attainment Test 37
4.1 Limiting modeled 8-hour daily maxima chosen to calculate RRF. 37
4.2 How Do I Apply The Recommended 8-Hour Ozone Modeled Attainment Test? 40

5.0 What Are The Modeled Attainment Tests For The Two $PM_{2.5}$ NAAQS? 43
5.1 What Is The Recommended Modeled Attainment Test For The Annual NAAQS? 43
5.1.1 What Ambient $PM_{2.5}$ Data is Used in the Attainment Test? 44
5.1.2 FRM Monitors that Don't Have Speciation Data 45
5.1.3 New Species Calculations and Definitions 46
5.1.4 Recommended Treatment of Species Data 47
5.2 What Is The Recommended Modeled Attainment Test For The 24-Hour NAAQS? 56
5.3 Special Considerations for Primary $PM_{2.5}$ 63
5.3.1 Analysis of Primary $PM_{2.5}$ Impacts at Monitors 64

6.0 What Is The Recommended Modeling Analysis for Regional Haze? 75
6.1 Regional Haze Rule Background 75
6.2 Uniform Rate of Progress 76
6.3 How are Regional Haze Trends Measured? 77
6.4 How Do I Apply A Model To Calculate Changes in Visibility ? 83
6.5 How Do I Select Appropriate Inputs For The Uniform Rate of Progress Analysis? 93

7.0 How Can Additional Analyses Can Be Used to Support the Attainment Demonstration? 98
7.1 What Types of Additional Analyses Should Be Completed as Part of the Attainment Demonstration? 99
7.2 What Is Entailed In A Weight Of Evidence Determination? 105
7.3 What Role Should Weight Of Evidence Play In Visibility-Related Analyses? 107

8.0 What Additional Analyses and Data Collection Can Be Completed to Assess Progress towards Attainment 110

9.0 What Documentation Do I Need To Support My Attainment Demonstration? 113

Part II. How Should I Apply Air Quality Models To Produce Results Needed To Help Demonstrate Attainment? 120

10.0 How Do I Apply Air Quality Models?-- An Overview 121

11.0 How Do I Get Started?- A "Conceptual Description" 126
11.1 What Is A "Conceptual Description"? 126
11.1.1 8-Hour Ozone NAAQS 126
11.1.2 Annual $PM_{2.5}$ NAAQS 128
11.1.3 Example Regional Haze Application 131

12.0 What Does A Modeling/Analysis Protocol Do, And What Does Developing One Entail? 133
12.1 What Is The Protocol's Function? 133
12.2 What Subjects Should Be Addressed In The Protocol? 133

13.0 What Should I Consider In Choosing An Air Quality Model? 135
13.1 Types of Air Quality Models 135
13.2 What Prerequisites Should An Air Quality Model Meet To Qualify For Use In An Attainment Demonstration? 136
13.3 What Factors Affect My Choice of A Model For A Specific Application? 137
13.4 What Are Some Examples Of Air Quality Models Which May Be Considered? 138

14.0 How are the Meteorological Time Periods (Episodes) Selected? 140
14.1 What Are The Most Important Criteria For Choosing Episodes? 141
14.1.1 8-Hour Ozone NAAQS 142
14.1.2 Recommendations for the 24-Hour $PM_{2.5}$ NAAQS 147
14.1.3 Annual $PM_{2.5}$ 148
14.1.4 Regional Haze 149
14.2 What Additional, Secondary Criteria May Be Useful For Selecting Time Periods to Model? 150

15.0 What Should Be Considered When Selecting The Size And Horizontal/Vertical Resolution Of The Modeling Domain? 152
15.1 How is the Size of the Modeling Domain Chosen? 152
15.2 How are the Initial and Boundary Conditions Specified? 153
15.3 What Horizontal Grid Cell Size Is Necessary? 155
15.3.1 Horizontal Resolution Recommendations for Ozone 157
15.3.2 Horizontal Resolution Recommendations for the $PM_{2.5}$ NAAQS . . 157
15.3.3 Horizontal Resolution Recommendations for Regional Haze 158
15.4 How Should the Vertical Layers Be Selected? 158

16.0 How are the Meteorological Inputs Prepared for Air Quality Modeling? 160
16.1 What Issues are Involved in the Generation and Acquisition of Meteorological Modeling Data? 160
16.2 How Should the Meteorological Modeling Analysis be Configured? 161
16.3 How Should the Performance of the Meteorological Modeling Be Evaluated? 163

17.0 How Are the Emission Inputs Developed? 168
17.1 What Emission Inventory Data are Needed to Support Air Quality Models? 168
17.2 Can The National Emission Inventory Be Used As a Starting Point? 172
17.3 What Other Data are Needed to Support Emissions Modeling? 172
17.4 How Are Inventory Data Converted Into Air Quality Model Input? 174
17.5 Are there Other Emissions Modeling Issues? 177
17.6 How Are Emissions Estimated for Future Years? 179
17.6.1 Identify and prioritize key sectors and regions for projections 180
17.6.2 Collect available data and models for each sector needing projections 180
17.6.3 Evaluate and refine data for projections in modeling region 187
17.6.4 Create future-year inventories and air quality model inputs 188
17.7 Emissions Preparation for a Local Area Analysis 188

18.0 What are the Procedures for Evaluating Model Performance and What is the Role of Diagnostic Analyses? 190
18.1 How Do I Assess Model Performance And Make Use of Operational and Diagnostic Evaluations? 190
18.2 What Pollutants Are of Interest for Model Performance? 191
18.3 What Are the $PM_{2.5}$ Measurement versus Model Concentration Issues? . 192
18.4 How Should the Operational Evaluation of Performance Be Completed? 197
18.4.1 Ozone Operational Evaluation 198
18.4.2 PM/RH Operational Evaluation 202
18.5 How Do I Assess Model Performance and What Types of Analyses Can be Done to Evaluate the Accuracy of the Model Response: Diagnostic Evaluations? 208
18.5.1 Observational models 209
18.5.2 Probing Tools 210
18.5.3 Receptor Models (Multi-variate Source Apportionment) 213
18.5.4 Retrospective Analyses 214
18.5.5 Alternative Base Cases (Sensitivity tests) 215
18.6 How Should the Results of the Model Evaluation be Assessed (Interpreted)? 216

REFERENCES 218

Glossary 239

APPENDIX A 240

Appendix B: Summary of recent model performance evaluations conducted by other modeling groups 243

FOREWORD

The purpose of this document is to provide guidance to EPA Regional, State, and Tribal air quality management authorities and the general public, on how to prepare 8-hour ozone and $PM_{2.5}$ attainment demonstrations and regional haze uniform rate of progress analyses using air quality models and other relevant technical analyses. This guidance is designed to implement national policy on these issues. This document does not substitute for any Clean Air Act provision or EPA regulation, nor is it a regulation itself. Thus, it does not impose binding, enforceable requirements on any party, nor does it assure that EPA will approve all instances of its application. The guidance may not apply to a particular situation, depending upon the circumstances. The EPA and State decision makers retain the discretion to adopt approaches on a case-by-case basis that differ from this guidance where appropriate. Any decisions by EPA regarding a particular State Implementation Plan (SIP) demonstration will only be made based on the statute and regulations, and will only be made following notice and opportunity for public review and comment. Therefore, interested parties will be able to raise questions and objections about the contents of this guidance and the appropriateness of its application for any particular situation.

This guidance is a living document and may be revised periodically. Updates, revisions, and additional documentation will be provided at *http://www.epa.gov/ttn/scram/*. Any mention of trade names or commercial products in this document is not intended to constitute endorsement or recommendation for use. Users are cautioned not to regard statements recommending the use of certain procedures or defaults as either precluding other procedures or information, or providing guarantees that using these procedures or defaults will result in actions that are fully approvable. As noted above, EPA cannot assure that actions based upon this guidance will be fully approvable in all instances, and all final actions will only be taken following notice and opportunity for public comment. The EPA welcomes public comments on this document and will consider those comments in any future revisions of this guidance document, providing such approaches comply with all applicable statutory and regulatory requirements.

1.0 Introduction

This document describes how to estimate if an emissions control strategy will lead to attainment of annual and 24-hour national ambient air quality standards (NAAQS) for particles smaller than 2.5 µm in diameter ($PM_{2.5}$) and the 8-hour NAAQS for ozone. We also describe how to use modeled and monitored data to estimate the visibility improvement in *Class I areas* (e.g., national parks, wilderness areas) as part of a uniform rate of progress analysis[1].

The document describes how to apply air quality models to generate the predictions later used to evaluate attainment and/or uniform rate of progress assessments. Modeling to show attainment of the NAAQS primarily applies to nonattainment areas[2] for which modeling is needed, or desired. Modeling to assess uniform rate of progress for regional haze applies to all States[3]

The guidance consists of two major parts. Part I addresses the question, "how should I use the results of models and other analyses to help demonstrate attainment and/or assess uniform rate of progress?" We begin by describing a *modeled attainment test* for the 8-hour ozone NAAQS, the annual $PM_{2.5}$ NAAQS, and the 24-hour $PM_{2.5}$ NAAQS. We also recommend a modeled test to assess *uniform rate of progress* goals to reduce regional haze. We explain what is meant by a *modeled attainment demonstration*, a *modeled attainment test*, and a *weight of evidence determination*. We also identify additional data which, if available, can enhance the credibility of model results. Part I concludes by identifying what documentation States/Tribes should include as part of an attainment demonstration.

Part II of the guidance describes how to apply air quality models. The recommended procedure for applying a model has nine steps. The results of this process are then used to apply the modeled attainment test to support an attainment demonstration, as described in Part I of the guidance.

---

[1]Modeling is not used to determine reasonable progress goals and does not determine whether reasonable progress has been met. Modeling is one part of the reasonable progress analysis. The modeling results are used to determine if future year visibility at Class I areas are estimated to be on a (glide)path towards reaching natural background. This is called a uniform rate of progress analysis or "glidepath" analysis. The uniform rate of progress analysis is described in more detail in section 6.

[2]While this guidance document is primarily directed at modeling applications in nonattainment areas, it may also be useful as a guide for modeling in maintenance areas or to support other rules or sections of the Clean Air Act.

[3]The Regional Haze rule assumes that all 50 States either contain a Class I area or impact visibility within a Class I area. Therefore, a regional haze State Implementation Plan (SIP) is required for all States (40 CFR 50.308(b)).

1. Develop a conceptual description of the problem to be addressed.
2. Develop a modeling/analysis protocol.
3. Select an appropriate model to support the demonstration.
4. Select appropriate meteorological time periods to model.
5. Choose an appropriate area to model with appropriate horizontal/vertical resolution and establish the initial and boundary conditions that are suitable for the application.
6. Generate meteorological inputs to the air quality model.
7. Generate emissions inputs to the air quality model.
8. Run the air quality model with basecase emissions and evaluate the performance. Perform diagnostic tests to improve the model, as necessary.
9. Perform future year modeling (including additional control strategies, if necessary) and apply the attainment test.

Model applications require a substantial effort. States/Tribes 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.

1.1 What Is The Purpose Of This Document?

This document has two purposes. The first is to explain how to interpret whether results of modeling and other analyses support a conclusion that attainment of the ozone and/or $PM_{2.5}$ NAAQS, and/or uniform rate of progress for regional haze will occur by the appropriate date for an area. The second purpose is to describe how to apply an air quality model to produce results needed to support an attainment demonstration or a uniform rate of progress analysis.

The guidance herein should be viewed as recommendations rather than requirements. Although this guidance attempts to address issues that may arise in attainment demonstrations, situations which we have failed to anticipate may occur. These should be resolved on a case by case basis in concert with the appropriate U.S. EPA Regional Office.

1.2 Does The Guidance In This Document Apply To Me?

This guidance applies to all locations required to submit a State Implementation Plan (SIP), or Tribal Implementation Plan (TIP) revision with an attainment demonstration designed to achieve attainment of the ozone or $PM_{2.5}$ NAAQS. The guidance also applies to SIPs developed to address regional haze rule requirements. Areas required to submit an attainment demonstration and/or a reasonable progress demonstration are encouraged to follow the procedures described in this document. Details on when a State is required to submit a modeled attainment or reasonable progress demonstration can be found in the regional haze rule[4], the 8-

[4] 40 CFR 50.308

hour ozone implementation rule[5] and the $PM_{2.5}$ implementation rule[6].

Implementation plan revisions for ozone and $PM_{2.5}$ are due three years from the effective designation date after an area is designated "nonattainment" (e.g., June 15, 2007 for areas whose effective designation dates are June 15, 2004). Regional haze SIPs (for the first planning period) are due no later than December 17, 2007 (40 CFR 50.308(b)) Attainment and uniform rate of progress analyses supporting these revisions should be completed in time to allow sufficient time to complete the rulemaking process by the SIP due date.

1.3 How Does The Perceived Nature Of Ozone, PM, and Regional Haze Affect My Attainment Demonstration?

Guidance for performing attainment demonstrations needs to be consistent with the perceived nature of the pollutant. In this section, we identify several premises regarding ozone, $PM_{2.5}$, and regional haze. We then describe how the guidance accommodates each.

1.3.1 Ozone, PM, and Regional Haze

Premise 1. There is uncertainty accompanying model predictions. "Uncertainty" is the notion that model estimates will not perfectly predict observed air quality at any given location, neither at the present time nor in the future. Uncertainty arises for a variety of reasons, for example, limitations in the model's formulation which may be due to an incomplete representation in the model of physiochemical processes and/or meteorological and other input data base limitations, and uncertainty in forecasting future levels of emissions. States/Tribes should recognize these limitations when preparing their modeled attainment demonstrations.

We recommend several qualitative means for recognizing model limitations and resulting uncertainties when preparing an attainment demonstration. First, we recommend using models in a relative sense in concert with observed air quality data (i.e., taking the ratio of future to present predicted air quality and multiplying it times an "ambient" design value)[7]. As described later, we believe this approach should reduce some of the uncertainty attendant with using absolute model predictions alone. Second, we recommend that a modeling analysis be preceded by analyses of available air quality, meteorological, and emissions data to gain a qualitative understanding of an area's nonattainment problem. Such a description should be used to help guide a model application and may provide a reality check on the model's predictions. Third, we recommend that States/Tribes use several model outputs, as well as other supporting analyses, to provide corroborative evidence concerning the adequacy of a proposed strategy for meeting the

---

[5] 40 CFR 50.908

[6] 40 CFR 50.1007

[7] Ambient design values are based on observations made at monitor locations.

NAAQS. Modeling results and other supporting analyses can be weighed to determine whether or not the resulting evidence suggests a proposed control strategy is adequate to meet the NAAQS. Finally, we identify several activities/analyses which States/Tribes could undertake, if they so choose, to apply models and corroborative approaches in subsequent reviews and analyses of a control strategy, such as mid-course reviews. These subsequent reviews are useful for determining whether a SIP is achieving progress as expected.

Premise 2. For many areas, nested regional/urban model applications will be needed to support the attainment demonstration. Available air quality data suggest ozone and $PM_{2.5}$ concentrations approach levels specified in the NAAQS throughout much of the eastern U.S. and in large parts of California (U.S. EPA 2004a, U.S. EPA 2004b). A number of analyses (U.S.EPA, 1998b and U.S. EPA, 2005a and 2005b) show that regional ozone and PM transport can impact areas several hundred miles or more downwind. The regional extent of ozone and PM transport patterns and distances in some areas will likely necessitate nested regional model applications.

This guidance identifies several *modeling systems*[8] with nesting capabilities to resolve meteorological parameters, emissions, chemistry, and transport. We believe it is not beneficial to identify any modeling system as the preferred, or "guideline model" for ozone, $PM_{2.5}$[9] or regional haze modeling. States/Tribes may use any appropriate modeling system provided that the requirements of 40 CFR 51.112 are met. In this guidance, we provide certain criteria to assist States/Tribes in justifying the use of such modeling systems. These criteria apply equally to U.S.EPA models and alternative air quality model(s). The guidance also provides recommendations for developing meteorological, air quality and emissions inputs used in nested regional model applications, and makes suggestions for quality assuring inputs and evaluating performance of emissions, meteorological and air quality models.

Premise 3. Resource intensive approaches may often be needed to support an adequate attainment demonstration. This follows from the regional nature of ozone and $PM_{2.5}$ concentrations in the Eastern U.S. While we believe that existing and future regional reductions in NOx and SO_2 emissions will reduce ozone and $PM_{2.5}$ over much of the eastern U.S., elevated regional ozone and $PM_{2.5}$ concentrations will continue to affect local strategies needed to attain the NAAQS in the remaining nonattainment areas.

---

[8]A modeling system includes a chemical model, an emissions model and a meteorological model. Terms, such as this one, which are introduced using italics are defined more fully in a glossary at the back of this guidance. "Modeling system" and "air quality model" are used interchangeably. "Air quality model" means "modeling system" in this guidance.

[9]In this context, we are not referring to dispersion modeling of primary $PM_{2.5}$ sources subject to New Source Review (NSR) or Prevention of Significant Deterioration (PSD) analyses. We are only referring to $PM_{2.5}$ analyses for SIP attainment demonstrations.

This guidance recommends using regional modeling domains. Regional modeling applications require coordination, quality assurance and management of data bases covering large areas of the country. Resources used to run recommended models for generating meteorological and emissions inputs and the air quality model itself can be substantial. States/Tribes facing the need to develop an attainment demonstration requiring resource intensive techniques may wish to consider pooling resources in some manner. Examples might include delegating responsibilities for certain parts of the analyses to a single State/Tribe which can "specialize" in that kind of analysis. Another example might be formation of a regional technical center to perform analyses as directed by its client group of States/Tribes (e.g., multi-state and tribal organizations such as the Regional Planning Organizations (RPO) [10], LADCO, and the Ozone Transport Commission (OTC)).

Premise 4. High concentrations of ozone, $PM_{2.5}$ and regional haze often have a common origin. Ozone formation and formation of secondary particulates result from several common reactions and reactants. Secondary particulates are a major part of $PM_{2.5.}$ Often similar sources contribute precursors to both ozone and $PM_{2.5}$. In some regions of the U.S., high regional ozone and secondary particulates are observed under common types of meteorological conditions. Reducing $PM_{2.5}$ is the principal controllable means for improving regional haze. Reducing $PM_{2.5}$ precursors can also lead to reductions in ozone. Models intended to address secondary particulate matter problems need also to be capable of simulating ozone formation and transport. U.S. EPA policy is to encourage "integration" of programs to reduce ozone, $PM_{2.5}$ and regional haze to ensure they do not work at cross purposes and to foster maximum total air quality benefit for lower costs.

1.3.2 How is Measuring and Modeling Particulate Matter Different (and Often More Complicated) than Ozone?

Before modeling the expected $PM_{2.5}$ benefits from emissions controls, it is important to understand the unique and complicated aspects of measuring and modeling particulate matter. For many reasons, $PM_{2.5}$ and regional haze modeling presents many more difficulties compared to ozone modeling. In this section, we identify attributes of $PM_{2.5}$ that are applicable to most attainment or uniform rate of progress demonstrations. There are of course exceptions. As we discuss in Section 11.0, 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.

Premise PM1. 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

---

[10]EPA provides funding to five regional planning organizations to address regional haze and related issues. http://www.epa.gov/air/visibility/regional.html

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.

Premise PM2. In most parts of the country, "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 usually 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, 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 formation may take hours or days, as do some of the removal processes. Thus, many of the sources of measured secondary particulate matter may not be local emitted sources. 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 sources of precursors may not be necessary. Individual treatment of precursor emissions from relatively nearby large sources of primary $PM_{2.5}$ may be needed on a case by case basis.

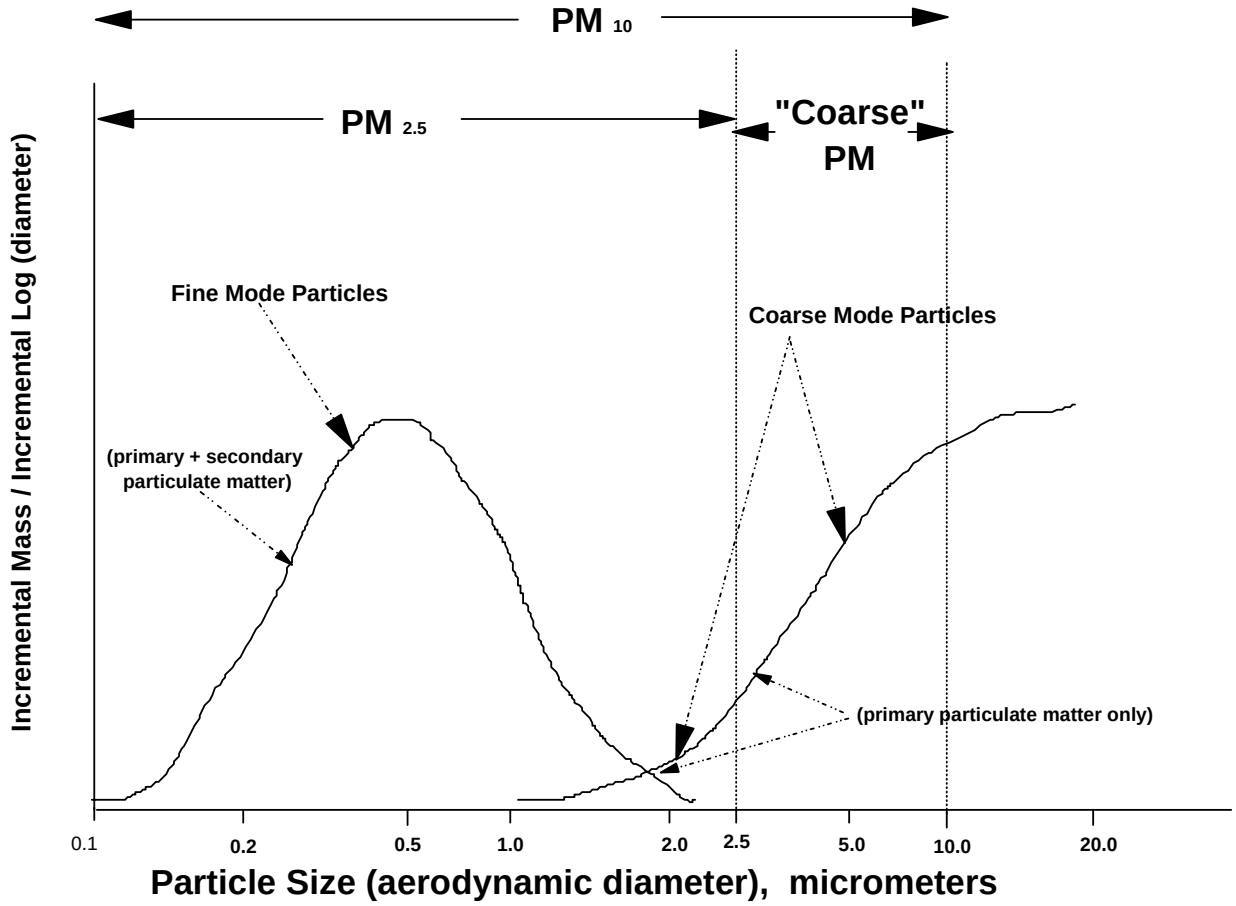


Figure 1.1. Conceptual Diagram Of Particulate Matter Properties

Premise PM3. 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 the 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 primary particles composed of soil material. Light extinction is also 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 can be enhanced 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 uniform rate of progress for regional haze will need to address secondary particulate matter. This, in turn, means that large modeling domains will be necessary. Class I areas are generally likely to be far removed from most sources of precursors for secondary particulate matter. Additionally, the measure of visibility (deciviews) with which we are most concerned, addresses maximum range of visibility. This measure reflects an effect which is integrated over a relatively large distance. Thus, the need for a large domain, the prevalence of 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 regional haze may be done without a fine degree of spatial resolution.

Premise PM4. Sampling anomalies associated with the Federal Reference Method for $PM_{2.5}$ need to be considered as part of the model attainment demonstration. 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 $PM_{2.5}$ (40 CFR Part 50, Appendix L). Since the FRM is included in the definition of the NAAQS, the modeled attainment test we recommend uses the measured concentrations of FRM $PM_{2.5}$ as the "ground truth". The FRM sampling protocol is known to be a precise measurement, but it may not accurately measure $PM_{2.5}$ in the air. There are many known positive and negative sampling artifacts associated with the measurement and analysis protocol. Among them are positive artifacts associated with organics and water retained on the Teflon filter and negative artifacts associated with volatile nitrates and organics (Frank, 2006).

The attainment test recommends using speciated ambient data to estimate $PM_{2.5}$ components. These measurements are from the Speciation and Trends Network (STN) for urban sites and the Interagency Monitoring of PROtected Visual Environments (IMPROVE) network for Class I areas. The Speciation measurements use different protocols than FRM measurements. Therefore, the species measured by STN and IMPROVE are often not directly comparable to the FRM $PM_{2.5}$. The attainment test recommends a set of default procedures to adjust the speciation measurements to make them more comparable to FRM measurements.

Premise PM5. 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, 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 may be 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 $PM_{2.5}$. 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 PM6. 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, wildfires, agricultural burning, prescribed burning 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 regional haze rule 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 a full year (or multiple years) or a representative portion of days may need to be modeled from each season.

Modeling for the 24-hour $PM_{2.5}$ NAAQS is further complicated by the fact that violations of the 24-hour standard may occur during different times of the year and may be caused by a variety of different $PM_{2.5}$ species and precursors.

Premise PM7. 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 PM8. 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) add to the uncertainty of the model's ability to accurately predict concentrations of $PM_{2.5}$ and its components at a given time and location.

The preceding paragraph has several implications for using models to demonstrate future attainment of a NAAQS for $PM_{2.5}$ or assessing uniform rate of progress 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.

1.4 What Topics Are Covered In This Guidance?

This guidance addresses two broad topics: Part I, "How do I use results of models and other analyses to help demonstrate attainment?", and Part II, "How should I apply air quality models to produce results needed to help demonstrate attainment?". Part I is divided into 8 sections (i.e., Sections 2-9). Part II consists of 9 sections (Sections 10-18).

Part I ("How do I use results of models and other analyses to help demonstrate attainment?") begins in Section 2 with an overview of the procedure for using modeling results to help demonstrate attainment of the 8-hour ozone NAAQS and the annual and 24-hour NAAQS for $PM_{2.5}$. Section 2 also summarizes the uniform rate of progress modeling analysis.

Section 3 describes the recommended modeled attainment tests.

Section 4 describes the attainment test for ozone in more detail.

Section 5 describes the attainment tests for each of the two NAAQS for $PM_{2.5}$.in more detail.

Section 6 outlines a recommended analysis to assess uniform rate of progress at Class I areas.

Section 7 describes how supporting analyses should be performed to complement the attainment test and the uniform rate of progress analysis, as well as how it should be used in a weight of evidence determination.

Section 8 identifies several data gathering activities and analyses which States/Tribes could undertake to enhance the credibility of the modeling and corroborative analyses to support subsequent reviews on progress toward attainment and/or visibility improvements.

Section 9 identifies the documentation necessary to adequately describe the analyses

used to demonstrate attainment of the ozone NAAQS.

Part II ("How should I apply air quality models to produce results needed to help demonstrate attainment?") begins in Section 10 with an overview of the topics to be covered.

Section 11 identifies a series of meteorological, emissions and air quality data analyses which should be undertaken to develop a qualitative description of an area's nonattainment problem prior to a model application. As we describe, this qualitative description should be used to guide the subsequent model application.

Section 12 describes the purpose, function, and contents of a modeling protocol.

Section 13 addresses what criteria should be considered in choosing a model to support the ozone and PM attainment demonstrations and uniform rate of progress analyses. Several guidelines are identified for accepting the use of a model for this purpose.

Section 14 provides guidance for selecting suitable time periods (or episodes) to model for an attainment demonstration or uniform rate of progress analysis. Topics include a discussion of the form of the NAAQS and its resulting implications for episode selection.

Section 15 identifies factors which should be considered in choosing; a model domain, the horizontal and vertical resolution, and the initial/boundary conditions for an air quality modeling application.

Section 16 addresses how to develop and evaluate meteorological inputs for use in a modeling exercise supporting an attainment demonstration or uniform rate of progress analysis..

Section 17 discusses how to develop appropriate emissions estimates for use in the selected air quality model.

Section 18 outlines the structure of model performance evaluations and discusses the use of diagnostic analyses.

The guidance concludes with references and a glossary of important terms which may be new to some readers.

This Page is intentionally left blank

Part I. How Do I Use Results Of Models And Other Analyses To Help Demonstrate Attainment?

2.0 What Is A Modeled Attainment Demonstration?--An Overview

A modeled attainment demonstration consists of (a) analyses which estimate whether selected emissions reductions will result in ambient concentrations that meet the NAAQS and (b) an identified set of control measures which will result in the required emissions reductions. As noted in Section 1, this guidance focuses on the first component of an attainment demonstration, that is, completion and interpretation of analyses to estimate the amount of emission reduction needed to attain the NAAQS. 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/Tribes should estimate the amount of emission reduction needed to demonstrate attainment of the NAAQS using the *modeled attainment test*. We recommend a similar analysis for assessing progress toward reducing regional haze. In addition, a State/Tribe should consider a broader set of model results, as well as perform a set of other corroboratory analyses to further support whether a proposed emission reduction will lead to attainment of the NAAQS or uniform rate of progress.

2.1 What Is The Recommended Modeled Attainment Test for Ozone and $PM_{2.5}$?-An Overview

A *modeled attainment test* is an exercise in which an air quality model is used to simulate current and future air quality. If future estimates of ozone and/or $PM_{2.5}$ concentrations are less than the NAAQS, then this element of the attainment test is satisfied[11]. Our recommended test is one in which model estimates are used in a "relative" rather than "absolute" sense. That is, we take the ratio of the model's future to current (baseline) predictions at monitors. We call these ratios, *relative response factors*. Future ozone and/or $PM_{2.5}$ concentrations are estimated at existing monitoring sites by multiplying a modeled relative response factor at locations "near" each monitor by the observation-based, monitor-specific, "baseline" design value. The resulting predicted "future concentrations" are compared to NAAQS.

The test for ozone is based on the calculation of a single mean ozone RRF for each monitor. The $PM_{2.5}$ attainment test is more complicated and reflects the fact that $PM_{2.5}$ is a mixture. In the test, ambient $PM_{2.5}$ is divided into major components. These are

- mass associated with sulfates
- mass associated with nitrates
-mass associated with ammonium
- mass associated with organic carbon

---

[11]As detailed in Section 7, additional corroborative analyses are still needed to supplement the modeled attainment test, even when predicted ozone and/ or $PM_{2.5}$ concentrations are less than the NAAQS.

- mass associated with elemental carbon
-mass associated with particle bound water
- mass associated with "other"primary inorganic particulate matter
-and passively collected mass

A separate RRF is calculated for each of the $PM_{2.5}$ components (except passive mass). We call each of these site-specific ratios, component-specific RRFs. Future $PM_{2.5}$ design values are estimated at existing monitoring sites by multiplying modeled relative response 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 future year values of the seven $PM_{2.5}$ components. If all future site-specific $PM_{2.5}$ design values are $\leq$ the concentration specified in the NAAQS, the test is passed.

2.2 What Is The Recommended Modeling Assessment For Regional Haze?- An Overview

The recommended modeled test for assessing uniform rate of progress in reducing regional haze is similar conceptually to the recommended tests for the two NAAQS for $PM_{2.5}$. Models are used to develop relative response 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 fine soil (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 (measured in deciviews). This ranking is used to identify the 20% of days with worst and 20% of 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 response factors to estimate future concentrations for each component on "best" and "worst" days. Future component concentrations are then inserted into 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 assess the future year visibility improvement.

2.3 What Does A Recommended Supplemental Analysis/Weight Of Evidence Determination Consist Of? --An Overview

As we describe in more detail in Section 7, States/Tribes should always perform complementary analyses of air quality, emissions and meteorological data, and consider modeling outputs other than the results of the attainment test. Such analyses are instrumental in guiding the conduct of an air quality modeling application. Sometimes, the results of corroboratory analyses may be used in a *weight of evidence determination* to show that attainment is likely despite modeled results which may be inconclusive. The further the attainment test is from being passed, the more compelling contrary evidence produced by corroboratory analyses must be to draw a conclusion differing from that implied by the modeled attainment test results. If a conclusion differs from the outcome of the modeled test, then the need for subsequent review (several years hence) with more complete data bases is increased. If the test is failed by a wide margin (e.g., future design values outside the recommended range at an individual site or multiple sites/locations), it is far less likely that the more qualitative arguments made in a weight of evidence determination can be sufficiently convincing to conclude that the NAAQS will be attained. Table 2.1 contains guidelines for assessing when corroboratory analyses and/or weight of evidence determinations may be appropriate.

Table 2.1 Guidelines For Weight of Evidence Determinations

Results of Modeled Attainment Test			Supplemental Analyses
Ozone	**Annual $PM_{2.5}$**	**24-Hour $PM_{2.5}$**	
Future Design Value < 82 ppb, all monitor sites	**Future Design Value <14.5 ug/m3, all monitor sites**	**Future Design Value <62 ug/m3, all monitor sites**	**Basic supplemental analyses should be completed to confirm the outcome of the modeled attainment test**
Future Design Value 82 - 87 ppb, at one or more sites/grid cells	**Future Design Value 14.5-15.5 ug/m3, at one or more sites/grid cells**	**Future Design Value 62-67 ug/m3, at one or more sites/grid cells**	**A weight of evidence demonstration should be conducted to determine if aggregate supplemental analyses support the modeled attainment test**
Future Design Value ≥ 88 ppb, at one or more sites/grid cells	**Future Design Value ≥ 15.5 ug/m3, at one or more sites/grid cells**	**Future Design Value ≥ 68 ug/m3, at one or more sites/grid cells**	**More qualitative results are less likely to support a conclusion differing from the outcome of the modeled attainment test[12].**

[12] Regional modeling completed by EPA indicates that, on average, considerable amounts of precursor control (e.g., 20-25 percent or more) will be needed to lower projected

In a weight of evidence (WOE) determination, States/Tribes should review results from several diverse types of air quality analyses, including results from the modeled attainment test. As a first step, States/Tribes should note whether or not the results from each of these analyses support a conclusion that the proposed strategy will meet the air quality goal. Secondly, States/Tribes should weigh each type of analysis according to its credibility, as well as its ability to address the question being posed (i.e., is the strategy adequate for meeting the NAAQS by a defined deadline?). The 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. If it is concluded that a strategy is inadequate to demonstrate attainment, a new strategy is selected for review, and the process is repeated. States/Tribes should provide a written rationale documenting how and why the conclusion is reached regarding the 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 will be met.

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

The procedure we recommend for estimating needed emission reductions differs from that in past guidance (U.S. EPA, 1996a) for ozone and PM in two major respects. First, we recommend a modeled attainment test in which model predictions are used in a relative rather than absolute sense. Second, the role of the weight of evidence determination, when used, has been expanded. That is, these results can now be used as a rationale for concluding that a control strategy will meet the NAAQS, even though the modeled attainment test alone may not be conclusive. There are several reasons why we believe these changes are appropriate.

1. Starting with an observed concentration as the base value reduces problems in interpreting model results. If a model under (or over) predicts an observed daily maximum concentration, the appropriate target prediction is not as clear as might be desired. For example, if an 8-hour daily maximum ozone concentration of 120 ppb were observed and a model predicted 100 ppb on that day, should the target for the day still be 84 ppb? In the relative attainment test, observed data is used to define the target concentration. This has the effect of anchoring the future concentrations to a "real" ambient value. Although good model performance remains a prerequisite for use of a model in an attainment demonstration, problems posed by less than ideal model performance on individual days are reduced by the new procedure.

2. The form of the 8-hour ozone and $PM_{2.5}$ NAAQS necessitates such an attainment test. The NAAQS for ozone and $PM_{2.5}$ requires ambient data to be averaged over three consecutive years. This presents difficulties using the resource-intensive Eulerian models we believe are

---

ozone design values by 3 ppb or $PM_{2.5}$ annual design values by 0.5 ug/m3 or 24-hour $PM_{2.5}$ design values by 3 ug/m3.

necessary to capture spatially differing, complex non-linearities between ambient ozone and precursor emissions. That is, it is difficult to tell whether or not a modeled exceedance obtained on one or more days selected from a limited sample of days is consistent with meeting the NAAQS. To do so would require modeling several years and, perhaps, many strategies. This problem is reduced by using the *monitored* design value, as an inherent part of the modeled attainment test.

3. $PM_{2.5}$ 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 possibility is reduced if the model responses are instead applied to components of $PM_{2.5}$ which are derived from measurements.

4. Model results and projections will continue to have associated uncertainty. The procedure we recommend recognizes this by including modeling plus other analyses to determine whether all available evidence supports a conclusion that a proposed emission reduction plan will suffice to meet the NAAQS. For applications in which the modeled attainment test is not passed, a weight of evidence analysis may be used to support a determination that attainment will be achieved, despite the results of the modeled attainment test. The weight of evidence determination includes several modeling results which are more difficult to relate to the form of the 8-hour ozone and 24-hour $PM_{2.5}$ NAAQS. These results address relative changes in the frequency and intensity of high modeled ozone or $PM_{2.5}$ concentrations on the sample of days selected for modeling. If corroboratory analyses produce strong evidence that a control strategy is unlikely to meet the NAAQS, then the strategy may be inadequate, even if the modeled attainment test is passed.

5. Focusing the modeled attainment test only at monitoring sites could result in control targets which are too low if the monitoring network is limited or poorly designed. We recommend a test which includes a review of the strategy's impact at locations without monitors. This exercise provides a supplemental test to determine whether there is a need for further action despite passing the modeled attainment test at all monitoring sites. While this test may indicate potential unmonitored violations, ultimately, the best way to account for a limited or poorly designed monitoring network is to use the model results, or other available analyses, to help determine locations where additional monitors should be sited.

3.0 What Is The Recommended Modeled Attainment Test?

In Section 2, we provided an overview of the recommended modeled attainment test. In this section we provide more details on the calculation of baseline design values, the definition of "nearby grid cells", and an "unmonitored area analysis" which provides estimates of future year values in unmonitored areas. Sections 4, 5, and 6 provide more specific details and examples for attainment tests for ozone and $PM_{2.5}$, and the uniform rate of progress assessment for regional haze respectively.

We begin by establishing the basic equation on which all of the tests are built upon. Equation (3.1) describes the recommended modeled attainment test, applied near monitoring site I.

$$\mathbf{(DVF)_I = (RRF)_I\ (DVB)_I} \qquad \mathbf{(3.1)}$$

where

$(DVB)_I$ = the baseline concentration monitored at site I, units in ppb or ug/m3[13];

$(RRF)_I$ = the relative response factor, calculated near site I, unitless
The relative response factor is the ratio of the future concentration predicted near a monitor (averaged over multiple days) to the baseline concentration predicted near the monitor (averaged over the same days), and

$(DVF)_I$ = the estimated future design value for the time attainment is required, ppb or ug/m3.

Equation (3.1) looks simple enough. However, several issues must be resolved before applying it.

(1) How is a "site-specific" baseline design value ($(DVB)_I$) calculated?

(2) In calculating the $(RRF)_I$, what do we mean by "near" site I?

(3) Several surface grid cells may be "near" the monitor, which one(s) of these should be used to calculate the $(RRF)_I$?

(4) How do you calculate future design values in unmonitored areas?

(5) Which base year emissions inventory should be projected to the future for the purpose of calculating RRFs?

---

[13]The units for ozone is ppb and the units for $PM_{2.5}$ is ug/m3.

(6) Which future year should emissions be projected to in order to assess attainment using the modeled attainment test?

3.1 Calculating site-specific baseline concentrations.

The modeled attainment test is linked to the form of the ozone and $PM_{2.5}$ NAAQS through the use of monitored design values. Official design values are calculated in different ways for 8-hour ozone, annual average $PM_{2.5}$, and 24-hour average $PM_{2.5}$. The following is a brief description of the calculations:

Ozone

The 8-hour ozone design value is calculated as the 3 year average of the fourth highest monitored daily 8-hour maximum value at each monitoring site. The standard is met if the design value is <0.8 ppm (in practice, this translates to < 85 ppb). The detailed description of the methodology can be found in 40 CFR Part 50.10, and Appendix I to Part 50.

Annual Average $PM_{2.5}$

The annual average $PM_{2.5}$ standard 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 $\mu g/m^3$. The annual arithmetic mean at a monitoring site is calculated by averaging the four quarterly arithmetic mean concentrations observed during a calendar year at the site. The annual mean concentration averaged over three years must be $\leq$ 15.0 $\mu g/m^3$ at all monitoring sites. The details can be found in 40CFR Part 50, Appendix N

In some cases, two nearby sites can be spatially averaged to determine the annual average design value for both sites. For details on spatial averaging requirements see 40 CFR Part 58, Appendix D, Section 2.8.

24-Hour Average $PM_{2.5}$

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 $\mu g/m^3$.[14] 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. See 40CFR Part 50, Appendix N for details. Attainment is reached if the design value at the site is $\leq$ 65 $\mu g/m^3$.

---

[14]See 40CFR Part 50, Appendix N for a discussion of how data completeness affects the determination of attainment status.

How is a "site-specific" baseline design value ($(DVB)_I$) calculated?

The baseline measured concentrations at each monitoring site is the anchor point for future year projected concentrations. The baseline design values are projected to the future using RRFs. In practice, the choice of the baseline design value can be critical to the determination of the estimated future year design values. Therefore, careful consideration should be given to the calculation of baseline values. The baseline design values should have the following attributes:

1) Should be consistent with the form of the applicable NAAQS.
2) Should be easy to calculate.
3) Should represent the baseline inventory year.
4) Should take into account the year-to-year variability of meteorology.
5) Should take into account the year-to-year variability of emissions.

Several possible methodologies to calculate baseline design values are:

1) The designation design value period (i.e. 2001-2003).
2) The design value period that straddles the baseline inventory year (e.g., the 2001-2003 design value period for a 2002 baseline inventory year).
3) The highest (of the three) design value periods which include the baseline inventory year (e.g. the 2000-2002, 2001-2003, and 2002-2004 design value periods for a 2002 baseline inventory year).
4) The average (of the three) design value periods which straddle the baseline inventory year.

For the modeled attainment tests we recommend using the average of the three design value periods which include the baseline inventory year (choice number 4 from above). Based on the attributes listed above, the average of the three design value periods best represents the baseline concentrations, while taking into account the variability of the meteorology and emissions (over a five year period).

The three design values that are averaged in the calculation cover a five year period, but the average design value is not a straight five year average. It is, in effect, a weighted average of the annual averages. For example, given a baseline inventory year of 2002, the years used to calculate the average design value range from 2000-2004. In the average of the 2000-2002, 2001-2003, and 2002-2004 periods, 2002 is "weighted" three times, 2001 and 2003 are weighted twice, and 2000 and 2004 are weighted once. This has the **desired** effect of weighting the projected ozone or $PM_{2.5}$ values towards the middle year of the five year period, which is the emissions year (2002 in this example). The average design value methodology is weighted towards the inventory year (which is the middle year) and also takes into account the emissions and meteorological variability that occurs over the full five year period (although the emissions and meteorology from the other years are weighted less than the middle year of the 5 year

period)[15]. Because of this, the average weighted design value is thought to be more representative of the baseline emissions and meteorology period than other methodologies such as choosing the highest single design value period.

Additionally, the average design value will be more stable (less year to year variability) than any single design value period. An analysis of recent ambient design value data at 471 ozone monitors, over the period 1993-2004, shows that the median standard deviation of design values was 3.3 ppb whereas the standard deviation of the 5 year weighted average design values was 2.4 ppb (Timin, 2005a). Also, moving from the period ending in 2003 to the period ending in 2004, the median change in the ozone design values was 4.0 ppb. The median change in the 5 year weighted average ozone design values was only 0.8 ppb (there was not a long enough data record to do the same analyses for $PM_{2.5}$) . These analyses show that the average design values are clearly more stable and will therefore provide a "best estimate" baseline year design value (DVB_I) for use in future year model projections.

The recommended averaging technique assumes that at least five complete years of ambient data is available at each monitor. In some cases there will less than five years of available data (especially at relatively new monitoring sites). In this case we recommend that data from the monitor is used if there is at least three consecutive years of data. If there are three years of data then the baseline design value will be based on a single design value. If there are four years of data then the baseline design value will be based on an average of two design value periods. If a site has less than three years of data, then the site should not ordinarily be used in the attainment test.

Calculating site-specific "baseline" design values[16] (DVB) to use in the attainment test

Example 3.1

Given:

(1) The baseline inventory year is 2002 (i.e., 2002 emissions are being modeled).

---

[15]The emissions trends and meteorological data during the five year period should be evaluated to determine if any large emissions changes and/or "extreme" meteorological conditions have occurred during the period. It is especially important to consider the emissions and meteorology during the middle year of the period. When choosing a base emissions year and calculating baseline design values, extreme conditions that may lead to either abnormally high or abnormally low concentrations should be discussed with the appropriate EPA Regional Office. See section 14 for more information on selecting modeling time periods.

[16]The "baseline design value" is an average of several design values and thus is technically not a design value. The guidance continues to refer to the average design values as "design values" even though they are based on <u>averages</u> of observed design values.

(2) For purposes of illustration, suppose the area contains only three ozone monitors.

Find: The appropriate site-specific baseline design values to use in the modeled attainment test.

Solution: Since the inventory reflects 2002, we need to examine monitored design values for overlapping 3-year periods that include 2002. The three design values are then averaged for each site. These are the values for site-specific baseline design values (DVB) in the modeled attainment test. The procedure is shown in Table 3.1.

Table 3.1 Example Illustrating Calculation Of Baseline Design Values

Monitor	2000-2002 Design Value, ppb	2001-2003 Design Value, ppb	2002-2004 Design Value, ppb	Baseline Design Value (DVB) Used In The Modeled Attainment Test, ppb
1	88	87	90	88.3[17]
2	86	84	91	87.0
3	88	86	85	86.3

3.2 Identifying surface grid cells near a monitoring site.

There are four reasons why we believe it is appropriate, in the modeled attainment test, to consider cells "near" a monitor rather than just the cell containing the monitor. First, one consequence of a control strategy may be "migration" of a predicted peak. If a State were to confine its attention only to the cell containing a monitor, it might underestimate the RRF (i.e., overestimate the effects of a control strategy). Second, monitor siting guidelines recommend spatial scales (for ozone and $PM_{2.5}$) which are generally urban or neighborhood scale. In many cases, the representative spatial scale of a monitor is larger than a single grid cell. Third, we believe that uncertainty in the formulation of the model and the model inputs is consistent with

[17] The average design value should carry one significant figure to the right of the decimal point for ozone and 24-hour $PM_{2.5}$ and 2 significant figures for annual $PM_{2.5}$. There are several calculations in the modeled attainment test which carry the tenths digit or hundreths digit. We have found that rounding and/or truncating concentrations and RRFs can lead to an overestimate or underestimate of the impact of emissions controls. In some cases, a few tenths of a ppb change (or a few tenths of a percent reduction) can be meaningful. Rounding or truncating can make the change appear to be equal to a full ppb (or 1%) or equal to zero change. For 8-hour ozone, it is recommended to round to the tenths digit until the last step in the calculation when the final future design value is truncated. For annual $PM_{2.5}$ it is recommended to round to the hundreths digit until the last step when the final future design value is rounded to the nearest tenths place. For daily $PM_{2.5}$ it is recommended to round to the tenths digit until the last step when the final future design value is rounded to the nearest whole number.

recognizing some leeway in the precision of the predicted location of daily average $PM_{2.5}$ and daily maximum ozone concentrations. Finally, standard practice in defining a gridded modeling domain is to start in the southwest corner of the domain, and determine grid cell location from there. Considering several cells "near" a monitor rather than the single cell containing the monitor diminishes the likelihood of inappropriate results which may occur from the geometry of the superimposed grid system.

Earlier ozone modeling guidance (U.S. EPA,1996a) has identified 15 km as being "near" a site. This is also consistent with the broad range of intended representativeness for urban scale ozone monitors identified in 40CFR Part 58, Appendix D. 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. For $PM_{2.5}$, sites having "neighborhood" or "urban" scales of representativeness are generally most suitable for estimating exposure or compliance with the NAAQS. Because of the rather long sampling 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 we originally defined for ozone modeling (U.S.EPA, 1999a).

For ease in computation, States/Tribes may assume that a monitor is at the center of the 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 a 15 km radius of) a monitor is a function of the size of the grid cells used in the modeling. Table 3.2 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 with 4 km grid cells, "nearby" is defined by a 7 x 7 array of cells, with the monitor located in the center cell.

The use of an array of grid cells near a monitor may have a large impact on the RRFs in "oxidant limited" areas (areas where NOx decreases may lead to ozone or secondary PM increases) and in areas with large gradients in primary $PM_{2.5}$. The array methodology could lead to unrealistically small or large RRFs, depending on the specific case. Care should be taken in identifying an appropriate array size for these areas. States/Tribes may consider the presence of topographic features, demonstrated mesoscale flow patterns (e.g., land/sea, land/lake interfaces), the density of the monitoring network, the density of emissions (especially primary PM)[18], and/or other factors to deviate from our default definitions for the array of "nearby" grid cells, provided the justification for doing so is documented.

---

[18]It may be inappropriate to use an array as large as 7 X 7 in areas with large gradients of primary $PM_{2.5}$ concentrations. Averaging of multiple grid cells may mask changes in local concentrations between the base year and future years.

Table 3.2. Default Recommendations For Nearby Grid Cells Used To Calculate RRF's

Size of Individual Cell, km	Size of the Array of Nearby Cells, unitless
4 - 5[19]	**7 x 7**
>5 - 8	**5 x 5**
>8 - 15	**3 x 3**
>15	**1 x 1**

3.3 Choosing model predictions to calculate a relative response factor $(RRF)_I$ near a monitor.

8-Hour Ozone NAAQS

Given that a model application produces a time series of estimated 1-hour ozone concentrations (which can be used to calculate running 8-hour averages), what values should be chosen from within the time series? We recommend choosing predicted 8-hour daily maximum concentrations from each modeled day (excluding "ramp-up" days) for consideration in the modeled attainment test. The 8-hour daily maxima should be used, because they are closest to the form of concentration specified in the NAAQS.

The second decision that needs to be made is, "which one(s) of the 8-hour daily maxima predicted in cells near a monitor should we use to calculate the RRF?" We recommend choosing the nearby grid cell with the highest predicted 8-hour daily maximum concentration with baseline emissions for each day considered in the test, and the grid cell with the highest predicted 8-hour daily maximum concentration with the future emissions for each day in the test. Note that, on any given day, the grid cell chosen with the future emissions need not be the same as the one chosen with baseline emissions.

We believe selecting the maximum (i.e., peak) 8-hour daily maxima on each day for subsequently calculating the relative response factor (RRF) is preferable for several reasons. First, it is likely to reflect any phenomenon which causes peak concentrations within a plume to migrate as a result of implementing controls. Second, it is likely to take better advantage of data produced by a finely resolved modeling analysis.

The relative response factor (RRF) used in the modeled attainment test is computed by taking the ratio of the **mean** of the 8-hour daily maximum predictions in the future to the **mean** of the 8-hour daily maximum predictions with baseline emissions, over all relevant days. The

[19]The appropriate size of the array for horizontal grid cells < 4km should be discussed with the appropriate U.S EPA Regional Office.

procedure is illustrated in Example 3.2.

Example 3.2

Given: (1) Four primary days have been simulated using baseline and future emissions.

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

(3) In each of the 9 grid cells "near" a monitor site I, the maximum daily predicted future concentrations are 87.2, 82.4, 77.5, and 81.1 ppb.

(4) In each of the 9 grid cells "near" a monitor site I, the maximum daily predicted baseline 8-hour daily maximum ozone concentrations are 98.3, 100.2, 91.6, and 90.7 ppb.

Find: The site-specific relative response factor for monitoring site I, $(RRF)_I$

Solution:

(1) For each day and for both baseline and future emissions, identify the 8-hour daily maximum concentration predicted near the monitor. Since the grid cells are 12 km, a 3 x 3 array of cells is considered "nearby" (see Table 3.2).

(2) Compute the mean 8-hour daily maximum concentration for (a) future and (b) baseline emissions.

Using the information from above,

(a) $(\text{Mean 8-hr daily max.})_{\text{future}} = (87.2 + 82.4 + 77.5 + 81.1)/4 = 82.1$ ppb
and

(b) $(\text{Mean 8-hr daily max.})_{\text{baseline}} = (98.3 + 100.2 + 91.6 + 90.7)/4 = 95.2$ ppb

(3) The relative response factor for site I is

$(RRF)_I = (\text{mean 8-hr daily max.})_{\text{future}}/(\text{mean 8-hr daily max.})_{\text{baseline}}$

$= 82.1/95.2 = 0.862$

Figure 3.1. Choosing Ozone Predictions To Estimate RRF's

(a) Predictions With Baseline Emissions

Day 1			Day 2			Day 3			Day 4		
97.2	95.5	96.2	100.2	98.5	98.1	87.8	90.1	89.9	85.9	87.9	88.9
97.1	95.2	89.1	100.0	99.1	97.3	90.9	91.6	88.7	87.9	90.5	90.7
97.2	98.3	97.6	99.5	95.4	97.9	88.5	89.4	90.2	86.9	87.3	88.4
98.3			100.2			91.6			90.7		

Mean Baseline Ozone Concentration = (98.3 + 100.2 + 91.6 + 90.7) / 4 = 95.2 ppb

---

(b) Predictions With Future Emissions

Day 1			Day 2			Day 3			Day 4		
86.1	85.4	86.8	82.2	80.8	81.2	72.1	76.1	75.5	75.4	78.8	79.8
86.2	84.5	84.3	82.4	79.9	80.5	74.6	77.5	74.3	80.8	79.5	80.9
85.8	87.2	86.9	81.4	77.8	80.1	76.9	77.4	75.6	80.4	76.9	81.1
87.2			82.4			77.5			81.1		

Mean Future Ozone Concentration = (87.2 + 82.4 + 77.5 + 81.1) / 4 = 82.1 ppb

$PM_{2.5}$ NAAQS

For $PM_{2.5}$ applications, we recommend choosing predicted 24-hour average concentrations from each modeled day (excluding "ramp-up" days) for consideration in the modeled attainment test[20]. The 24-hour average values should be used, because they are the basis for design value calculations.

The second decision that needs to be made is, "which of the 24-hour average values predicted in cells near a monitor should we use to calculate the RRF?" For the annual PM2.,5 NAAQS, we recommend taking the spatially averaged value of the nearby predictions (mean value of the grid cell array). Each component-specific relative response 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

[20]This includes modeled days for which there is no corresponding monitored data.

emissions. The procedure is the same as the ozone example presented above, except the spatial mean is used instead of the highest concentration near the monitor.

For the 24-hour $PM_{2.5}$ NAAQS and for annual $PM_{2.5}$ NAAQS sites which exhibit strong spatial concentration gradients of primary $PM_{2.5}$, we recommend calculating an RRF based on the single value in the grid cell which contains the monitor. Particularly in areas with strong primary PM concentration gradients, the use of a single grid cell is more likely to accurately capture the local change in $PM_{2.5}$ components.

3.4 Estimating design values at unmonitored locations: what is an unmonitored area analysis and why is it needed?

An additional review is necessary, particularly in nonattainment areas where the ozone or $PM_{2.5}$ monitoring network just meets or minimally exceeds the size of the network required to report data to Air Quality System (AQS). This review is intended to ensure that a control strategy leads to reductions in ozone or $PM_{2.5}$ at other locations which could have baseline (and future) design values exceeding the NAAQS were a monitor deployed there. The test is called an "unmonitored area analysis". The purpose of the analysis is to use a combination of model output and ambient data to identify areas that might exceed the NAAQS if monitors were located there.

The unmonitored area analysis should identify areas where future year design values are predicted to be greater than the NAAQS. The unmonitored area analysis for a particular nonattainment area is intended to address potential problems within or near that nonattainment area. The analysis should include, at a minimum, all nonattainment counties and counties surrounding the nonattainment area (located within the State). In large States, it is possible that unmonitored area violations may appear in counties far upwind or downwind of the local area of interest. In those cases, the distance to the nonattainment area and ability of the modeling to represent far downwind areas should be evaluated on a case by case basis. In order to examine unmonitored areas in all portions of the domain, it is recommended to use interpolated spatial fields of ambient data combined with gridded modeled outputs.

3.4.1 Why does the unmonitored area analysis need to use both ambient data and model output?

Ambient data can be interpolated to provide a set of spatial fields. The spatial fields will provide an indication of concentrations in monitored and unmonitored areas. But a simple interpolation of the ambient data cannot identify unmonitored areas with higher concentrations than those measured at monitors. The interpolated concentration between monitors will generally be the same or lower than the measured concentration at the monitors (assuming that more sophisticated statistical techniques are not used, such as adding a nugget effect or a trend surface). The interpolation technique does not account for emissions or chemistry information that may be needed to identify potential unmonitored violations.

The gridded model output (absolute) concentrations can also be used to examine unmonitored area concentrations. The model provides an hourly concentration for every grid cell. The concentrations can be analyzed to determine unmonitored areas where the model predicts high values. But the absolute predictions from the model may not be entirely accurate. The model output is only as accurate as the emissions and meteorological input. But unlike the interpolated ambient data, the model output explicitly accounts for emissions, chemistry, and meteorology over the entire domain.

Both the interpolated ambient data and the model outputs have major weaknesses. But they also both have strengths. We can take advantage of the strengths of each dataset by combining the two types of data. The interpolated spatial fields of ambient data provide a strong basis for estimating accurate pollutant concentrations at monitors and near monitors. Given that information, the model outputs can be used to adjust the interpolated spatial fields (either up or down) so that more accurate estimates can be derived in the unmonitored areas[21]. The best way to use the model to adjust the spatial fields is to use modeled *gradients.* It is preferable to assume that the model is predicting areas of generally high or low ozone or PM, as compared to assuming that the absolute predictions from the model are correct. For example, in areas where the model predicts relatively high ozone or PM concentrations, the spatial fields can be adjusted upward. In areas where the model predicts relatively low ozone or PM concentrations, the spatial fields can be adjusted downward. In this way, it may be possible to predict unmonitored areas that may have high concentrations. At the same time, concentrations in rural areas, (which may be overly influenced by high monitored concentrations near urban areas), may be adjusted downward. It may also be possible to predict high $PM_{2.5}$ concentrations in unmonitored areas which contain large sources of primary PM emissions. The combination of interpolated spatial fields and modeled output will be referred to as "gradient adjusted spatial fields"

3.4.2 Implementation of Gradient adjusted Spatial Fields

Gradient adjusted spatial fields are first created for the base year. Future year estimates can then be created by applying gridded RRFs to the gradient adjusted spatial fields. The basic steps are as follows:

1) Interpolate base year ambient data to create a set of spatial fields.
2) Adjust the spatial fields using gridded model output gradients (base year values).
3) Apply gridded model RRFs to the gradient adjusted spatial fields.
4) Determine if any unmonitored areas are predicted to exceed the NAAQS in the future.

EPA has developed a software package called "Modeled Attainment Test Software" (MATS) which will spatially interpolate data, adjust the spatial fields based on model output gradients and multiply the fields by model calculated RRFs (steps 1-3 above) (ABT, 2007).

---

[21]The accuracy of interpolated fields can be tested by removing sets of monitors to see how well the interpolation scheme estimates known concentrations at ambient monitoring sites.

States will be able to use the EPA-provided software or are free to develop alternative techniques that may be appropriate for their areas or situations.

Step 1

The first step in the analysis is to interpolate ambient data. Ideally, design values should be interpolated. The same 5 year weighted average design values that are used in the monitor based model attainment test can be used in the development of ambient spatial fields. Care should be taken so that the interpolated fields are not unduly influenced by monitoring sites that do not have complete data. Since the design values can vary significantly from year to year, it is important to use a consistent set of data. In some cases, it may be preferable to interpolate individual years of data or individual design values, and then average those up to get the 5 year weighted average.

We are not recommending a single interpolation technique. EPA has provided example analyses in the past using the Kriging interpolation technique (U.S.EPA, 2004c). Alternatively, EPA's BenMAP software, which was used to create interpolated fields for the Clean Air Interstate Rule (CAIR) uses the Voronoi Neighbor Averaging (VNA) technique (Abt, 2003)[22].

For ozone analyses, a single spatial field of ozone values is needed. But for the $PM_{2.5}$ NAAQS, we recommend creating interpolated spatial fields for $PM_{2.5}$ (FRM) and for each component of $PM_{2.5}$. For the annual $PM_{2.5}$ NAAQS, a set of quarterly average spatial fields can be created. The 4 quarters are averaged to get an annual average set of fields.

For the 24-hour $PM_{2.5}$ standard, a spatial field can be created using the high end of the distribution of 24-hour $PM_{2.5}$ concentrations. This is best represented by interpolating the measured high values from each quarter. To be consistent with the attainment test, we recommend interpolating the $PM_{2.5}$ concentration in each quarter which is equal to or less than the 98th percentile value for the year. For the $PM_{2.5}$ component species, we recommend interpolating the high $PM_{2.5}$ days in each quarter. This can be based on the highest monitored days in each quarter.

Interpolated data should be evaluated to quantify the errors associated with the interpolation technique. Additional information on evaluation of spatial fields can be found in (U.S. EPA, 2004c).

Step 2

The second step in the process involves the use of gridded model output to adjust the spatial fields. The BenMAP software contains an example of this technique called eVNA. It uses seasonal average model output data to adjust interpolated spatial fields (MATS also uses the

[22]MATS also uses a version of VNA.

same technique). The eVNA technique has been used in health benefits assessments (U.S. EPA, 2005a). A specific metric is needed to determine the model predicted gradient in concentrations for each of the NAAQS. For ozone, a logical metric is the 4th highest ozone prediction in each grid cell[23]. For the annual $PM_{2.5}$ NAAQS, the model predicted quarterly mean concentrations for $PM_{2.5}$ and $PM_{2.5}$ species can be used to adjust the ambient spatial fields. For the 24-hour $PM_{2.5}$ NAAQS, the gradient adjusted fields can be derived from the high end of the distribution of daily averages in each quarter. This could be for all days with 24-hour predicted (or measured) $PM_{2.5}$ > 65 ug/m3 (or some relatively high value) or simply the top 10% or 25% of all ($PM_{2.5}$) days in each quarter.

Step 3

The next step is to create future year fields by multiplying the base year gradient adjusted spatial fields by model derived gridded RRFs. The RRFs for the unmonitored area analysis are calculated in the same way as the monitored based attainment test (except that the grid cell array is not used in the spatial fields based analysis). The future year concentrations are equal to the base year concentration times the RRF in each grid cell. The future year gradient adjusted spatial fields are then analyzed to determine if any grid cells are predicted to remain above the NAAQS.

For ozone, a single spatial field is multiplied by a single set of model derived RRFs. For $PM_{2.5}$, the RRFs for each of the species, for each quarter, are multiplied by the spatial fields for each species, for each quarter.

3.4.3 Using the Results of the Unmonitored Area Analysis

It should be stressed that due to the lack of measured data, the examination of ozone and $PM_{2.5}$ concentrations as part of the unmonitored area analysis is more uncertain than the monitor based attainment test. As a result, the unmonitored area analysis should be treated as a separate test from the monitor based attainment test. The results from the unmonitored area analysis should, at a minimum, be included as a supplemental analysis. While it is expected that additional emissions controls are needed to eliminate predicted violations of the monitor based test, the same requirements may not be appropriate in unmonitored areas.

The cause of predicted violations of the unmonitored area analysis may not be obvious. While careful analysis of the emissions in the area may reveal likely sources of an ozone violation, elimination of the violation may not be simple or straightforward. Alternatively, predicted violations of the unmonitored area analysis for $PM_{2.5}$ may, in some cases be simple to diagnose. They may occur in areas with high emissions of primary $PM_{2.5}$. Elimination of the

---

[23]The metric should approximate the measured design values at monitoring sites. Depending on the days modeled, other metrics, such as the 2nd or 3rd highest predicted ozone values or an average of several days may be a better proxy for the design value.

predicted violation may be possible by placing controls on one or more sources. Additional dispersion modeling may be helpful in diagnosing the source(s) of high primary $PM_{2.5}$ in a particular grid cell. The results of a more refined dispersion modeling analysis may help identify emissions controls or locations where additional monitoring may be appropriate.

It is recommended that predicted violations of the unmonitored area analysis are carefully scrutinized to determine whether they are likely to exist in the ambient air or whether they may be caused by an error or uncertainty in the modeling system. At a minimum, it may be appropriate to commit to additional deployment of ambient monitors in areas where the unmonitored area analysis predicts future violations[24]. This monitoring would allow a better assessment in the future of whether the NAAQS is being met at currently unmonitored locations.

Violations of the unmonitored area analysis should be handled on a case by case basis. As such, additional analyses and/or tracking requirements may be needed depending on the nature of the problem and the uncertainty associated with the potential violation(s).

3.5 Which base year emissions inventory should be projected to the future for the purpose of calculating RRFs?

The modeled attainment test adjusts observed concentrations during a baseline period (e.g., 2000-2004) to a future period (e.g., 2009) using model-derived "relative response factors". It is important that emissions used in the attainment test correspond with the period reflected by the chosen baseline design value period (e.g., 2000-2004). Deviations from this constraint will diminish the credibility of the relative response factors. Therefore, it is important to choose an appropriate baseline emissions year. There are potentially two different base year emissions inventories. One is the base case inventory which represents the emissions for the meteorology that is being modeled. These are the emissions that are used for model performance evaluations. For example, if a State is modeling a base year of 1998 (or episodes from 1998), "base case" emissions and meteorology would be for 1998. As described in Section 18, it is essential to use base case emissions *together with* meteorology occurring in the modeled episode(s) in order to evaluate model performance.

Once the model has been shown to perform adequately, it is no longer necessary to model the base case emissions. It now becomes important to model emissions corresponding to the period with a recent observed design value. The second potential base year inventory corresponds to the middle year of the baseline average design value (e.g 2002 for a 2000-2004 average design value). This is called the baseline inventory. The baseline emissions inventory is the inventory that is ultimately projected to a future year.

---

[24]It would also be appropriate to commit to additional emissions controls in lieu of additional monitoring in unmonitored areas.

In section 17 we recommend using 2002 as the baseline inventory year for the current round of ozone, $PM_{2.5}$, and Regional Haze SIPs. If States/Tribes model the full year of 2002 (or episodes from 2002) then the base case and baseline inventory years will be the same[25]. But if States/Tribes model other years or episodes or full seasons from other years, then the base case inventories should be projected (or "backcasted") to 2002 to provide a common starting point for future year projections.

Alternatively, the baseline emissions year could be earlier or later than 2002, but it should be a relatively recent year (preferably within the 5 year design value window). In order to gain confidence in the model results, the emissions projection period should be as short as possible. For example, projecting emissions from 2002 to 2009 (with a 2000-2004 baseline average design value) should be less uncertain than projecting emissions from 1995 to 2009 (with a 1993-1997 baseline average design value). Use of an older baseline average design value period is discouraged.

It is desirable to model meteorological time periods occurring during the period reflected by the baseline design value (e.g., 2000-2004). However, modeling time periods need not be selected from the period corresponding to the baseline design value, provided they are representative of meteorological conditions which commonly occur when exceedances of the ozone or $PM_{2.5}$ standard occur (or represent typical annual concentrations of $PM_{2.5}$) . The idea is to use selected representative time periods to capture the sensitivity of predicted ozone or $PM_{2.5}$ to changes in emissions during commonly occurring conditions. There are at least three reasons why using time periods outside the period with the baseline design value may be acceptable: (1) availability of air quality and meteorological data from an intensive field study, (2) the desire to use meteorological data which may be "more representative" of typical conditions compared to the baseline design value period and (3) availability of a past modeling analysis in which the model performed well.

Under the regional haze rule, the period for establishing baseline visibility is defined as 2000-2004. Therefore, emissions and meteorological time periods should be chosen from the base time period. Modeling the same year (or time period) for $PM_{2.5}$ NAAQS applications and uniform rate of progress analyses is a logical way to conserve resources.

3.6 Choosing a year to project future emissions.

States/Tribes should project future emissions to the attainment year or time period, based on the area's classification. The "Final Rule to Implement the 8-Hour Ozone National Ambient Air Quality Standard, Phase 1" provides a schedule for implementing emission reductions

---

[25] The year may be the same, but the emissions may still differ. The base case inventory may include day specific information (e.g. wildfires, CEM data) that is not appropriate for using in future year projections. Therefore the baseline inventory may need to replace the day specific emissions with average or "typical" emissions (for certain types of sources).

needed to ensure attainment by the area's attainment date (40 CFR 51.908). Specifically, it states that emission reductions needed for attainment must be implemented by the beginning of the ozone season immediately preceding the area's attainment date. The $PM_{2.5}$ implementation rule contains similar provisions. It states that emissions reductions should be in place by the beginning of the year preceding the attainment date (40 CFR 51.1007).

The logic for requiring emissions reductions by the year (or season) immediately preceding the attainment year follows from language in the Clean Air Act. The Clean Air Act (172(a)(2)(C)) identifies a process by which attainment is expected to occur. Shortly after the attainment deadline, EPA is to evaluate whether areas have attained. If the area is clean for the last three years, then the area has met the attainment deadline. But if the area was above the standard for earlier years/seasons but has clean data for the single year/season **immediately preceding** the attainment date (and if the area has met planning requirements), the area is eligible for a 1-year extension of the attainment deadline. If the air is clean again the following year, the area is eligible for a second 1-year extension of the attainment deadline. The expectation is then that the third year's data, in combination with data from years 1 and 2, will show attainment (This is commonly referred to as the "clean data policy".) Therefore, emissions in the year **preceding** the attainment year should be at a level that is consistent with attainment. It also follows that the year preceding the attainment year should be modeled for attainment planning purposes.

As an example, based on the language in the Clean Air Act and the ozone and $PM_{2.5}$ rules, areas with an attainment date of no later than June 15th 2010 (for moderate ozone nonattainment areas) or April 5th, 2010 (for $PM_{2.5}$ nonattainment areas)[26], need to have emission reductions implemented no later than the beginning of 2009 (for $PM_{2.5}$) or the beginning of the 2009 ozone season (for 8-hour ozone). Therefore, modeling the year before the attainment deadline (2009 in this case) is generally appropriate for both ozone and $PM_{2.5}$ attainment demonstrations. However, attainment dates are expressed as "no later than" three, five, six, or nine years after designation and nonattainment areas are required to attain <u>as expeditiously as practicable</u>. Therefore, additional considerations are necessary before a modeling year can be established.

Using moderate ozone nonattainment areas as an example, these areas have an attainment date of no later than June 15, 2010, or as expeditiously as practicable. States/Tribes are required to conduct a Reasonably Available Control Measures (RACM) analysis to determine if they can

---

[26]One key difference between ozone and $PM_{2.5}$ planning is that all $PM_{2.5}$ nonattainment areas have a single classification and a single set of attainment deadlines. The $PM_{2.5}$ nonattainment areas are required to attain the standard within 5 years of designation or as expeditiously as practicable. They can also request up to a 5 year extension of the attainment deadline. The procedures for justifying an extension of the attainment date are contained in the PM implementation rule (40 CFR 50.1004). In this guidance document, we will assume that areas will model a future year based on attaining within the first 5 year period. Areas that request and are granted attainment date extensions should consult with their Regional Office regarding future year modeling analyses and requirements

advance their attainment date by at least a year[27]. A RACM analysis is also required for $PM_{2.5}$ nonattainment areas[28]. Since areas are required to attain as expeditiously as practicable and perform a RACM analysis, results of the analysis may indicate attainment can be achieved earlier, (e.g., 2008). In this case, the timing of implementation of control measures should be used to determine the appropriate projection year. For example, if emission reductions (sufficient to show attainment) are implemented no later than the beginning of the 2008 ozone season, then the attainment year and the future projection year should be no later than 2008. In all cases, the selection of the future year(s) to model should be discussed with the appropriate EPA Regional Office as part of the modeling protocol development process.

For regional haze assessments, the first review of progress will occur in 2018. This will likely be based on monitored data from the 2013-2017 period. Therefore, a logical future period to model is the middle of the five year review period (2015). But the uniform rate of progress level can be calculated for any year in the future. Therefore, it is appropriate to model any future year between 2015 and 2018. Unless there are reasons to the contrary, any of the years between 2015 and 2018 can be considered to be representative of the regional haze future planning period. The calculation of the uniform rate of progress should be consistent with the emissions projection period. For example, if a future year of 2015 is modeled, the uniform rate of progress calculation should be based on the number of years between the base period and 2015. Similarly, if a future year of 2018 is modeled, then the uniform rate of progress should be based on the number of years between the base period and 2018.

.

[27] 40 CFR 51.912(d)

[28] 40 CFR 51.1010(b)

4.0 Ozone Attainment Test

Section 3 contained the details of the modeled attainment test that are common to both ozone and $PM_{2.5}$. This section describes several additional details related to the 8-hour ozone attainment test. The section concludes with an example calculation.

4.1 Limiting modeled 8-hour daily maxima chosen to calculate RRF.

On any given modeled day, meteorological conditions may not be similar to those leading to high concentrations (i.e., values near the site-specific design value) at a particular monitor. If ozone predicted near a monitor on a particular day is much less than the design value, the model predictions for that day could be unresponsive to controls (e.g., the location could be upwind from most of the emissions in the nonattainment area on that day). Using equation (3.1) could then lead to an erroneously high projection of the future design value.

In order to examine this issue, we analyzed modeled baseline and future emissions for 30 episode days during 1995 using a grid with 12 km x 12 km cells and 9 vertical layers[29]. We examined modeled RRF's computed near each of 299 monitoring sites in the eastern half of the United States (Timin, 2005b). The study examined the day to day variability of (daily) RRFs at each site. One purpose of the study was to assess the extent to which a relative response factor (RRF) is dependent on the magnitude of modeled current 8-hour daily maxima.

Figure 4.1 shows an example of the raw data from the analysis for all of the monitoring sites in the Baltimore region. The plot shows the daily RRFs vs. the base case daily maximum modeled concentrations for all days (above 60 ppb in this case). In this example, it can be seen that the model tends to respond more to emissions reductions (lower RRFs) at higher predicted ozone concentrations. There appears to be a general pattern in the model results such that the model predicts less benefit from emissions reductions at lower concentrations. The greater model response at higher concentrations is likely due to more "controllable ozone" at higher concentrations. On days with high ozone concentrations, there is a relatively high percentage of locally and/or regionally generated ozone compared to days with low basecase concentrations. Days with low concentrations likely have a high percentage of ozone due to background and boundary conditions (background ozone is uncontrollable, although boundary conditions may be controllable, depending on the size of the domain.) Since we are generally interested in the model response on high ozone days, these results tend to suggest that the RRF calculation should be limited to days when the model predicts high ozone concentrations.

---

[29]See http://www.epa.gov/cair/pdfs/finaltech02.pdf for documentation of the base case and future year modeling.

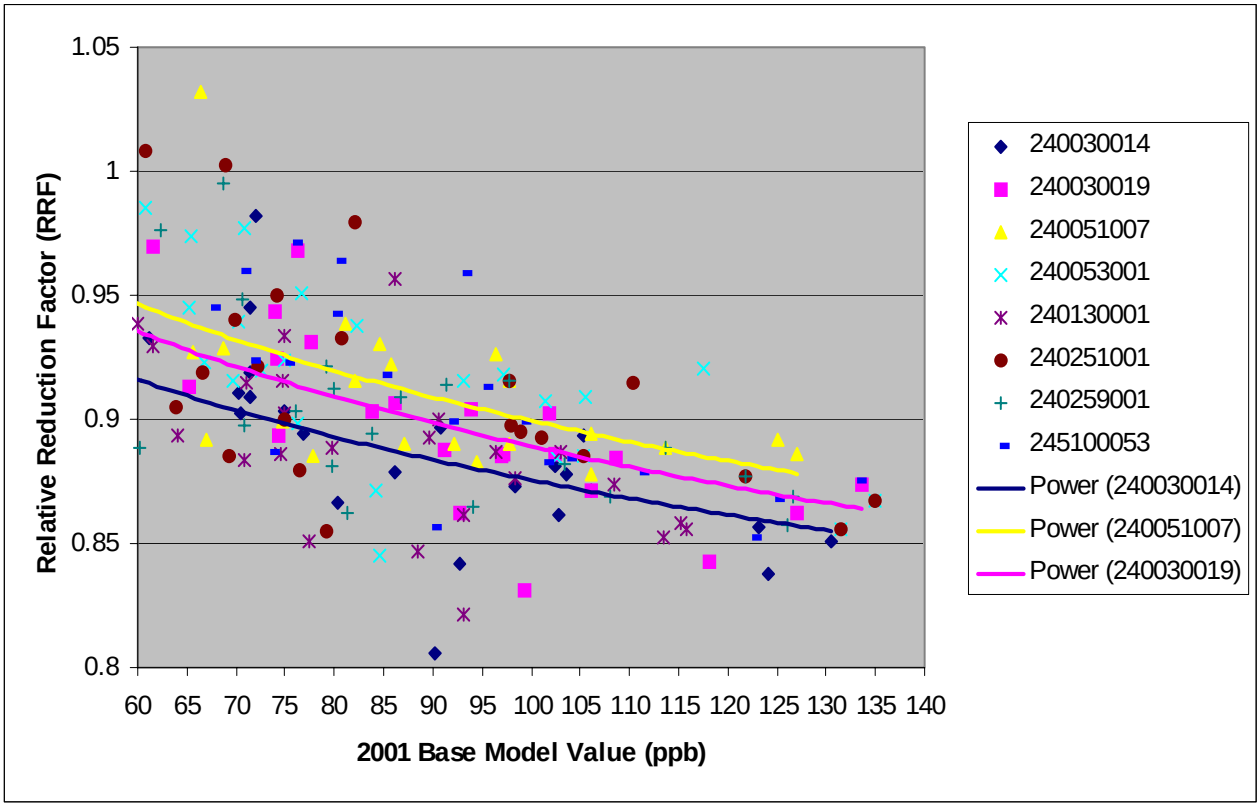


Figure 4.1 - Daily relative response factors as a function of daily maximum base modeled concentrations for monitors in the Baltimore nonattainment area.

The analysis examined daily RRFs, but in practice, the RRFs are not calculated on a daily basis[30]. A mean RRF is calculated based on the mean base case concentration (across model days) divided by the mean future case concentration (across the same model days). As such, we also calculated mean RRFs using various minimum concentration thresholds. The minimum concentration thresholds examined ranged from 70-85 ppb in 5 ppb increments[31]. The

---

[30]Although it is recommended to calculate mean RRFs, examination of daily RRFs can provide important information. Information gleaned from daily RRFs can be used to examine model response under different meteorological regimes. Additionally, days that are particularly unresponsive to emissions controls can indicate problems in the model inputs and/or formulation and may also provide a focus for the development of control strategies. For example, controlling specific upwind sources may provide benefits on particular days that appear to be otherwise unresponsive.

[31]It was clear from the plots that a minimum threshold value of less than 70 ppb would not be appropriate. An upper threshold of 85 ppb was examined because it is equal to the NAAQS. We are generally concerned about the model response on days that exceed the

monitoring sites were screened to eliminate sites that had a limited number of "high" modeled ozone days. All sites that had less than 10 days with an 8-hour average maximum modeled concentration of 85 ppb were dropped from the analysis. This left 206 sites in the analysis[32].

Table 4.1 shows that the mean RRF (averaged across all sites) is sensitive to the minimum ozone threshold. As the threshold is raised from 70 ppb to 85 ppb, the RRFs become smaller (larger percent decrease in ozone). On average, the model predicts a 0.2% greater ozone reduction for every 5 ppb increase in the minimum threshold. Additionally, the variability of the daily RRFs (as measured by the standard deviation of the daily RRFs) is reduced as the threshold is increased. This is an important finding because lower variability in day to day RRFs indicates lower uncertainty in the mean RRFs, and hence the attainment test results.

Minimum Threshold	Mean RRF	Mean Standard Deviation
70 ppb	0.879	0.030
75 ppb	0.876	0.029
80 ppb	0.874	0.028
85 ppb	0.872	0.026

Table 4.1- Mean RRFs and standard deviations as a function of various minimum thresholds.

As a result of the sensitivity in model response at concentrations below the NAAQS, combined with the increased variability of the RRFs at lower concentrations, we recommend using a minimum concentration threshold of 85 ppb[33]. This will result in RRFs that adequately represent high ozone days and also provide for more robust RRFs and future design values. For a more detailed description of selecting modeling days for ozone, see Section 14.1.1.

NAAQS. But at the same time, it is necessary to have a sufficient number of days in the mean RRF calculation to develop a robust estimate (this is addressed in more detail in section 14.1.1).

[32]The same set of sites was used for each of the threshold concentrations (70-85 ppb).

[33]The analysis suggests a threshold of greater than 85 ppb may be appropriate, but 85 ppb was chosen as the upper end of the threshold because it is equal to the NAAQS. For areas with very high baseline design values (>110 ppb) it may be necessary to use a higher threshold.

Example 4.1

Example 4.1 illustrates how to apply the minimum concentration threshold.

Given: The same simulations as performed in Example 3.2 yield low predictions near site I with baseline emissions on day 3, such that the 8-hour daily maximum ozone concentration predicted for that day is 65.0 ppb (rather than the 91.6 ppb shown in Example 3.2).

Find: The relative response factor near site I ($(RRF)_I$).

Solution: (1) Calculate the mean 8-hour daily maximum ozone concentration obtained near site I for baseline and future emissions. Exclude results for day 3 from the calculations. From Example 3.2,

(a) $(\text{mean 8-hr daily max})_{future} = (87.2 + 82.4 + 81.1)/3 = 83.6$ ppb

(b) $(\text{mean 8-hr daily max})_{baseline} = (98.3 + 100.2 + 90.7)/3 = 96.4$ ppb.

(2) Compute the relative response factor by taking the ratio of future/baseline.

$(RRF)_I = 83.6/96.4 = 0.867$

4.2 How Do I Apply The Recommended 8-Hour Ozone Modeled Attainment Test?

States/Tribes should apply the modeled attainment test at all monitors within the nonattainment area plus other nearby counties within the State[34]. Inputs described in Section 3.1 are applied in Equation (3.1) to estimate a future design value at all monitor sites and grid cells for which the modeled attainment test is applicable. When determining compliance with the 8-hour ozone NAAQS, the standard is met if, over three consecutive years, the average 4th highest 8-hour daily maximum ozone concentration observed at each monitor is $\leq$ 0.08 ppm (i.e., $\leq$ 84 ppb using rounding conventions)[35]. Thus, if all resulting predicted future design values (DVF) are $\leq$ 84 ppb, the test is passed. The modeled attainment test is applied using 3 steps.

---

[34]States are responsible for submitting SIPs for all areas of their State. As such, the monitored and unmonitored area attainment tests should be applied both within and outside of the nonattainment area. In the modeling protocol, States should identify the appropriate areas to apply the respective tests. States should work with their EPA Regional Office to help determine the appropriate areas.

[35]40CFR Part 50.10, Appendix I, paragraph 2.3

Step 1. Compute baseline design values. Compute site-specific baseline design values (DVBs) from *observed* data by using the average of the design value periods which include the baseline inventory year.

This is illustrated in Table 3.1 for specific sites. The values in the right hand column of Table 3.1 are site-specific baseline design values.

Step 2. Estimate relative response factors. Use air quality modeling results to estimate a relative response factor for each grid cell near a monitoring site.

This step begins by computing the mean 8-hour daily maximum ozone concentrations for future and baseline emissions. The relative response factor for site I is given by Equation 4.1.

$$(RRF)_I = (\text{mean 8-hr daily max})_{future} / (\text{mean 8-hr daily max})_{baseline} \quad (4.1)$$

Using Equation (4.1), the relative response factor is calculated as shown in example 3.2. Note that the RRF is calculated to three significant figures to the right of the decimal place. The last significant figure is obtained by *rounding*, with values of "5" or more rounded upward. For the illustration shown in Table 4.2, we have assumed that the same four days described previously in Example 4.1 have been simulated. Note that on day 3, model baseline 8-hour daily maximum ozone concentration was < 85 ppb. As discussed in Section 4.1, predictions for this day are not included in calculating the mean values shown in the last row of the table. We have also assumed that the monitored baseline design value (DVB) at site I is 102.0 ppb.

Step 3. Calculate future design values for all monitoring sites in the nonattainment area. Multiply the observed baseline design values obtained in Step 1 times the relative response factors obtained in Step 2.

In Table 4.2, we see (column (2)) that the baseline observed design value at monitor site I is 102.0 ppb. Using Equation (3.1), the predicted future design value for monitor site I is,

$$(DVF)_I = (102.0 \text{ ppb})\ (0.867) = 88.434 \text{ ppb} = 88 \text{ ppb}$$

Note that the final future design value is *truncated*[36] and in this example, the modeled attainment test is not passed at monitor site I.

[36]This effectively defines attainment in the modeled test as <= 84.999 ppb and nonattainment as >= 85.0 ppb.

Table 4.2 Example Calculation of a Site-Specific Future Design Value $(DVF)_I$

Day	Calculated baseline design value, $(DVB)_I$, (ppb)	Baseline 8-hr daily max. concentration at monitor (ppb)	Future predicted 8-hr daily max. concentration at monitor (ppb)	Relative response factor(RRF),	Future design value, $(DVF)_I$, (ppb)
1		98.3	87.2	-	-
2		100.2	82.4	-	-
3		65.0	Not Considered	-	-
4		90.7	81.1	-	-
Mean	102.0	96.4	83.6	0.867 (i.e., 83.6/96.4)	88.434= 88 ppb

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

In this section, we begin by illustrating the monitored tests for the annual and 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 primary $PM_{2.5}$ analysis which may be used to better capture the contributions of primary $PM_{2.5}$ to nonattainment and the benefits of controlling primary $PM_{2.5}$ in areas with relatively large contributions from primary sources.

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

As described in section 2, the monitored attainment test for $PM_{2.5}$ utilizes both $PM_{2.5}$ and individual $PM_{2.5}$ component species. We therefore call the attainment test for $PM_{2.5}$, the Speciated Modeled Attainment Test (SMAT). A separate RRF is calculated for each $PM_{2.5}$ species. In order to perform the recommended modeled attainment test, States should divide observed mass concentrations of $PM_{2.5}$ into 7 components (plus passive mass):

- mass associated with sulfates
- mass associated with nitrates
-mass associated with ammonium
- mass associated with organic carbon
- mass associated with elemental carbon
-mass associated with particle bound water
- mass associated with “other”primary inorganic particulate matter
-and passively collected mass

To apply the test, States must first have run an air quality model at least twice--to simulate current emissions and to simulate future year emissions. We recommend a modeled attainment test which has 4 basic steps. Additional details are provided later in this section.

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

Derive current quarterly mean concentrations[37] for each of the major components of $PM_{2.5}$. This is done by multiplying the monitored quarterly mean concentration of Federal Reference Method

[37]Concentrations should be calculated based on 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 data used to calculate design values is averaged on a calendar quarter basis before calculating annual averages.

(FRM) (EPA, 1997a) derived $PM_{2.5}$ by the monitored fractional composition of $PM_{2.5}$ species for each quarter. (e.g., 20% sulfate x 15.0 ug/m^3 $PM_{2.5}$ = 3.0 ug/m^3 sulfate).

Step 2. Using model results, derive component-specific relative response factors (RRF) at each monitor for each quarter.

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

The relative response 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 baseline design value for annual $PM_{2.5.}$

$C_{j,\ projected}$ is the future year quarterly mean concentration predicted at or near the monitoring site.

(e.g., given model predicted base year sulfate of 10.0 ug/m^3 and future year concentration of 8.0 ug/m^{3,} then RRF for sulfate is 0.800).

Step 3. Apply the component specific RRFs to observed air quality to obtain projected quarterly species estimates.

For each quarter, multiply the current quarterly mean component concentration (step 1) times the component-specific RRF obtained in step 2. This leads to an estimated future quarterly mean concentration for each component. (e.g., 3.0 ug/m^3 sulfate x 0.800 = future sulfate of 2.4 ug/m^3).

Step 4. Calculate a future year annual average $PM_{2.5}$ estimate.

Sum the quarterly mean components to get quarterly mean $PM_{2.5}$ values. Then average the quarterly mean $PM_{2.5}$ concentrations to get a future year annual average $PM_{2.5}$ estimate for each FRM site.

Compare the projected average annual arithmetic mean $PM_{2.5}$ concentration obtained in Step 4 with 15.0 μg/m^3. If all values are $\leq$ 15.0 μg/m^3, the test is passed.

5.1.1 What Ambient $PM_{2.5}$ Data is Used in the Attainment Test?

$PM_{2.5}$ data collected at FRM sites is used for nonattainment designations. Therefore we

recommend using FRM data as the base data for projecting future $PM_{2.5}$ concentrations. As can be seen from the list of steps, the modeled attainment test is critically dependent on the availability of species component mass at FRM sites. This raises several issues. First, the FRM filter measurements and $PM_{2.5}$ speciation measurements do not always measure the same mass (Frank, 2006). There are numerous known issues with positive and negative sampling artifacts. And second, the majority of FRM sites do not have co-located STN[38] speciation samplers[39]. Both of these issues are addressed below.

5.1.2 FRM Monitors that Don't Have Speciation Data

There are approximately 1200 FRM sites and only ~200 urban speciation monitoring sites. This makes it difficult to apply the attainment test at the majority of FRM sites. Species concentration data and/or species fractions are needed in order to apply SMAT. There are several possible ways to estimate species concentrations at FRM monitors that lack speciation data. Among them are:

1) Use concurrent data from a nearby speciation monitor to estimate species concentrations and/or fractions at one or more FRM sites.

2) Use representative speciation data (from a different time period) collected in an area to estimates species data at FRM sites.

3) Use interpolation techniques to create spatial fields using ambient speciation data.

4) Use interpolation techniques to create spatial fields using ambient speciation data and gridded model outputs to gradient adjust the species concentrations.

In general, we recommend using spatial interpolation techniques to estimate species concentrations at FRM sites that do not have speciation data (numbers 3 and 4 above). But in some cases, interpolating data from nearby sites may not be appropriate, may not be feasible, or simply may not be necessary. For the CAIR analysis, a relatively simple interpolation technique was used to estimate species concentrations at all FRM sites in the country. The analysis used a Voronoi Neighbor Averaging (VNA) technique. Other interpolations have been done using Kriging and other more complicated methodologies. States are encouraged to explore techniques that are most appropriate for their area and situation. EPA's MATS software contains routines to interpolate data and model output using the VNA technique.

---

[38]References to STN monitors in this document always refer to the overall speciation network which includes both trends sites and SLAMS sites.

[39]There are ~1100 FRM measurement sites and ~200-250 urban speciation sites (trends and SLAMS).

For areas which contain one or more speciation sites, and where FRM sites exhibit very little spatial gradients, it may be appropriate to assume that the speciation site is representative of the entire area. Areas which exhibit strong spatial gradients in $PM_{2.5}$ concentrations should strongly consider using more sophisticated techniques to estimate species concentrations. Combining ambient data with model output concentrations (number 4 above) may help adjust concentrations in areas with strong gradients, but limited speciation data. This technique has great promise but has not been used extensively for $PM_{2.5}$ analyses. The technique is further limited by uncertainties in the representativeness of the model outputs and emissions inventories.

5.1.3 New Species Calculations and Definitions

Recent data analyses (Frank, 2006) have noted that the FRM monitors do not measure the same components and do not retain all of the $PM_{2.5}$ that is measured by routine speciation samplers and therefore cannot be directly compared to speciation measurements from the Speciation Trends Network (STN)[40]. By design, the FRM mass measurement does not retain all ammonium nitrate and other semi-volatile materials (negative sampling artifacts) and includes particle bound water associated with sulfates, nitrates and other hygroscopic species (positive sampling artifacts). This results in concentrations (and percent contributions to $PM_{2.5}$ mass) which may be different than the ambient levels of some $PM_{2.5}$ chemical constituents.

The FRM data is used to determine compliance with the NAAQS. Therefore, it is critical to estimate the species composition as measured on the FRM filters. In addition, for the purposes of predicting changes in $PM_{2.5}$ components, constructed $PM_{2.5}$ mass should match the composition of mass retained by the FRM. As such, we are recommending changes to the calculation and definition of $PM_{2.5}$ species compared to the (original 2001) draft $PM_{2.5}$ modeling guidance[41].

We recommend a SMAT technique which uses an FRM mass construction methodology which results in reduced nitrates (relative to the amount measured by routine speciation networks), higher mass associated with sulfates and nitrates (reflecting water included in gravimetric FRM measurements) and a measure of organic carbonaceous mass which is derived from the difference between measured $PM_{2.5}$ and its non-organic carbon components (Frank,

---

[40]The information in this section applies to the most common samplers in the STN network. Some networks use alternative special purpose samplers to collect both $PM_{2.5}$ and $PM_{2.5}$ speciation data. The characteristics of the sampler and the analytical procedures used to produce chemical speciation data should be considered in determining which, if any adjustments are appropriate to make the data useful for comparison to FRM data.

[41]The draft $PM_{2.5}$ modeling guidance recommended a mass reconstruction scheme that was identical to the IMPROVE technique. We are still recommending use of IMPROVE assumptions for regional haze modeling (see section 6).

2006). This characterization of $PM_{2.5}$ mass also reflects "other" primary inorganic $PM_{2.5}$ and other minor constituents. Frank (2006) terms this approach "sulfate, adjusted nitrate, derived water, inferred carbonaceous material balance approach (SANDWICH)". The resulting characterization provides a complete mass balance. It does not have any unknown mass which is sometimes presented as the difference between measured $PM_{2.5}$ mass and the characterized chemical components derived from routine speciation measurements. The recommended SMAT characterizations should yield more accurate assessments of future $PM_{2.5}$ concentrations as measured by FRM monitors.

5.1.4 Recommended Treatment of Species Data

As noted above, special treatment of species component mass is recommended in the SMAT methodology. The goal is to reconstruct the measured species components so that they add up to the measured FRM mass. This concept can generally be represented by the following equation:

$RCFM_{FRM}$ = [Ammoniated Sulfate Mass] + [Retained Nitrate Mass] + [Retained Carbonaceous Mass] + [Other Primary $PM_{2.5}$] + [Other Components] [5.1]

In the above characterization, RCFM equals reconstructed fine mass and all of the listed chemical components reflect those retained during sampling and equilibration on the FRM's Teflon filter. Sulfate and nitrate mass include associated ammonium which may be different than assumed ammonium sulfate and ammonium nitrate compounds. Also, sulfates, nitrates and carbonaceous mass includes particle bound water associated with these hygroscopic aerosols. In this characterization, other primary $PM_{2.5}$ mass is intended to be a more general term that includes fine soil, and oxides that result from other PM emissions. The following section describes the recommended treatment for each of the components.

Retained Nitrate Mass

The first step in the procedure for identifying FRM mass components is to estimate the retained nitrate mass on the FRM filters. The FRM does not capture all of the semi-volatile components of the ambient air, such as ammonium nitrate. The retained amount of nitrate ion, however, can be accurately estimated by a simple thermodynamic model that involves 24-hr ambient nitrate speciation concentrations (as measured by a standard speciation sampler using a nylon filter preceded by a HNO3 denuder) together with hourly ambient temperature and humidity. Atmospheric nitrates are higher during the cooler months. Retention on the FRM is also higher during the cooler months and essentially all the nitrates are lost during the summer. The retention does not appear to depend on ambient NH3 or HNO3. More NO3 is retained at low temperatures and high humidity which varies by sampling location and time of year.

Because nitrate retention varies by site and season, an ammonium nitrate equilibrium model can be used to predict the amount of nitrates retained on the FRM Teflon filter. As used by Hering (Hering, 1999; Zhang, 1992),

$$\text{delta NO3 (ug/m3)} = 745.7/T_R * 1/24 * \sum_{i=1}^{24} (\mathbf{K}_i^{1/2}) \qquad [5.2]$$

where delta NO3 is the amount of volatilized nitrate mass, T_R is the reference temperature for the sampled air volume in degrees Kelvin and $\mathbf{K_i}$ is the dissociation constant for ammonium nitrate evaluated at the ambient temperature for hour i. The nitrate loss can be predicted for each day, based on hourly temperature and relative humidity data. The SMAT analysis for CAIR used National Weather Service temperature and relative humidity data for the closest station. Other sources of meteorological data may also be appropriate. Further details on the nitrate loss calculations can be found in (Frank, 2006).

Ammonium Associated with Sulfates and Retained Nitrates

To determine the mass associated with nitrates, we first assume retained nitrate is probably all ammonium nitrate. Thus the ammonium associated with nitrates can be derived directly from the measured or predicted $NO3_{FRM}$ as

$$NH4_{NO3} = 0.29 * NO3_{FRM} \qquad [5.3]$$

The difference between total FRM NH4 (amount associated with nitrates and sulfates), termed $NH4_{FRM}$, and the measured NH4, termed $NH4_{STN,}$, is needed to determine the ammoniated form of sulfates as described by equation 5.1. Recent measurement study by Collett (Collett, 2004) shows that NH4 may not be completely retained during collection on nylon filters preceded by a nitric acid denuder. During sampling conditions associated with nitrate volatilization, ammonium nitrate dissociates but the HNO3 downstream of the denuder is recaptured on the basic nylon media and the result is reported as particle nitrate. On the other hand, the NH4+ volatilizes to gaseous NH3 and apparently passes thru the filter. At several FRM study sites, the STN NH4 which is adjusted for evaporated NH4NO3 tends to more closely correspond to the measured NH4 from the FRM Teflon filter. However, for other sites, the measured STN NH4 appears to agree with FRM NH4.

In the CAIR analysis, 50% of the ammonium associated with volatilized nitrate was assumed to also volatilize. But more recent information suggests that using measured ammonium (assuming none is volatilized) may be a better assumption. We recommend using measured ammonium, but further analysis of this issue is warranted.

There are several ways to estimate ammonium mass for use in SMAT. The most direct way is to use measured ammonium concentrations from the STN network (IMPROVE does not measure ammonium ion).

A second, more indirect method is to calculate the ammonium associated with sulfate and the degree of neutralization of sulfate (DON), and then use the resulting information to calculate ammonium mass. Due to uncertainties associated with the ammonium measurements and the lack of ammonium measurements in rural areas, this indirect method was used for the CAIR

analysis. The ambient data is such that all of the ammonium data is from urban sites (STN), but the sulfate and nitrate data is from both urban (STN) and rural (IMPROVE) sites. This leads to an overestimation of ammonium concentration in rural areas when ammonium is directly interpolated. Therefore, in the CAIR analysis, calculated DON, SO4 and $NO3_{FRM}$ were interpolated to get quarterly average concentrations at each FRM site. The interpolated species concentrations were then used to calculate $NH4_{FRM}$ using the following equation:

$$NH4_{FRM} = DON * SO4 + 0.29*NO3_{FRM} \quad [5.4]$$

The indirect calculation of ammonium mass from interpolated fields tends to smooth out the gradients in mass. This was deemed to be beneficial, due to the uncertainty in the measurements.

Particle Bound Water

Because ammoniated sulfate and ammonium nitrate are hygroscopic, the retained sulfate and nitrate mass will include water. Particle bound water (PBW) can be estimated using an equilibrium model. The CAIR analysis used the Aerosol Inorganic Model (AIM) *(Clegg, 1998)* to calculate PBW. PBW was derived from quarterly average FRM concentrations of sulfate, ammonium, and nitrate as describe above. Estimated hydronium ion, H+, needed to achieve ionic balance was derived from the latter values. The model enables the distribution of water and ions to be calculated between liquid, solid and vapor phases for specific temperature and relative humidity conditions. Typical FRM filter equilibration conditions of 35% RH and 22 deg C (295 deg K) temperature were used.

Application of AIM at the specified FRM filter equilibration conditions show that PBW is much more dependent on sulfate concentration compared to nitrate and that the relationship varies somewhat by season. There is proportionally less estimated PBW water for wintertime aerosol which has higher NO3 and lower SO4. The PBW concentrations are also sensitive to the degree of neutralization of the sulfate particles (determined by the relative concentration of $NH4_{FRM}$).

For computational convenience, a polynomial regression equation was fit to the calculated water mass from AIM and the three input values that fed into AIM (sulfate, nitrate and ammonium). A polynomial equation can then be used in all SMAT analyses to estimate water.

We recommend calculating PBW as a component of $PM_{2.5}$ mass. The AIM model (or other equilibrium models) can be used, or a regression equation can be developed to simplify the process.

Other Primary $PM_{2.5}$

The terms “crustal”, “fine soil”, “major metal oxides”, “inorganic particulates”, and

"other PM" are sometimes used interchangeably. For $PM_{2.5}$ NAAQS calculations we will refer to this material as "other primary $PM_{2.5}$"(OPP). For regional haze calculations, we will continue to refer to this material as fine soil.

For the purposes of estimating other primary $PM_{2.5}$ for SMAT, all measured non-carbon mass that is not organic in nature (not associated with sulfate and/or nitrate) should be counted. As with the other $PM_{2.5}$ components measured on the FRM filter, there is uncertainty associated with this estimate. The "crustal" or "fine soil" definition from IMPROVE can be used to estimate other primary $PM_{2.5}$ [42] or an alternative formula can be defined which better estimates the urban nature of the FRM measurements. The IMPROVE definition of "fine soil" accounts for the typical crustal components (and attached mass) that would be expected in remote Class I areas. Fine soil is represented as five elements (Al, Si, Fe, Ti, and Ca) with coefficients to represent various oxides and material which may be attached to or associated with the major elements. In urban areas, inorganic PM which is not elemental carbon, or associated with sulfate or nitrate, may come from many sources such as re-suspended dust or industrial sources (stack or fugitives). It is generally "crustal" in nature (dominated by silicon), but urban PM is more likely to contain heavy metals and industrial components.

Although the composition of inorganic PM may differ between urban and rural (remote) areas, we recommend using an equation similar to the IMPROVE fine soil equation to estimate other primary $PM_{2.5}$ for the $PM_{2.5}$ attainment test. The recommended equation is suggested by (Frank, 2006) and uses only four elements. The equation removes aluminum (and accounts for associated mass by increasing the coefficient for Si), due to the fact that aluminum is often missing from the speciation measurements. This allows for more complete data.

The recommended equation is as follows:

Other primary $PM_{2.5}$ mass = $3.73 \times [Si] + 1.63 \times [Ca] + 2.42 \times [Fe] + 1.94 \times [Ti]$

<u>Blank mass</u>

The other quantifiable components of $PM_{2.5}$ mass include passively collected mass, represented by a field blank concentration of 0.3-0.5ug/m3 (U.S. EPA, 2002a). This appears to constitute a contamination of the filter resulting from handling or contact with the FRM cassette. This value is deemed to be an important constituent of $PM_{2.5}$ mass (it is assumed to not be dependent on pollutant emissions). We recommend using a default nominal blank mass value of 0.5 ug/m3. This value can be modified based on local FRM blank mass measurements. The blank mass is assumed to remain constant through time (RRF=1.0).

---

[42]IMPROVE estimates fine soil as: $2.2 \times [Al] + 2.49 \times [Si] + 1.63 \times [Ca] + 2.42 \times [Fe] + 1.94 \times [Ti]$

Calculation of Carbonaceous Mass

Organic carbon mass is typically estimated from blank corrected speciation data, where organic carbonaceous mass is first estimated by multiplying the organic carbon concentrations by 1.4 or alternative factors to account for the oxygen, hydrogen and other elements associated with ambient carbon particles.

There are many uncertainties in estimating carbonaceous mass from carbon measurements (Turpin, 2001; Chow, 2004). Uncertainties include differences in carbon measurement protocol between urban and rural monitoring locations; a relatively "bumpy" surface of urban carbon concentrations as derived from urban and rural organic carbon measurements; and lack of carbon measurements at all FRM locations. We therefore recommend an alternative approach to estimate the organic carbon contribution to $PM_{2.5}$ mass.

The recommended SMAT approach estimates organic carbon by mass balance. Precisely measured FRM $PM_{2.5}$ mass (U.S. EPA, 2003a) is compared to the sum of its non-organic carbon components. The latter are sulfates, ammonium, nitrates, estimated particle bound water, elemental carbon, estimated other primary $PM_{2.5}$ material plus 0.5 ug/m3 blank mass as discussed earlier.

This approach estimates retained organic carbon FRM mass and explicitly accounts for the following important and difficult to estimate carbon mass properties: (1) regional and urban-rural differences in the mix of carbonaceous aerosols, i.e. the amount of oxygen, hydrogen, etc; (2) retained water associated with hygroscopic carbon compounds (Saxena, 1996; Yua, 2004); (3) volatile carbonaceous material measured by speciation samplers, but not retained in FRM mass; and (4) uncertainties associated with blank corrections of measured organic carbon.

Organic Carbon Mass by mass balance **(OCM_{mb})** is defined as ,

$$\mathbf{OCM_{mb}} = PM_{2.5} - \{ [SO4] + [NO3_{FRM}] + [NH4_{FRM}] + [water] + [EC] + [OPP] + [0.5] \} \quad [5.5]$$

In this expression, all of the above components represent the mass retained on FRM Teflon filters.

This approach completely accounts for FRM mass[43] and OCMmb is often greater than the

[43]The OC by mass balance technique assumes that all other mass is accounted for and therefore all remaining mass is OCM. This may not always be a good assumption. The results of the technique should be carefully evaluated to ensure that OC mass is not overestimated (and therefore other mass components are underestimated). This may be a problem in areas that do not have nearby speciation data and have relatively large concentrations of primary $PM_{2.5}$. The OC by mass balance technique may inadvertently apportion mass to organic carbon which may actually be EC or "other" primary $PM_{2.5}$ mass. All available ambient data and modeling data

amount that would be derived directly from speciation measurements. Because of uncertainties in speciation measurements and their estimates from interpolated surfaces, setting a lower limit (floor) may be necessary so that the OCMmb is not unreasonably low. In the CAIR analysis, the floor was set so that OCMmb could not be more than 30% lower than measured OCM (OC*1.4). We used interpolated measured values of OCM to calculate the floor. The lower limit was equal to interpolated OC * 1.4 * 0.7. If the OCMmb concentration was less than the lower limit, it was set equal to the lower limit. The recommended approach is to simply set the OC floor equal to measured OC.

There may also be situations where an OCM "ceiling" is needed. In remote urban areas with relatively high FRM concentrations that may be surrounded by rural background concentrations, the OC by mass balance technique may apportion 95% or more of the $PM_{2.5}$ mass to OCM. If this is not a reasonable assumption, then a ceiling may be needed to cap the OCM as a percentage of $PM_{2.5}$ mass. Based on measured data (FRM sites with co-located speciation data), it appears that on a quarterly average basis, OCM is rarely more than 80% of total $PM_{2.5}$ mass. This may be a reasonable default ceiling, but a lower value (or in rare circumstances a higher value) may be more appropriate in many regions of the country.

Summary of $PM_{2.5}$ Composition Calculations

The terms of equation 5.6 reflect the final estimated composition of the particles measured by the FRM (for each quarter). Quarterly average FRM mass is equal to the sum of the seven species plus blank mass.

$$PM_{2.5FRM} = \{ [OCMmb] + [EC] + [SO4] + [NO3_{FRM}] + [NH4_{FRM}] + [water] + [OPP] + [0.5] \} \quad [5.6]$$

The recommended order to generate the data is as follows:

1) Calculate adjusted nitrate using hourly meteorology and 24-hour average nitrate measurements.

2) Calculate quarterly averages for adjusted nitrate, sulfate, elemental carbon, ammonium (or degree of sulfate neutralization (DON)), OPP mass , and measured OCM.[44]

3) Quarterly average ammonium is calculated from the adjusted nitrate, sulfate, and DON values (if measured ammonium is not used directly).

4) Calculated ammonium, sulfate, and nitrate values are input into the polynomial water equation to derive particle bound water concentrations.

---

should be used to evaluate the species apportionment results.

[44]The measured OCM is only used to calculate the "floor" for OCMmb

5) Carbon mass by difference (OMCmb) is calculated from the $PM_{2.5}$ mass, adjusted nitrate, ammonium, sulfate, water, elemental carbon, other primary $PM_{2.5}$, and blank mass values. The sum of the 7 species plus blank mass is equal to the FRM mass.

We illustrate application of the recommended test in example 5.1.

Example 5.1

Given: (1) Area "C" has 2 monitoring sites.

(2) Monitored FRM air quality data show the following average quarterly mean $PM_{2.5}$ concentrations based on a 5 year weighted average of observations from 2000-2004 at each site (values are in $\mu g/m^3$).

Table 5.1

Site	Quarter 1	Quarter 2	Quarter 3	Quarter 4
1	17.21 ug/m3	16.34	20.30	14.76
2	15.39	17.98	18.23	13.76

(3) The area has 1 urban speciation site which is co-located with FRM site 2. The speciation data is only available for 2003 and 2004. Therefore, the species data for the two years has been interpolated to derive estimated species concentrations at each site. The species concentrations (derived from 2003 and 2004 data) are matched up with FRM data for 2003 and 2004 to derive representative species fractions.

Average monitored air quality for the 3rd quarter of 2003 and 2004 at site 1 is as follows[45]:

Table 5.2

FRM Mass (ug/m3)	Blank Mass (ug/m3)	Non-blank Mass (ug/m3)	Sulfate (ug/m3)	Nitrate (ug/m3)	Organic Carbon Mass (ug/m3)	Elemental Carbon (ug/m3)	Water (ug/m3)	Ammonium (ug/m3)	OPP (ug/m3)
22.57	0 .5	22.07	8.51	1.11	5.21	0.91	2.31	3.31	0.71

The blank mass is subtracted before species fractions are calculated because the blank mass is held constant at 0.5 ug/m3 throughout the analysis. In the example, the measured FRM mass for quarter 3 is 22.57 ug/m3. The non-blank FRM mass is 22.07 ug/m3. The mass of the seven

[45]The species concentrations can be derived from co-located speciation data, nearby speciation data or from interpolated spatial fields. In this example, FRM mass is not interpolated.

species add up to the non-blank mass.

(4) Species fractions are calculated for each quarter for each species. In the example below, a fraction of non-blank mass is calculated for each of the seven species. Blank mass remains fixed at 0.5 ug/m3.

Table 5.3

FRM Mass (ug/m3)	Blank Mass (ug/m3)	Non-blank Mass (ug/m3)	% Sulfate	% Nitrate	% Organic aerosol	% Elemental Carbon	% Water	% Ammonium	% OPP
22.5	0 .5	22.0	38.6	5.0	23.6	4.1	10.5	15.0	3.2

The percentages in table 5.3 above are the relative composition for the 3rd quarter of 2003 and 2004. It is assumed that these species fractions are representative of the 2000-2004 time period[46].

(5) The weighted quarterly average FRM design values are used as the baseline FRM value for each monitoring site (2000-2004). The species fractions from the 2003/2004 speciation data were used to estimate the species concentrations for the baseline year FRM $PM_{2.5}$ data. The percentage compositions for 2003/2004 are applied to the quarterly weighted average design values as shown in table 5.4. In the example below, the weighted average design value for the 3rd quarter for the site from table 5.1 is 20.30 ug/m3. This leads to the following concentrations of $PM_{2.5}$ species:

Table 5.4 Calculation of the "current" species concentrations

Weighted Avg. FRM Mass (ug/m3)	Blank Mass (ug/m3)	Non-blank Mass (ug/m3)	Sulfate (ug/m3)	Nitrate (ug/m3)	Organic aerosol (ug/m3)	Elemental Carbon (ug/m3)	Water (ug/m3)	Ammonium (ug/m3)	OPP (ug/m3)
20.30	0 .5	19.8	7.64	0.99	4.67	0.81	2.08	2.97	0.63

This procedure is repeated for each $PM_{2.5}$ site and quarter to complete the calculation of current (baseline) ambient concentrations used as the basis for future estimates of $PM_{2.5}$ mass and its components.

[46]If there are less than 5 years of speciation data, then an assumption of representativeness needs to be made. It is generally assumed that the inter-annual variability of the species fractions is small compared to the variability of species concentrations. In the near future, 5 full years of concurrent FRM and speciation data will be available.

(6) Modeled results show the following relative response factors (RRF) in predicted mass of 5 components of $PM_{2.5}$ for the 3rd quarter:

Table 5.5

RRF Sulfate	RRF Nitrate	RRF Organic aerosol	RRF Elemental Carbon	RRF OPP
0.876	0.943	0.971	0.932	1.042

(7) The quarterly mean RRFs from table 5.5 are multiplied by the weighted quarterly average species concentrations from table 5.4 to derive future year concentrations.

From the example above, the future year 3rd quarter concentrations are:

$Sulfate_{Future}$ = 7.64 * 0.876 = 6.69 ug/m3
$Nitrate_{Future}$= 0.99 * 0.943 = 0.93 ug/m3
Organic carbon $mass_{Future}$= 4.67 * 0.971 = 4.53 ug/m3
Elemental $Carbon_{Future}$= 0.81 * 0.932 = 0.75 ug/m3
OPP_{Future}= 0.63 * 1.042 = 0.66 ug/m3

(8) The future year concentrations derived in step 5 are used to calculate the future year concentration of ammonium (if the direct ammonium RRF is not used) and particle bound water.

The future year ammonium concentrations are calculated from the sulfate, nitrate, and (current year) DON values. Assuming that the DON is unchanged from the current year[47], the ammonium is calculated using the following formula:

$NH4_{future} = DON * SO4_{future} + 0.29*NO3_{future}$,

In the example above, assuming the base year DON is 0.336,

[47]In the CAIR analysis, the DON was assumed to stay constant through time due to the uncertainty in the ammonium measurements. The water calculation is sensitive to the ammonium (and therefore the DON value) concentrations. Keeping the DON constant allows for the future year ammonium and water values to be solely a function of the change in sulfate and nitrate concentrations. Otherwise, the water concentration can go up when the sulfate and nitrate concentrations go down. This may occur if sulfate becomes more neutralized in the future. It is a somewhat illogical outcome (although scientifically possible) and is highly dependent on an uncertain measurement (ammonium). Therefore, use of a constant DON creates a more stable set of calculations. If the measured and modeled ammonium concentrations are believed to be accurate and respond in a reasonable way to emissions controls, then it would be more scientifically credible to use the model predicted change in ammonium. Otherwise, it is a reasonable assumption to keep the DON constant over time.

$Ammonium_{Future}$= 0.336 * 6.69 + 0.29 * 0.93 = 2.52 ug/m3

The $NH4_{future}$, $SO4_{future,}$ and $NO3_{future}$ concentrations can then be input into an equilibrium model (AIM or another alternative model) or through a polynomial equation to predict future year particle bound water concentration.

(9) The future species concentrations at each FRM site are then summed over the seven species plus blank mass to estimate the future quarterly average $PM_{2.5}$ concentration.

In the example above, the total $PM_{2.5Future}$=
6.69 + 0.93 + 4.53 + 0.75 + 0.66 + 2.52 + 1.73 + 0.5 = 18.31 ug/m3
(assuming that the future year water concentration is equal to 1.73 ug/m3)

(10) The same calculations are completed for the other 3 quarters to get a future year $PM_{2.5}$ concentration for each quarter. The 4 quarters are then averaged to get a final future year annual average $PM_{2.5}$ concentration for each FRM site.

(11) The future year annual average concentration is compared to 15.0 ug/m3[48].

5.2 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 5 years). In the test, we are interested in reducing the current design values at each site to $\leq$ 65 $\mu g/m^3$ [49,50].

---

[48]In the final step, the future year concentration should be rounded to the tenths digit. A (rounded) value of 15.0 ug/m3 meets the NAAQS. A value of 15.1 ug/m3 or greater violates the NAAQS.

[49]$PM_{2.5}$ nonattainment areas are not specifically designated for the annual or 24-hour standard. Therefore, the attainment demonstration must show that the area will meet both forms of the standard. Areas that only violate the annual standard and whose design value is well below the 24-hr standard may not need to apply the 24-hour modeled attainment test. Attainment test requirements should be discussed with the appropriate EPA Regional Office.

[50]The $PM_{2.5}$ 24-hour standard was lowered to 35 ug/m3 in 2006 (71 FR 61224) (40 CFR 50.13) . Planning requirements for areas that violated the old standard are unchanged. We expect the attainment test for the new standard to be the same or similar to the procedures in this version of the modeling guidance. As part of the implementation process for the revised standard, the procedures in the guidance will be re-evaluated. If necessary, we will release a revised modeling guidance document that addresses any changes to the 24-hour test for the new

Ideally, the modeled attainment test should reflect model results obtained for days in each season having observed $PM_{2.5}$ concentrations slightly below 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 response 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 $\mu g/m^3$. Alternatively, the test can focus on the high end of the distribution of days in each quarter, (e.g. the top 25% of $PM_{2.5}$ days) assuming that many of the high days violate the NAAQS (or are representative of days that violate the NAAQS). As with the annual NAAQS (and for the same reasons), the preferred approach is to develop relative response 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, 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 5 days per quarter then the analysis should focus on the high end of the distribution of $PM_{2.5}$ days in each quarter (e.g. the top 25% of $PM_{2.5}$ days[51]).

The 24-hour attainment test should be based on the same 5 year weighted average methodology that was used for the annual standard, with some slight modifications. The 24-hour design values are calculated from the 98th percentile value for each year. We recommend applying the attainment test on an annual basis and then averaging the annual average results up to 3 year averages and then averaging again to get a 5 year weighted average.

$PM_{2.5}$ is a complex mixture of components that have significant seasonal variations. If the observed 98th percentile value came from the 1st quarter, it is possible that an observed concentration from a different quarter, which may have a slightly lower concentration, could become the 98th percentile value in the future. It is important to demonstrate that the standard will be met for all seasons. Therefore we recommend applying the test by quarter for each year[52].

---

standard.

[51]The top 25% of days may seem like too many days per quarter to use for a 98th percentile based standard, but for sites with a 1 in 6 day sampling schedule meeting minimal completeness criteria (assuming 11 samples per quarter), the top 25% of days is only 3 days per quarter. For most sites, the top 25% of monitored days per quarter will represent between 3 and 8 days.

[52]In some areas it may not be necessary to model and evaluate the NAAQS for all quarters. For example, if $PM_{2.5}$ concentrations only exceed the NAAQS in the 1st and 4th quarters, and concentrations in the 2nd and 3rd quarters are very low, then it may not be necessary

Similar to the annual $PM_{2.5}$ attainment test, we recommend interpolation techniques for FRM monitors that do not have co-located speciation data. Because the 24-hour standard is a 98th percentile based value, the component composition on high concentration days may be highly variable from day to day and from site to site. Therefore, while interpolation techniques may need to be used, we strongly recommend collecting speciation data at all FRM sites that violate the 24-hour NAAQS. A precise estimate of the $PM_{2.5}$ components at violating sites will help reduce uncertainty in the future year concentration estimates.

We recommend a modeled attainment test for the 24-hour $PM_{2.5}$ NAAQS with 6 steps.

Step 1. Compute *observed* 98th percentile values for each year and next highest concentrations for each quarter.

The first step in projecting the daily design value is to identify the maximum daily average $PM_{2.5}$ concentration in each quarter that is less than or equal to the annual 98th percentile value over the entire year. This results in data for each year (for five years) for each site which contains one quarter with the 98th percentile value and three quarters with the maximum values from each quarter which are less than or equal to the 98th percentile value[53].

Step 2. Identify the species composition for each quarter for each monitor.

These quarterly $PM_{2.5}$ concentrations are then separated into their component species by multiplying the quarterly maximum daily concentration at each site by the estimated fractional composition of $PM_{2.5}$ species, by quarter and year, based on observed species fractions for the days > 65 ug/m3 or from the days in the high end of the distribution of $PM_{2.5}$ concentrations (using the same methodology as used in the annual average calculations[54]) (e.g., 30% nitrate x 70.0 ug/m^3 $PM_{2.5}$ = 21.0 ug/m^3 nitrate).

---

to model the full year. But for areas that have monitored violations (or high values that are close to the NAAQS) in all 4 seasons, the entire year should be evaluated.

[53]The test should be performed for each monitoring site that meets the data completeness criteria for calculating a 98th percentile value under the 24-hr NAAQS. The highest concentration from the other 3 quarters (less than or equal to the 98th percentile) is taken from any values that are measured at the site. It is necessary to examine the concentrations in other quarters which are less than the design value because PM species may have different responses to emissions controls in different parts of the year. Basecase concentrations that are less than the design value concentrations may end up being higher than the design value concentration in the future (after RRFs are applied). There may not always be data available for all four quarters.

[54]The ambient speciation data for the 24-hour attainment test should use the same adjustments that were recommended for the annual test. The adjusted data should be carefully examined. Adjustments to daily concentrations of speciation data (used for the 24-hr test) may in some cases be large compared to adjusted quarterly average data.

Step 3. Using model results, derive component-specific relative response factors (RRF) at each monitor for each quarter.

For each quarter, apply an air quality model to estimate current and future concentrations for each of the components of $PM_{2.5}$. Take the ratio of future to current predictions for each component on the modeled days with $PM_{2.5}$ > 65 ug/3 or the high end of the distribution of modeled days. The result is a component-specific relative response factor (RRF). (e.g., given model predicted base year nitrate of 20.0 ug/m^3 and future year concentration of 15.0 ug/$m^{3,}$ then RRF for nitrate is 0.750).

The relative response 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 mean concentration (for the high modeled days for each quarter) predicted at or near the monitoring site with emissions characteristic of the period used to calculate the baseline design value for 24-hour $PM_{2.5.}$

$C_{j,\ projected}$ is the future year mean concentration (for the high modeled days for each quarter) predicted at or near the monitoring site with emissions characteristic of the future year period.

Step 4. Apply the component specific RRFs to observed air quality to obtain projected quarterly species estimates.

For each quarter, multiply the quarterly component concentration (step 2) times the component-specific RRF obtained in step 3. This leads to an estimated future quarterly concentration for each component. (e.g., 21.00 ug/m^3 nitrate x 0.750 = future nitrate of 15.75 ug/m^3).

Step 5. Sum the quarterly components to get quarterly "potential" 98th percentile $PM_{2.5}$ values.

Sum the quarterly components to get quarterly "potential" 98th percentile $PM_{2.5}$ values. Repeat this procedure for each quarter and for each of the 5 years of ambient data. The highest daily value (from the 4 quarterly values) for each year at each monitor is considered to be the estimated 98th percentile value for that year.

Step 6. Calculate future year 5 year weighted average 24-hour design values and compare to the NAAQS.

The estimated 98th percentile values for each of the 5 years are averaged over 3 year intervals to create design values and then the 3 design values are averaged to create the 5 year weighted average for each monitor. Compare the projected 24-hour design values to 65 ug/m3.

The recommended test is illustrated in example 5.2. In the interest of brevity, some of the steps are only illustrated for quarter 1. The actual calculations would also use analogous data for the other three quarters.

Example 5.2

Given: (1) Area "D" has 1 monitoring site. The 98th percentile values for 5 years of observed data are as follows (year 3 is being modeled for the attainment test):

Year 1- 80 ug/m3
Year 2- 68 ug/m3
Year 3- 71 ug/m3
Year 4- 58 ug/m3
Year 5- 74 ug/m3

Step 1- Compute *observed* 98th percentile values for each year and next highest concentrations for each quarter.

The maximum quarterly values from the 5 years which are equal to or less than the 98th percentile concentration are as follows:

	Year 1	**Year 2**	**Year 3**	**Year 4**	**Year 5**
Quarter 1	80.0 ug/m3	65.0	57.0	58.0	74.0
Quarter 2	67.6	45.5	46.4	45.3	52.2
Quarter 3	41.7	47.6	32.5	39.4	45.3
Quarter 4	52.8	68.4	71.3	49.5	64.2

Step 2- Identify the species composition for each quarter for each monitor.

The mean species composition[55] for the observed high $PM_{2.5}$ days in quarter 1 is:

Nitrate- 30%
Sulfate- 10%
Organic carbon- 23%
Elemental carbon- 5%
Other primary $PM_{2.5}$- 5%
Ammonium- 15%
Water- 12%

This leads to the following baseline concentrations for quarter 1 (ug/m3):

	$PM_{2.5}$	**Nitrate**	**Sulfate**	**OCM**	**EC**	**OPP**	**Ammonium**	**Water**
Year 1	79.5	23.85	7.95	18.29	3.98	3.98	11.93	9.54
Year 2	64.5	19.35	6.45	14.84	3.23	3.23	9.68	7.74
Year 3	56.5	16.95	5.65	13.00	2.83	2.83	8.48	6.78
Year 4	57.5	17.25	5.75	13.23	2.88	2.88	8.63	6.90
Year 5	73.5	22.05	7.35	16.91	3.68	3.68	11.03	8.82

Step 3- Using model results, derive component-specific relative response factors (RRF) at each monitor for each quarter.

The following are RRFs for the 1st quarter derived from the modeled baseline and future year concentrations for the high $PM_{2.5}$ days.

	Nitrate	**Sulfate**	**OCM**	**EC**	**OPP**
RRF	0.942	0.843	0.917	0.930	1.040

[55]The percentages are derived from several steps in the processing. The nitrate, sulfate, elemental carbon, and other primary PM2.5 are derived from measured or interpolated data. The ammonium is calculated from interpolated sulfate, nitrate, and DON (or directly interpolated ammonium). Water is calculated from ammonium, nitrate, and sulfate using a model or equation. The OMC is calculated by mass balance.

Step 4- Apply the component specific RRFs to observed air quality to obtain projected quarterly species estimates.

The quarter 1 concentrations for each year are multiplied by the component RRFs. Ammonium concentrations are calculated from the future year sulfate, nitrate, and (baseline) DON value. Water concentrations are calculated from the future year nitrate, sulfate, and ammonium concentrations, using an empirical equation.

	Nitrate	**Sulfate**	**OCM**	**EC**	**OPP**	**Ammonium**	**Water**
Year 1	22.47	6.70	16.77	3.70	4.14	10.41	8.41
Year 2	18.23	5.44	13.60	3.01	3.36	8.45	6.83
Year 3	15.97	4.76	11.92	2.63	2.94	7.40	5.98
Year 4	16.25	4.85	12.13	2.68	3.00	7.53	6.09
Year 5	20.77	6.20	15.50	3.43	3.83	9.62	7.78

Step 5. Sum the quarterly components to get quarterly "98th percentile" $PM_{2.5}$ values.

The year 1 $PM_{2.5}$ concentration is equal to the sum of the 7 components plus 0.5 ug/m3 blank mass:

Year 1= 22.47 + 6.70 + 16.77 + 3.70 + 4.14 + 10.41 + 8.41 + 0.5 = 72.6

Filling in the quarter 1 values for the other 4 years, plus the other 3 quarters for each year; the full matrix of quarterly values is as follows:

	1999	**2000**	**2001**	**2002**	**2003**
Quarter 1	72.6 ug/m3	58.9	51.6	52.5	67.1
Quarter 2	57.5	38.7	39.4	38.5	44.4
Quarter 3	38.8	44.3	30.2	36.6	42.1
Quarter 4	46.6	60.5	63.0	43.6	57.1

Step 6. Calculate future year 5 year weighted average 24-hour design values and compare to the NAAQS.

In the example, the highest quarterly value for each year in step 5 represents the future year

annual 98th percentile. Therefore, the 3 year average design values are:

DV1 = (72.6 + 60.5 + 63.0)/3 = 65.4
DV2 = (60.5 + 63.0 + 52.5)/3 = 58.7
DV3 = (63.0 + 52.5 + 67.1)/3 = 60.9

And finally, the 5 year weighted average is equal to the average of the 3 future year design values:

5 year weighted average future DV= (65.4 + 58.7 + 60.9)/3 = 61.7 = **62 ug/m3**

5.3 Special Considerations for Primary $PM_{2.5}$

Primary particulate matter does not undergo 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. Therefore, 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 (OPP) and elemental carbon (EC) components of ambient $PM_{2.5}$ (these are often called "hotspots"). 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.

The $PM_{2.5}$ NAAQS rulemaking (U.S. EPA 1997a) as well as ambient monitoring rule language[56] indicates that high $PM_{2.5}$ concentrations occuring in "unique" population oriented areas should not be compared to the annual NAAQS and should only be compared to the 24-hour NAAQS (especially when caused by a single source). Furthermore, high $PM_{2.5}$ concentrations measured in non-population oriented areas should not be compared to either of the $PM_{2.5}$ NAAQS. $PM_{2.5}$ monitor siting guidance (U.S. EPA, 1997b) defines the appropriate scales of influence for the $PM_{2.5}$ monitoring network.

High primary PM concentrations can occur at (or near) existing monitors or in unmonitored areas. The modeled attainment test is primarily a monitor based test. As such, the

[56] 40 CFR part 58 subpart D 58.30

focus of the attainment test is whether attainment can be reached at existing monitors. To address the issue of $PM_{2.5}$ concentrations in unmonitored areas, we have recommended an "unmonitored area analysis" (see section 3.4). The unmonitored area analysis is intended to be the primary means for identifying high $PM_{2.5}$ concentrations outside of traditionally monitored locations. The spatial resolution of the modeling that is the underlying basis of the unmonitored area analysis will determine how well it addresses primary PM hotspot issues. The finer the resolution of the grid model, the more likely that primary PM hotspots will be recognized. Based on the monitoring guidance, we believe that an unmonitored area analysis conducted at 12km or finer resolution is sufficient to address unmonitored $PM_{2.5}$ for the annual NAAQS. Conducting the unmonitored analysis at 4km or finer resolution will provide an even more detailed analysis of the spatial gradients of primary $PM_{2.5}$, especially when evaluating violations of the 24-hr. NAAQS.

5.3.1 Analysis of Primary $PM_{2.5}$ Impacts at Monitors

Most $PM_{2.5}$ monitors are located at "urban scale" sites. These sites tend to represent relatively large spatial scales and do not have large gradients compared to other monitors in the area. Some sites are classified as "neighborhood" scale and may be influenced by more local sources. For reasons stated above, local influences creating spatial gradients are likely to consist mostly of primary $PM_{2.5}$ (OC, EC, and other primary particulates). These sources may be local point sources, or they may be nearby roads or other mobile or area sources.

It may be necessary to evaluate the impact of local primary PM sources for contributions to both the 24-hour and annual NAAQS. As stated earlier, it may not be appropriate to compare population oriented FRM sites that are dominated by point sources, to the annual NAAQS. But there are numerous cases where the impact from local sources is not dominant, but a sizable contributor to total $PM_{2.5}$ (~10-30% of total annual average $PM_{2.5}$). In these cases, a more refined analysis of the contribution of local primary $PM_{2.5}$ sources to $PM_{2.5}$ at the monitor(s) will help explain the causes of nonattainment at the monitor and may lead to the more efficient ways to attain the standard by controlling emissions from local sources.

There are several modeling tools that can be used to evaluate contributions of local $PM_{2.5}$ sources. A grid model can be run at very high resolution (down to 1 or 2 km) or a Gaussian dispersion can be used. Grid based models usually simulate chemical transformation and complex meteorological conditions, while dispersion models are generally more simplistic; being limited to a local-scale and using Gaussian approximations with little or no chemistry. Therefore, while dispersion models may not be an appropriate tool for determining secondary $PM_{2.5}$ concentrations, they work well for use in determining local primary $PM_{2.5}$ impacts in a small area. The model(s) and model setup should be evaluated to determine the most appropriate tools for a specific situation.

Regardless of which type of models are used to evaluate changes in primary $PM_{2.5}$ at monitors, we recommend that the model results be used in a relative manner. This is consistent with the already described attainment test. If a grid model is used at very high resolution, the

attainment test as described in Section 5.1 should be followed. If a Gaussian dispersion model is used then the application of the attainment test will vary slightly. The test will need to combine results from the photochemical grid based modeled attainment test and the results from local-scale dispersion modeling. If such an approach is taken, the suggested variations on the attainment test are discussed in the following section.

5.3.2 Analysis of Primary $PM_{2.5}$ Impacts at Monitors using a Gaussian Dispersion Model

To apply a dispersion model in an attainment test, it becomes important to determine the local component of primary $PM_{2.5}$ at the monitor and the sources that are contributing to that component. There is no single, simple method for quantifying this contribution, but detailed analysis of ambient data and advanced data analysis techniques, such as receptor modeling, may help quantify the contribution. The simplest method for identifying the local component of $PM_{2.5}$ is to examine local ambient $PM_{2.5}$ concentrations. For this analysis, it is important to identify the local contributions from as small an area as possible. This will make it easier to identify contributing sources. It is most appropriate to examine monitored concentration differences between urban monitors (with the highest concentrations) and more suburban measurements. This is likely to be representative of the excess contribution from a relatively local area. "Urban excess" calculations which pair an urban monitor with a rural background monitor (U.S. EPA, 2004b) are more likely to indicate "local" contributions that may be more representative of an entire metropolitan area. In most cases, the local component will include contributions from more than one source.

Sources identified as contributing to the monitor will be modeled with a dispersion model (or alternatively, a very high resolution grid model) . It is common practice to run dispersion models for limited numbers of point sources in relatively small areas (out to a distance of ~50 km). Dispersion models can also be run with all sources of primary PM in a limited area (including mobile, non-road, and area sources).

When applying the model to evaluate changes in primary $PM_{2.5}$, one should determine the $PM_{2.5}$ species that make up the local primary particulate matter[57]. We recommend that the individual components of primary $PM_{2.5}$ are tracked separately in the dispersion model. This will allow more thorough evaluation of the dispersion modeling results (comparing model output to speciated ambient data) and may aid in the development of primary $PM_{2.5}$ control strategies.

The majority of the primary PM will consist of the primary portion of the organic carbon

---

[57]The $PM_{2.5}$ emissions should be broken out into individual species using the best information available. Typically, SCC specific speciation profiles are used to "speciate" the PM into individual components. Local, source specific information should be used whenever possible.

component, elemental carbon (EC), and the general category of "other primary particulate matter" (OPP). In some cases, directly emitted sulfate (and in rare cases nitrate) may also be a significant component of the local primary $PM_{2.5}$.

The dispersion modeling results should be evaluated to ensure adequate model performance. Similar to grid modeling, the dispersion model results should be compared to ambient data to ensure that the model is working well. Although section 18 of this guidance is geared towards evaluating grid models, many of the same statistical calculations can be made for primary $PM_{2.5}$ and $PM_{2.5}$ components predicted by a Gaussian dispersion model. Since secondary $PM_{2.5}$ is often a large component of total $PM_{2.5}$ concentrations, it may be difficult to separate the primary and secondary components of ambient $PM_{2.5}$. EC and OPP should be considered to be primary $PM_{2.5}$. Much of the rest of the primary $PM_{2.5}$ concentration will be primary OC.

As part of the analysis, an estimated concentration of primary OC is needed. There are several methods available for estimating the primary vs. secondary portion of ambient OC. Among these are the EC tracer method and receptor modeling. The EC tracer method is the most common method used to estimate secondary and primary OC concentrations (Turpin, 1995), (Strader, 1999) (Cabada, 2004), (Chu, 2005), (Saylor, 2006) . The method uses measurements of OC and EC and calculated OC to EC ratios to identify periods when OC is likely to be mostly primary. This information is then used to calculate the secondary contribution to OC . Receptor models such as CMB and PMF have also been used to estimate secondary organic concentrations (Na, 2004), (Yuan, 2006).

The following sections discuss a suggested methodology for examining decreases or increases in PM concentrations due to local primary sources for the Annual and 24-hour $PM_{2.5}$ NAAQS. Because each nonattainment area has unique emissions sources and source-receptor relationships, States should work closely with their EPA Regional Office in developing local area analysis applications.

Methodology for the Annual $PM_{2.5}$ NAAQS

The suggested steps outlined below assume that the attainment test has been performed using the photochemical modeling and that the annual DV has been computed as discussed in Section 5.1.

(1) Identify local primary $PM_{2.5}$ sources that are thought to be are contributing to the monitor and causing non-attainment of the annual NAAQS.

(2) By quarter, identify at the monitor, primary versus secondary components of $PM_{2.5}$.

(3) Identify the percentage of total primary $PM_{2.5}$ at the monitor due to the local sources identified in (1). Convert the percentage to an estimated concentration in ug/m3

(4) Model with a dispersion model the current and future concentrations of these local sources using actual emissions.

-Place receptors at and near the monitor location(s)[58].

-The receptor spacing should be fine enough to get an accurate picture of the concentration at the monitor.

-Model the base year and future annual average

-For both base and future year use the same 1 year of met data – the same year as used for the photochemical modeling (if available, use more than 1 year of met data).

(5) From the modeling in (4), calculate quarterly RRFs[59] for total primary $PM_{2.5}$ modeled for the selected sources (using concentration values at receptors at monitors).

(6) Multiply the primary $PM_{2.5}$ identified in (3) by the RRFs from (5).

(7) Subtract the result in (6) from the primary $PM_{2.5}$ identified in (3)

(8) For each quarter, subtract the amount in (7) from the future year quarterly DV provided by the photochemical modeling.

(9) Average the four quarters to get an annual average DV.

Methodology for the 24-Hour NAAQS

The suggested steps outlined below assume that the attainment test has been performed using the photochemical modeling and that the quarterly DV's have been computed as discussed in Section 5.2. As with the annual design value, monitoring data for 5 years are used as the basis for the projection.

(1) Determine the local primary $PM_{2.5}$ sources that are thought to be contributing to the monitor for each of the 4 quarters[60].

---

[58]Additional receptors should be placed near monitors to examine predicted concentration gradients. Receptors should only be located in areas that are appropriate for placing FRM monitors for comparison to the annual NAAQS (see EPA, 1997b).

[59]RRFs should be calculated as a ratio of the base and future mean concentrations (a ratio of the means, not a mean of the daily ratios).

[60]The selection of these sources should be representative of the 5-year period used to calculate the 5 year weighted average design values. If large emissions changes have occurred over the 5 year period then it may be necessary to determine the local contribution of identified

(2) For each of the four representative quarters, determine the primary versus secondary components of $PM_{2.5}$ at the monitor on the 98th percentile day and all high days.

(3) Identify percentage of total primary $PM_{2.5}$ at monitor due to the local sources identified in (1). Convert the percentage to an estimated concentration in ug/m3.

(4) Model with a dispersion model the current and future concentrations of these local sources using actual emissions that are representative of emissions on high $PM_{2.5}$ days[61].

-Place receptors at and near the monitor location(s)[62].
-The receptor spacing should be fine enough to get an accurate picture of the concentration at the monitor.
-Model the base and future year.
-For both base and future year use the same 1 year of met data used for the photochemical modeling (if available, use more than 1 year of met data).
-Calculate the 24-hour average for each day. Average the high days[63] within each quarter to determine a modeled high concentration for each quarter in the base and future year.

(5) From the modeling in (4), calculate quarterly RRFs[64] for total primary $PM_{2.5}$ modeled for the selected sources.

(6) Multiply the primary $PM_{2.5}$ identified in (3) by the RRFs from (5).

---

sources for each year. The identification of the contribution to the monitor should be most influenced by the modeling year (the middle year of the 5 year period) since it is the year being modeled and it has the strongest weighting in the design value calculations.

[61]Because the test is relative, in most cases, actual emissions should be used. The actual emissions should be representative of emissions on high $PM_{2.5}$ days (days that exceed the NAAQS). Since the absolute predicted concentrations are not used directly, allowable emissions may overestimate the changes in concentrations due to the identified sources. This should be evaluated on a case-by-case basis. Allowable emissions may be appropriate if there are sources that often emit above their typical levels on high $PM_{2.5}$ days.

[62]Additional receptors should be placed near monitors to examine predicted concentration gradients. Receptors should only be located in areas that are appropriate for placing FRM monitors for comparison to the 24-hour NAAQS (see EPA, 1997b).

[63]High days may be all days > 65 ug/m3 or the high end of the distribution of modeled days (e.g top 25% days)

[64]RRFs should be calculated as a ratio of the base and future mean concentrations (a ratio of the means, not a mean of the daily ratios).

(7) Subtract the result in (6) from the primary $PM_{2.5}$ identified in (3)

(8) For each of the 5 years used to calculate the 5 year weighted average design value, subtract the result from (7) from each quarter for each year. The "new" annual 98th percentile value is the highest quarterly concentration in each year.

(9) Average the annual 98th percentile concentrations to get 3 design values and then average the 3 DV to get the final future year 5 year weighted average DV.

Annual NAAQS Example:

For quarter 4, a monitor has a 5 year weighted mean concentration of 17.0 ug/m3.

(1) It is believed that several steel mills in the area are contributing to the local component of $PM_{2.5}$ at the monitor

(2) Based on comparisons to other monitors in the area, it is believed that the local $PM_{2.5}$ component at the monitor for quarter 4 is 2 ug/m3.

(3) Using receptor modeling, it is determined that, on average, the steel mills are contributing a local $PM_{2.5}$ component in quarter 4 of approximately 1 ug/m3.

(4) Modeling with a dispersion model shows the following contribution at the monitor from the identified local sources (steel mills):
Base: 2.0 ug/m3
Future: 1.5 ug/m3

(5) Quarter 4 RRF_local = 1.5/2.0 = 0.750.

(6) 1 ug/m3 * 0.750 = 0.75 ug/m3.

(7) 0.75 ug/m3 - 1 ug/m3 = -0.25 ug/m3 (total Q4 reduction from identified local sources is 0.25 ug/m3)

(8) Quarter 4 New DV = 17.0 – 0.25 ug/m3 = 16.75 ug/m3

(9) Steps 1-8 are repeated for each quarter with similar results. The annual DV = (Q1 New DV + Q2 New DV + Q3 New DV + Q4 New DV)/4

24 Hour NAAQS Example:

The baseline design values are calculated from the 5 year base period centered around the modeling year. The dispersion model is run with emissions and meteorology for year 3 (the

center year of the 5 year period). The maximum quarterly values from the 5 years which are equal to or less than the 98th percentile concentration are as follows (from example 5.2):

	Year 1	**Year 2**	**Year 3**	**Year 4**	**Year 5**
Quarter 1	80.0 ug/m3	65.0	57.0	58.0	74.0
Quarter 2	67.6	45.5	46.4	45.3	52.2
Quarter 3	41.7	47.6	32.5	39.4	45.3
Quarter 4	52.8	68.4	71.3	49.5	64.2

The future year $PM_{2.5}$ concentrations from the photochemical modeling described in Section 5.2, Step 5 are as follows:

	Year 1	**Year 2**	**Year 3**	**Year 4**	**Year 5**
Quarter 1	72.6 ug/m3	58.9	51.6	52.5	67.1
Quarter 2	57.5	38.7	39.4	38.5	44.4
Quarter 3	38.8	44.3	30.2	36.6	42.1
Quarter 4	46.6	60.5	63.0	43.6	57.1

(1) It is believed that several steel mills in the area are contributing to the local component of $PM_{2.5}$ at the monitor

(2) Based on comparison to other monitors in the area, it is believed that the primary local $PM_{2.5}$ component at the monitor (on high PM days) is the following for each of the quarters in 2001:
Quarter 1: 15 ug/m3
Quarter 2: 13 ug/m3
Quarter 3: 8 ug/m3
Quarter 4: 20 ug/m3
It is believed that this is representative of the other four years within the 5-year period.

(3) Using receptor modeling, it is determined that the steel mills are contributing a local $PM_{2.5}$ component (on high PM days) of the following for each of the quarters in 2001:
Quarter 1: 8 ug/m3
Quarter 2: 4 ug/m3
Quarter 3: 4 ug/m3
Quarter 4: 5 ug/m3

It is believed that this is representative of the other four years within the 5-year period.

(4) Modeling with a dispersion model shows the following contribution at the monitor from the identified local sources (steel mills) for total primary $PM_{2.5}$ in each of the quarters in 2001:
Quarter 1: Base: 11.7 ug/m3 Future: 8.3 ug/m3
Quarter 2: Base: 6.3 ug/m3 Future: 4.4 ug/m3
Quarter 3: Base: 5.3 ug/m3 Future: 3.1 ug/m3
Quarter 4: Base: 8.3 ug/m3 Future: 6.2 ug/m3
For each quarter, we computed the 24-hr average for each day and computed the mean of the high days for each quarter to get the base and future year concentrations.

(5) Based on the results from the modeling in (4), the RRF values are:
Quarter 1: 8.3/11.7 = 0.709
Quarter 2: 4.4/6.3 = 0.698
Quarter 3: 3.1/5.3 = 0.585
Quarter 4: 6.2/8.3 = 0.747

(6) Multiplying the primary $PM_{2.5}$ identified in (3) by the RRFs from (5).
Quarter 1: 8 ug/m3 * 0.709 = 5.7 ug/m3
Quarter 2: 4 ug/m3 * 0.698 = 2.8 ug/m3
Quarter 3: 4 ug/m3 * 0.585 = 2.3 ug/m3
Quarter 4: 5 ug/m3 * 0.747 = 3.7 ug/m3

(7) Subtracting the result in (6) from the primary $PM_{2.5}$ identified in (3)
Quarter 1: 5.7 ug/m3 - 8 ug/m3 = -2.3 ug/m3
Quarter 2: 2.8 ug/m3 - 4 ug/m3 = -1.2 ug/m3
Quarter 3: 2.3 ug/m3 - 4 ug/m3 = -1.7 ug/m3
Quarter 4: 3.7 ug/m3 - 5 ug/m3 = -1.3 ug/m3

(8)Subtracting the result from (7) from each quarter for each of the five years:

Year 1:
Q1: 72.6 ug/m3 - 2.3 ug/m3 = 70.3 ug/m3
Q2: 57.5 - 1.2 = 56.3
Q3: 38.8 - 1.7 = 37.1
Q4: 46.6 - 1.3 = 45.3

Year 2:
Q1: 58.9 - 2.3 = 56.6
Q2: 38.7 - 1.2 = 37.5
Q3: 44.3 - 1.7 = 42.6
Q4: 60.5 - 1.3 = 59.2

Year 3:
Q1: 51.6 - 2.3 = 49.3
Q2: 39.4 - 1.2 = 38.2
Q3: 30.2 - 1.7 = 28.5
Q4: 63.0 - 1.3 = 61.7

Year 4:
Q1: 52.5 - 2.3 = 50.2
Q2: 38.5 - 1.2 = 37.3
Q3: 36.6 - 1.7 = 34.9
Q4: 43.6 - 1.3 = 42.3

Year 5:
Q1: 67.1 - 2.3 = 64.8
Q2: 44.4 - 1.2 = 43.2
Q3: 42.1 - 1.7 = 40.4
Q4: 57.1 - 1.3 = 55.8

The "new" 98th percentile values are:

Year 1: 70.3
Year 2: 59.2
Year 3: 61.7
Year 4: 50.2
Year 5: 64.8

(9) Averaging the 3 design values:
DV1 = (70.3 + 59.2 + 61.7)/3 = 63.7
DV2 = (59.2 + 61.7 + 50.2)/3 = 57.0
DV3 = (61.7 + 50.2 + 64.8)/3 = 59.9

Averaging the 3 DV to get the final future year 5 year weighted average DV:
(63.7 + 57.0 + 59.9)/3 = 60.2 **= 60 ug/m3**

In the above examples, emissions reductions from steel mills were modeled with a dispersion model, which provided additional reductions to the future year 5 year weighted average annual and 24-hour design values. Changes in design values derived from photochemical modeling should be evaluated to determine the magnitude (on concentrations) of emissions changes from the flagged steel mills. The change in concentration in the photochemical model from emissions controls on the steel mills may need to be subtracted from the final design value to ensure that there is no double counting of emission reductions.

Double Counting

The methodology discussed above may result in double counting of local emissions sources and reductions. The change in emissions from local sources is counted once in the photochemical modeling and again in the dispersion modeling. There are several ways to account for this issue.

1) Run the photochemical model without the primary $PM_{2.5}$ emissions from the flagged sources in both the base case and future case(s).

2) Run the grid model with a source tagging approach. The flagged sources can be tagged so that their contribution in the grid model can be explicitly tracked (Douglas, 2006) (Environ, 2006a), (Environ, 2006b).

3) Employ a methodology to combine the results of the grid and dispersion models so that double counting is eliminated or minimized. Several techniques have been developed to do this (Isakov,2007).

4) Assume that double counting is small enough to be ignored.

The grid resolution of the photochemical model and the number of flagged sources in the dispersion model will determine the magnitude of the potential double counting. The higher the resolution of the grid model, the more double counting will be a problem (because the local sources will be more finely resolved). And the more sources that are flagged, the more double counting will be a problem. Therefore, the nature of the issue should be evaluated on a case by case basis. The simplest way to evaluate the magnitude of the problem is to run the photochemical model with and without the flagged sources from the dispersion model. This will indicate the maximum impact the sources can have on the photochemical modeling results. A very small impact means that double counting is not a problem. A large impact indicates that double counting needs to be explicitly accounted for.

Quality Assurance

As with any modeling exercise, it is important to quality assure the model inputs and outputs. In particular, we recommend a focused quality assurance check on emissions from sources flagged for a dispersion modeling analysis. Prior to applying a model, States should review available information to ensure that there are no major discrepancies between modeled estimates and nearby monitored data for particulate matter. The emission factors, activity data, and speciation profiles of the $PM_{2.5}$ emissions should also be analyzed for accuracy. If a speciation monitor exists, the speciated data from the monitor can be compared to the speciation profiles of the flagged sources. Receptor models can also be used as a QA tool. Discrepancies between receptor modeling results (which are based on analyzing ambient data) and speciated

emissions may indicate a problem with the magnitude and/or the speciation of the emissions sources. If discrepancies are found, those implementing the modeling protocol (see section 12) 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 identified sources.

It is also important to quality assure the model outputs. Modeling each species of primary $PM_{2.5}$ separately within the dispersion model should aid in this analysis, especially if the selected control strategies applied in the future year do not control each primary species to the same degree. If a speciation monitor exists, the speciated data from the monitor may also help quality assure the dispersion model results. However, the relative application of the dispersion model results will help to reduce the impact of possible over- or under-estimations by the dispersion model due to emissions, meteorology, or general selection of other model input parameters.

6.0 What Is The Recommended Modeling Analysis for Regional Haze?

In this section, we describe the recommended modeled test to assess visibility improvement. The visibility analysis has many similarities to the attainment tests described in Section 5. The recommended visibility analysis and the attainment tests both use monitored data to define current air quality. They divide $PM_{2.5}$ into major species, and component-specific relative response factors are multiplied by current monitored values to estimate future concentrations.

Section 6.1 provides background information on the requirements of the Regional Haze rule and defines the components of a reasonable progress analysis. The rest of the section describes the recommended modeling analysis which supports the regional haze rule.

6.1 Regional Haze Rule Background

Section 169A of the Clean Air Act states "Congress hereby declares as a national goal the prevention of any future, and the remedying of any existing, impairment of visibility in mandatory class I Federal areas which impairment results from manmade air pollution." Section 169B calls for EPA to "carry out the Administrator's regulatory responsibilities under [section 169A], including criteria for measuring 'reasonable progress' toward the national goal."

In response to these mandates, EPA promulgated the regional haze rule (RHR) on July 1, 1999.[65] Under section 51.308(d)(1) of this rule, States must "establish goals (expressed in deciviews) that provide for reasonable progress towards achieving natural visibility conditions for each Class I area within a State". These reasonable progress goals must provide for an improvement in visibility for the most impaired days over the period of the implementation plan and ensure no degradation in visibility for the least impaired days over the same period.[66] Reasonable Progress Goals (RPGs), measured in deciviews, are interim goals that represent incremental visibility improvement over time toward the goal of natural background conditions and are developed in consultation with other affected States and Federal Land Managers.[67]

In determining what would constitute reasonable progress, section 169A(g) of the Clean Air Act requires States to consider the following four factors:

- The costs of compliance;
- The time necessary for compliance;
- The energy and non-air quality environmental impacts of compliance; and

---

[65] 64 FR 35714 (codified at 40 CFR 51.300-309).

[66] 40 CFR 51.308(d)(1).

[67] 40 CFR 51.308(d)(1)(iv) and 51.308(i).

- The remaining useful life of existing sources that contribute to visibility impairment.[68]

States must demonstrate in their SIPs how these factors are taken into consideration in selecting the RPG for each Class I area in the State. More details on the setting of RPGs and the appropriate use of the four factors is contained in "Guidance for Setting Reasonable Progress Goals Under the Regional Haze Program" (U.S. EPA, 2007).

The regional haze rule also establishes an additional analytical requirement in the process of establishing RPG. States are required to determine the rate of improvement in visibility needed to reach natural conditions by 2064, and to set each RPG taking this "glidepath" into account.[69] The glidepath is the amount of visibility improvement needed in each review period to stay on a linear path towards visibility improvement to natural conditions by 2064. The glidepath represents a linear or uniform rate of progress. Therefore, the amount of visibility improvement needed to stay on the glidepath is also referred to as "uniform rate of progress". This document uses the terms "glidepath" and "uniform rate of progress" interchangeably.

This document focuses on the modeling analysis required to assess future visibility improvement relative to the glidepath or uniform rate of progress (for each Class I area). As mentioned previously, this document does not address the process of setting and evaluating reasonable progress goals. The actual determination of reasonable progress is a multi-step process which looks at the uniform rate of progress, the predicted rate of progress, and the four factor test. In the end, the reasonable progress goal will be determined by the states for each Class I area by considering the factors and the modeling analysis. The results of the modeling analyses will be one of several components in the determination of reasonable progress.

6.2 Uniform Rate of Progress

Regional haze is calculated by estimating light scattering and absorption by components of $PM_{2.5}$. Baseline conditions represent the visibility conditions which exist for the best and worst days at the time the regional haze program is established for each Class I area. Once established, the baseline represents the starting point from which the uniform rate of progress will be measured. The RHR also requires an estimate of "natural conditions" for each Class I area which represents the visibility conditions that would exist in the absence of man-made impairment.

As explained in the RHR, the baseline for each Class I area is the average visibility (in deciviews) for the 20% most impaired days, or "worst days", and for the 20% least impaired days, or "best days," for the years 2000 through 2004[70].

---

[68] 42 U.S.C. § 7491(g)(1); 40 CFR 51.308(d)(1)(i)(A).

[69] 40 CFR 51.308(d)(1)(i)(B).

[70] 64 FR at 35730.

States are asked to calculate the reduction in deciviews (on the worst visibility days) by an amount which is proportional to the number of years between the base period and the first review period vs. the number of years between the base period and 2064. This establishes the glidepath or uniform rate of progress. 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 glidepath for 2018 would be to reduce visibility impairment on the worst days by 6.3 *deciviews* (i.e., [35 -8][(2018-2004)/(2064-2004)]). More detailed examples of this calculation are contained in (U.S. EPA, 2003b).

The goal of the regional haze program is to return to natural conditions by 2064, and States are required to demonstrate, by the end of the first planning period (2018), reasonable progress toward meeting that goal. The U.S. EPA has developed guidance on how to track visibility for the regional haze rule (U.S. EPA, 2003b). In Section 6.3, we describe how the ambient trends in regional haze will be determined. The modeling assessment for uniform rate of progress needs to provide a good predictor for what the trend in observed air quality will be in the future. This is most likely to happen if we reproduce the calculations which will be applied to the ambient observations as closely as practical. In Section 6.4, we present and recommend a modeling analysis which closely replicates the ambient data analysis. Section 6.5 addresses how to develop key inputs for the recommended modeling assessment of uniform rate of progress.

6.3 How are Regional Haze Trends 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 by several different methodologies, 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 (6.1) may be applied in each Class I area to estimate b_{ext} (Sisler, 1996, IMPROVE, 2000). The equation reflects empirical relationships derived between measured mass of particulate matter components and transmissometer measurements of b_{ext} at monitoring sites in Class I areas within the IMPROVE network.

$$\mathbf{b_{ext} = 3(\mathit{f(rh)})[SO_4] + 3(\mathit{f(rh)})[NO_3] + 4(\mathit{f'(rh)})[OC] + 10[EC] + 1[Fine\ Soil] + 0.6[CM] + b_{rayleigh}} \qquad \mathbf{(6.1)}$$

where

the numerical coefficients on the right hand side of the equation represent the light scattering or absorption efficiency, m^2/gm of the corresponding 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, ug/m^3,

NO_3 is the mass associated with nitrates, ug/m^3,

OC is the mass associated with organic carbon, ug/m^3,

EC is the mass associated with elemental carbon, ug/m^3,

Fine Soil is inorganic primary particulate matter (excluding primary sulfate and nitrate particles) associated with soil components with aerodynamic diameter $\leq$ 2.5 um, μg/m^3,

CM is coarse particulate matter with aerodynamic diameter > 2.5 um, but $\leq$ 10 um, ug/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}.

New IMPROVE Algorithm

The IMPROVE program has recently revised the IMPROVE algorithm (IMPROVE, 2006); (Hand, 2006) . The new algorithm is intended to reduce biases in light extinction estimates compared to the old algorithm. The new algorithm is somewhat more complicated than the old algorithm as follows:

$$
\begin{aligned}
\textbf{bext} \approx\ & \mathbf{2.2 \times f_s(RH) \times [Small\ Sulfate] + 4.8 \times f_L(RH) \times [Large\ Sulfate]} \\
& \mathbf{+ 2.4 \times f_s(RH) \times [Small\ Nitrate] + 5.1 \times f_L(RH) \times [Large\ Nitrate]} \\
& \mathbf{+ 2.8 \times \{Small\ Organic\ Mass] + 6.1 \times [Large\ Organic\ Mass]} \\
& \mathbf{+ 10 \times [Elemental\ Carbon]} \\
& \mathbf{+ 1 \times [Fine\ Soil]} \\
& \mathbf{+ 1.7 \times f_{ss}(RH) \times [Sea\ Salt]} \\
& \mathbf{+ 0.6 \times [Coarse\ Mass]} \\
& \mathbf{+ Rayleigh\ Scattering\ (site\ specific)} \\
& \mathbf{+ 0.33 \times [NO_2\ (ppb)]} \qquad \mathbf{[6.2]}
\end{aligned}
$$

The total sulfate, nitrate and organic carbon compound concentrations are each split into two fractions, representing small and large size distributions of those components. The organic mass concentration used in the new algorithm is 1.8 times the organic carbon mass concentration, changed from 1.4 times carbon mass concentration used for input into the old IMPROVE algorithm. New terms have been added for sea salt (important for coastal locations) and for absorption by NO_2 (only used where NO_2 data are available). Site-specific Rayleigh scattering is calculated for the elevation and annual average temperature of each IMPROVE monitoring site.

The apportionment of the total concentration of sulfate compounds into the concentrations of the small and large size fractions is accomplished using the following equations.

[Large Sulfate] = [Total Sulfate]/20 ug/m3 x [Total Sulfate], for [Total Sulfate] < 20 ug/m3

[Large Sulfate] = [Total Sulfate], for [Total Sulfate] ≥ 20 ug/m3

[Small Sulfate] = [Total Sulfate] - [Large Sulfate]

The same equations are used to apportion total nitrate and total organic mass concentrations into the small and large size fractions.

Sea salt is calculated as 1.8 x [Chloride] or 1.8 x [Chlorine] if the chloride measurement is below detection limits, missing or invalid. The algorithm also uses three new water growth adjustment terms. They are for use with the small size distribution and the large size distribution sulfate and nitrate compounds and for sea salt ($f_S(RH)$, $f_L(RH)$ and $f_{SS}(RH)$ respectively).

Testing of the new algorithm reveals that, in most regions of the country, the new equation gives somewhat higher light extinction, especially on the worst visibility days. But compared to the old equation, the relative percentage of extinction attributed to the various PM components is nearly unchanged. Therefore, the revised equation is not expected to make a large difference in the rate of visibility changes resulting from potential control strategy simulations. However, the new equation could lead to differences in the relative change in light extinction predicted by the models. Either of the IMPROVE equations are acceptable for use in reasonable progress analyses, although States should provide documentation concerning the choice of equation, especially in situations where large differences are noted between the new and old[71].

---

[71]Both algorithms can be independently used to estimate changes in extinction between a base period and a future year. But the same algorithm needs to be used in each calculation (don't use the old algorithm to calculate base extinction and the new algorithm to calculate future extinction or vice versa). Also, the same algorithm should be used to calculate the

Deciview Haze Index

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

$$\textbf{Deciview} = 10 \ln(b_{ext}/10) \qquad (6.3)$$

Where the units of b_{ext} and light scattering due to Rayleigh scattering[72] (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 track progress is given by Equation (6.4). A 1 deciview change is equivalent to ~10% change in b_{ext}.

$$\Delta dv = dv_{future} - dv_{baseline} \qquad (6.4)$$

A negative number indicates a reduction in deciviews, which is an improvement in visibility.

Estimating Mass Associated With Components Of Particulate Matter.

Regional haze calculations for the modeled uniform rate of progress analysis utilizes IMPROVE measurements at Class I areas. All Class I areas have on-site speciated $PM_{2.5}$ measurements or have been assigned a representative IMPROVE monitoring site with measurements[73]. Therefore, it is generally not necessary to interpolate measured $PM_{2.5}$ species data to a site. The existing IMPROVE database of $PM_{2.5}$ measurements should be adequate to provide data for all Class I areas[74].

FRM data is not used in the regional haze analysis. Therefore, it is not necessary for the sum of species components to equal gravimetric measurements obtained with a Federal

---

glidepath (using ambient data) and the modeled visibility changes.

[72]Even though the "new" IMPROVE equation uses a site specific Rayleigh Scattering value, the denominator in the deciview equation should always be 10. In this way, the deciview calculation will be consistent across all Class I areas. Under very clean conditions, the deciview value can be negative, but this should not be considered a problem.

[73]See U.S. EPA 2003b, Appendix A, Table A-2.

[74]There may be a few IMPROVE sites that don't have enough complete data to provide baseline condition information for the 2000-2004 period. It may be necessary to substitute data from other monitoring networks or to interpolate data to the site(s).

Reference or equivalent method for measuring $PM_{2.5}$. Therefore, for regional haze calculations, it is not necessary to use the speciated data adjustment procedures introduced in Section 3 (the "SANDWICH" adjustments).

Sisler (1996) has developed a set of default assumptions[75] for mass associated with each of the components of particulate matter for Equation (6.1). These are presented in Table 6.1. The components in Table 6.1 are similar to those used in the modeled attainment demonstrations for the $PM_{2.5}$ NAAQS. Notice however, that sulfate and nitrate mass are assumed to be fully neutralized (therefore, ammonium is not needed as a separate component); organic carbon is assumed to be equal to 1.4 times **measured** OC mass (1.8 times OC in the new IMPROVE equation); there is no water component; and there is a term for coarse particulate matter.

Table 6.1. Default Assumptions Used To Derive Aerosol Species Concentrations For Estimating Extinction Coefficients

(1) Species	(2) Formula	(3) Assumptions
Sulfate	**4.125 (measured sulfur) or 1.375 (measured SO4)**	**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) or 1.8* OC for the new IMPROVE equation**	**Average organic molecule is 70% carbon**
Fine 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**

[75]The default assumptions in table 6.1 are applicable to the "old" IMPROVE equation. The new IMPROVE equation assumes that OCM= 1.8 x [OC] and sea salt = 1.8 x [chloride].

Normalizing trends in regional haze- f(rh) factors. It is clear from equations 6.1 and 6.2 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. 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 reduce 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 (6.1), 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 Equations 6.1 and/or 6.2 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.

The need to use climatological f(rh) data for tracking progress of measured data is obvious. The measured relative humidity on the 20% worst days in a baseline period will be different than the measured relative humidity in a future period. Therefore the only way to normalize for relative humidity is to use a single climatological value. In the context of modeling, there is a choice. We can use climatological data or we can use the modeled (or measured) relative humidity data from the air quality model. Since the meteorological data is held constant when predicting future year air quality, the modeled relative humidity values would also be constant. This would be one way to normalize the data.

Throughout this section we make every attempt to be consistent with the "Guidance for Tracking Progress Under the Regional Haze Program"(hereafter referred to as the "Tracking Guidance") (U.S. EPA, 2003b). It will be easier to interpret the modeling results and compare them to future measured values, if the data are calculated in the same manner. We therefore recommend using climatological f(rh) values to calculate baseline and future visibility. Appendix A (Table A-1) from the Tracking Guidance displays the relationship between relative

humidity and f(rh) values. The f(rh) values were calculated from data reported by Tang (1996). The f(rh) values are 1.00 up to 36% relative humidity and grow to a maximum value of 7.40 at a relative humidity of 95%[76]. The Tracking Guidance contains monthly climatological average f(rh) values for each Class I area. These values were calculated using 10 years of meteorological data.

For calculations with the new IMPROVE equation, three separate f(rh) curves are needed. These can be found in (Hand, 2006). The monthly f(rh) values for both the old and new IMPROVE equations have been incorporated into precalculated daily extinction and deciview data that can be found on the Visibility Information Exchange Web System (VIEWS) website (http://vista.cira.colostate.edu/views/). The use of precalculated f(rh) and visibility data will make it more simple for States and RPOs to calculate visibility changes and will also provide for more consistency.

6.4 How Do I Apply A Model To Calculate Changes in Visibility ?

The purpose of a modeling assessment of uniform rate of progress is to determine if a proposed control strategy will result in uniform rate of progress being met when measured concentrations of particulate matter are used to estimate visibility impairment at some future date. The analysis we recommend has 6 steps.

1. For each Class I area, rank visibility (in deciviews) 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 deciviews for the 20% of days with worst and 20% of days with best visibility. For each Class I area, calculate the 5 year mean deciviews 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 response factors for each component of particulate matter identified on the right hand side of Equations (6.1) and/or (6.2).

4. Multiply the relative response factors times the measured species concentration data during the base period (for the measured 20% best and worst days). This results in daily future year species concentrations data.

5. Using the results in Step 4 and the IMPROVE algorithm (equation 6.1 or 6.2), calculate the future daily extinction coefficients for the 20% best and 20% worst visibility days in each of the five base years.

---

[76]For the regional haze calculations, the f(rh) factors have been capped at the f(rh) value associated with 95% relative humidity. Relative humidity value above 95% use the same value.

6. Calculate daily deciview values (from total daily extinction) and then compute the future average mean deciviews for the worst and best days for each year. Then average the 5 years together to get the final future mean deciview value for the worst and best days.

We describe each of these steps more fully below. The methodology follows the same basic procedures outlined in the "Tracking guidance." Further details (for steps 1 and 2) can be found in that document. We conclude this subsection with an example illustrating the recommended modeled uniform rate of progress analysis.

Step 1. Using monitored data, rank baseline 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[77]. The deciview (dv) should serve as the basis for ranking. Day-specific observations for mass associated with SO_4, NO_3, OC, EC, soil, and CM, as defined in Table 6.1, should be used to calculate b_{ext} for each day. The appropriate month- and area-specific climatological relative humidity adjustment factor(s) (*f(rh)*) should be used. Total b_{ext} for all components should be converted to deciviews for each day to get a daily deciview value.

Step 2. Calculate the average baseline deciviews 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 deciview value) to best (lowest deciview value) visibility. For each year, note the 20% of days with worst and the 20% of days with best visibility. Calculate the arithmetic mean deciview value for the identified 20% worst- and best visibility days in each year. Average the resulting 5 yearly mean deciview values reflecting worst visibility. This represents the value subject to improvement (i.e., reduction) to meet the glidepath for regional haze. Average the 5 yearly mean deciview values reflecting mean visibility on the days with best visibility.

Step 3. Estimate relative response 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[78]. Air quality corresponding with future

---

[77]Pre-calculated and ranked extinction and deciview calculations for all Class I areas for the base period 2000-2004 (using both the "old" and "new" IMPROVE algorithms) are available on the VIEWS website at http://vista.cira.colostate.edu/views/. This data can be used to satisfy steps 1 and 2.

[78]How many and which days to simulate is discussed in Section 6.3 and in Section 14.1.

emissions (reflecting effects of growth and controls) should be simulated for the same days. Take the (temporal) arithmetic mean concentration for each $PM_{2.5}$ component (and coarse mass) 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. A separate set of RRF values are calculated for the "worst" and "best" visibility days identified in step 2. The RRFs are calculated using the identified 20% best and 20% worst **monitored days** at each Class I area. This will likely be a different set of days at each monitor.

Step 4. Using the RRFs and ambient data, calculate future year daily concentration data for the best and worst days.

Multiply the relative response factors derived in Step 3 times measured daily concentration data for each component of $PM_{2.5}$ and CM to get future daily estimates of species concentrations for $PM_{2.5}$ components and CM on "worst visibility" and "best visibility"days. These multiplications produce future concentration estimates for SO_4, NO_3, OC, EC, Soil and CM for each of the previously selected "worst" and "best" visibility days. This calculation is performed for each best and worst day for the five year period using an RRF for each PM component (a separate set of RRFs for the best days and the worst days)
.

Step 5. Use the information developed in Step 4 to compute future year daily bext values for the best and worst days.

Use the future year concentration data calculated in step 4 to calculate future year daily b_{ext} values for each PM component for each of the best and worst days for the five year period. This is accomplished by applying either the old or new IMPROVE visibility algorithm (equations 6.1 and 6.2).

Step 6. Use the daily total bext values from step 5 to calculate future mean deciview values for the best and worst days.

The total daily b_{ext} for each day is converted to deciviews. This gives a future year daily deciview value for each of the best and worst days.

Next, compute the arithmetic mean future deciview value for the "worst" and "best" visibility days for each year. This leads to 5 future estimated mean deciview values for the "worst" and 5 future estimated mean deciview values for the "best" visibility days. Compute the arithmetic mean of the 5 mean values for deciviews on the "worst" days, and the arithmetic mean of the 5 mean deciview values estimated for the "best" visibility days.

The resulting 5 year average mean values for deciviews on "worst" and "best" visibility days can be compared to the baseline mean deciview values calculated in step 2. If the resulting change in deciviews is a negative number (future - base), this represents an improvement in visibility. Compare the change in visibility to the previously calculated uniform rate of progress

value for each Class I area.

Example 6.2

We use example 6.2 to illustrate the modeled uniform rate of progress assessment. 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 deciviews for the other four base years to illustrate subsequent steps in the test. The example uses the old IMPROVE algorithm. The procedure is the same for the new IMPROVE algorithm.

Given:

Ten days have measured components of PM in a Class I area during the first year of a 5-year base period. Table 6.2 below shows the measurements (in $\mu g/m^3$) for each of the 10 days. Table 6.2 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.

Table 6.2

Day	Date	*f(rh)*	SO_4 ($\mu g/m^3$)	NO_3 ($\mu g/m^3$)	OC ($\mu g/m^3$)	EC ($\mu g/m^3$)	Soil ($\mu g/m^3$)	CM ($\mu g/m^3$)
1	**2/15**	1.7	4.53	2.23	3.37	0.89	0.32	7.33
2	**3/15**	2.9	5.12	1.78	2.92	0.78	0.44	9.17
3	**4/15**	3.5	5.67	1.25	2.44	0.54	0.43	10.25
4	**5/15**	3.7	6.59	1.12	2.86	0.50	0.66	9.80
5	**6/15**	4.3	7.47	0.93	3.49	0.61	0.98	10.99
6	**7/15**	4.6	8.33	0.79	3.50	0.69	1.08	11.38
7	**8/15**	4.4	9.22	0.89	3.24	0.67	0.89	10.78
8	**9/15**	4.1	8.05	0.97	3.12	0.73	0.71	8.25
9	**9/30**	4.1	6.84	1.15	3.21	0.85	0.42	8.82
10	**10/30**	2.7	5.32	1.41	3.26	0.76	0.64	8.11

The following uniform rate of progress has been calculated: (a) mean visibility on the 20% of days with worst visibility 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 uniform rate of progress in this Class I area?

Solution:

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

First, estimate the extinction coefficient for each day with the needed PM measurements. This is done using the information in table 6.2 with Equation (6.1). For day 1 in year 1, the current extinction coefficient is:

b_{ext} = (3)(1.7)[4.53] + (3)(1.7)[2.23] + (4)(1)[3.37] + (10)[0.89] + (1)[0.32] + (0.6)[7.33] + 10

$b_{ext} = 71.6\ Mm^{-1}$

Then convert extinction into deciviews:

dv = 10 * ln(71.6/10) = 19.7 dv

Current extinction coefficients and deciviews 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 deciviews (i.e., worst visibility) is given a rank of "1". The results of these calculations are displayed in the table 6.3 below. 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

Table 6.3

Day	Date	*f(rh)*	SO_4 (Mm^{-1})	NO_3 (Mm^{-1})	OC (Mm^{-1})	EC (Mm^{-1})	Soil (Mm^{-1})	CM (Mm^{-1})	b_{ext} baseline (Mm^{-1})	Deci views	Rank
1	2/15	1.7	23.1	11.4	13.5	8.9	0.3	4.4	*71.6*	*19.7*	*10*
2	3/15	2.9	44.5	15.5	11.7	7.8	0.4	5.5	*95.5*	*22.6*	*8*
3	4/15	3.5	59.5	13.1	9.8	5.4	0.4	6.2	*104.4*	*23.5*	*7*
4	5/15	3.7	73.1	12.4	11.4	5.0	0.7	5.9	*118.6*	*24.7*	*6*
5	6/15	4.3	96.4	12.0	14.0	6.1	1.0	6.6	*146.0*	*26.8*	*4*
6	7/15	4.6	115.0	10.9	14.0	6.9	1.1	6.8	*164.7*	*28.0*	*2*
7	8/15	4.4	121.7	11.7	13.0	6.7	0.9	6.5	*170.5*	*28.4*	*1*
8	9/15	4.1	99.0	11.9	12.5	7.3	0.7	5.0	*146.4*	*26.8*	*3*
9	9/30	4.1	84.1	14.1	12.8	8.5	0.4	5.3	*135.3*	*26.1*	*5*
10	10/30	2.7	43.1	11.4	13.0	7.6	0.6	4.9	*90.7*	*22.0*	*9*

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

For year 1, mean worst visibility = (28.0 + 28.4) / 2 = 28.2 dv, and

mean best visibility = (19.7 + 22.0) / 2 = 20.9 dv

Mean worst and best visibility for years 2-5 is provided in table 6.4. Table 6.4 below summarizes mean worst and best visibility for each of the 5 years in the base period.

Table 6.4

Year	Mean $dv_{current}$, Worst Visibility Days	Mean $dv_{current}$, Best Visibility Days
1	**28.2**	**20.9**
2	**29.1**	**20.4**
3	**30.2**	**18.7**
4	**27.8**	**19.1**
5	**28.5**	**19.5**

The "average mean baseline worst and best visibility" 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

$dv_{baseline}$ = (28.2 + 29.1 + 30.2 + 27.8 + 28.5) / 5 = **28.8 dv**

The average mean best visibility is

$dv_{baseline}$ = (20.9 + 20.4 + 18.7 + 19.1 + 19.5) / 5 = **19.7 dv**

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

Tables 6.5 and 6.6 show the procedure for calculating component-specific relative response factors using an air quality model. Component-specific relative response factors are computed as described in Section 4.3. The example shows the calculation of RRFs for the worst days in the year modeled. The same calculation is repeated for the best days.

Table 6.5 Base year modeled species concentrations on worst days (in ug/m3)

Modeled Output	SO_4	NO_3	OC	EC	Soil	CM
Day 1	2.12	6.54	3.56	0.87	1.23	5.43
Day 2	8.33	2.11	4.67	1.23	0.34	7.32
Day 3	9.33	1.23	6.32	1.45	0.87	8.21
Day 4	6.43	1.67	4.56	0.54	1.07	4.67
Day 5	10.53	0.57	4.12	0.69	1.54	10.32
Mean Base Concentration	**7.35**	**2.42**	**4.65**	**0.96**	**1.01**	**7.19**

Table 6.6 Future year modeled species concentrations on worst days (in ug/m3)

Modeled Output	SO_4	NO_3	OC	EC	Soil	CM
Day 1	2.01	4.23	3.45	0.77	1.21	5.75
Day 2	7.56	1.89	4.55	1.01	0.32	7.54
Day 3	6.54	1.11	5.99	1.09	1.02	8.71
Day 4	5.32	1.45	4.23	0.35	1.1	5.02
Day 5	7.23	0.43	3.99	0.45	1.51	10.47
Mean Future Concentration	**5.73**	**1.82**	**4.44**	**0.73**	**1.03**	**7.50**

Worst days RRF	SO_4	NO_3	OC	EC	Soil	CM
RRF (mean future/mean base)	**0.780**	**0.752**	**0.956**	**0.768**	**1.022**	**1.043**

Step 4. Multiply the relative response factors times the measured daily species concentration data during the base period to compute future daily species concentrations.

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 as well as the best days RRFs given in table 6.7 below:

Table 6.7

Best days RRF	SO_4	NO_3	OC	EC	Soil	CM
RRF (mean future/mean base)	**0.870**	**0.823**	**0.976**	**0.784**	**1.112**	**1.058**

Worst Days

Day 6: $[SO_4]_{future} = (RRF)_{SO4}\ [SO_4]_{baseline} = (0.780)\ [8.33] = 6.50\ \mu g/m^3$
$[NO_3]_{future} = (0.752)\ [0.79] = 0.59\ \mu g/m^3$
$[OC]_{future} = (0.956)\ [3.50] = 3.35\ \mu g/m^3$
$[EC]_{future} = (0.768)\ [0.69] = 0.53\ \mu g/m^3$
$(Soil)_{future} = (1.022)\ [1.08] = 1.10\ \mu g/m^3$
$[CM]_{future} = (1.043)\ [11.38] = 11.87\ \mu g/m^3$

Day 7: $[SO_4]_{future} = (0.780)\ [9.22] = 7.19\ \mu g/m^3$
$[NO_3]_{future} = (0.752)\ [0.89] = 0.67\ \mu g/m^3$
$[OC]_{future} = (0.956)\ [3.24] = 3.10\ \mu g/m^3$
$[EC]_{future} = (0.768)\ [0.67] = 0.51\ \mu g/m^3$
$(Soil)_{future} = (1.022)\ [0.89] = 0.91\ \mu g/m^3$
$[CM]_{future} = (1.043)\ [10.78] = 11.24\ \mu g/m^3$

Best Days

Day 1: $[SO_4]_{future} = (0.870)\ [4.53] = 3.94\ \mu g/m^3$
$[NO_3]_{future} = (0.823)\ [2.23] = 1.84\ \mu g/m^3$
$[OC]_{future} = (0.976)\ [3.37] = 3.29\ \mu g/m^3$
$[EC]_{future} = (0.784)\ [0.89] = 0.70\ \mu g/m^3$
$(Soil)_{future} = (1.112)\ [0.32] = 0.36\ \mu g/m^3$
$[CM]_{future} = (1.058)\ [7.33] = 7.76\ \mu g/m^3$

Day 10: $[SO_4]_{future} = (0.870)\ [5.32] = 4.63\ \mu g/m^3$
$[NO_3]_{future} = (0.823)\ [1.41] = 1.16\ \mu g/m^3$
$[OC]_{future} = (0.976)\ [3.26] = 3.18\ \mu g/m^3$
$[EC]_{future} = (0.784)\ [0.76] = 0.60\ \mu g/m^3$
$(Soil)_{future} = (1.112)\ [0.64] = 0.71\ \mu g/m^3$
$[CM]_{future} = (1.058)\ [8.11] = 8.58\ \mu g/m^3$

Similar calculations (using the same model derived component specific RRFs) are performed for each of the "worst" and "best" days in each of the other 4 years in the base period[79].

Step 5. Using the results in Step 4, calculate the future year extinction coefficients for the 20% best and 20% worst visibility days in each of the five base years.

Using future PM components obtained in Step 4, we can estimate future daily total b_{ext}.

For year 1

Worst Days

Day 6: b_{ext} = (3)(4.6)[6.50] + (3)(4.6)[0.59] + (4)[3.35] + (10)[0.53] + 1.10 + (0.6)[11.87] + 10 = 134.8 Mm^{-1}

Day 7: b_{ext} = (3)(4.4)[7.19] + (3)(4.4)[0.67] + (4)[3.10] + (10)[0.51] + 0.91 + (0.6)[11.24] + 10 = 139.0 Mm^{-1}

Best Days

Day 1: b_{ext} = (3)(1.7)[3.94] + (3)(1.7)[1.84] + (4)[3.29] + (10)[0.70] + 0.36 + (0.6)[7.76] + 10 = 64.6 Mm^{-1}

Day 10: b_{ext} = (3)(2.7)[4.63] + (3)(2.7)[1.16] + (4)[3.18] + (10)[0.60] + 0.71 + (0.6)[8.58] + 10 = 81.4 Mm^{-1}

Step 6. Use the daily total bext values from step 5 to calculate future mean deciview values for the best and worst days.

Calculate daily deciview values (from total daily extinction) and then compute the future average mean deciviews for the worst and best days for each year. Then average the 5 years together to get the final future mean deciview value for the worst and best days.

For Year 1

Worst days

Day 6: 134.8 Mm^{-1} = 10* ln(134.8/10) = 26.0 dv

Day 7: 139.0 Mm^{-1} = 10* ln(139.0/10) = 26.3 dv

---

[79]The same mean RRFs (for each component) are applied to the concentrations on all of the worst days for each of the 5 years of data. The mean RRFs are derived from the modeled days. A separate set of RRFs are derived (in the same manner) for the best days.

Future mean visibility on worst days = (26.0 + 26.3) / 2 = 26.2 dv

Best Days

Day 1: 64.6 Mm^{-1} = 10* ln(64.6/10) = 18.7 dv

Day 10: 81.4 Mm^{-1} = 10* ln(81.4/10) = 21.0 dv

Future mean visibility on best days = (18.7 + 21.0) / 2 = 19.8 dv

Similar calculations are performed for previously selected "worst" and "best" days in each of years 2-5. To illustrate the uniform rate of progress assessment, assume these other calculations yield the following estimates for future mean dv on worst and best visibility days.

Table 6.7

Year	Future Mean dv On Worst Visibility Days	Future Mean dv On Best Visibility Days
1	**26.2**	**19.8**
2	**27.3**	**19.3**
3	**27.6**	**17.6**
4	**25.9**	**18.3**
5	**26.1**	**18.6**

Using results in table 6.7, we see that the estimated future average mean visibility for the 20% days with worst visibility is

$$\mathbf{dv_{future}} = (26.2 + 27.3 + 27.6 + 25.9 + 26.1) / 5 = \mathbf{26.6}$$

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

$$\mathbf{dv_{future}} = (19.8 + 19.3 + 17.6 + 18.3 + 18.6) / 5 = \mathbf{18.7}$$

The results generated in step 6 are then used to 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 difference = 26.6 - 28.8 = **- 2.2 deciviews**

For the 20% days with best visibility,

dv difference = 18.7 - 19.7 = **- 1.0 deciviews**

The future year strategy leads to an improvement in visibility of 2.2 deciviews. Thus, the visibility improvement for the 20% days with the worst visibility is predicted to exceed the uniform rate of progress (-2.2 dv vs. -2.0 dv). Also visibility is not predicted to deteriorate on the best days but, in fact, is estimated to improve (by 1.0 deciviews)

This information is used in the process of determining reasonable progress. The remaining steps in the process are beyond the modeling guidance. The modeling can only be used to determine the predicted improvement in visibility (and whether the improvement is on, above, or below the glidepath). It cannot determine the reasonable progress goals and it cannot determine whether reasonable progress is met.

6.5 How Do I Select Appropriate Inputs For The Uniform Rate of Progress Analysis?

In Section 6.4, we described the recommended modeled uniform rate of progress analysis. An important part of the analysis requires using component-specific relative response factors (RRF's), obtained with models, to estimate future concentrations of these components and, subsequently, future visibility. In this subsection, we address more details concerning the calculation of RRFs. Second, there are several assumptions inherent in the recommended modeled uniform rate of progress analysis. We identify these in this subsection and comment on their underlying rationale. More specifically, we address eight 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 response factors;
- how many and what kind of days to use to develop RRF values
- use of relative response factors which are specific for days with poor vs. days with good visibility
-alternative RRF calculations

Estimating baseline "worst" and "best" visibility in a Class I area without monitors. There are 156 Class I areas and there are 110 IMPROVE sites in or near Class I areas. Therefore, approximately 46 Class I areas do not have co-located IMPROVE monitors. EPA's "Tracking Guidance" recommends using IMPROVE data from a nearby site to represent the visibility at each Class I area that does not have ambient data. Table A-2 in Appendix A (of the

tracking guidance) lists the recommended monthly f(rh) values[80] for each Class I area as well as the representative site for each Class area. The representative IMPROVE site data will be used to track regional haze progress for the Class I areas. Therefore, it follows that visibility improvement should be predicted at the monitor, not at the actual Class I area. For the purposes of deriving ambient data for modeling, we recommend following the same representative site assignments contained in the tracking guidance. In this way, the 20% worst and best days can be derived for each Class I area from the network of 110 IMPROVE sites[81]. Similarly, the modeling results should be extracted for the location of the representative monitor, not the actual location of the Class I area[82].

Considering a base year with little or no monitored particulate matter or missing data. The Tracking Guidance recommends calculating baseline visibility values for sites with at least 3 out of 5 complete years of data. It further contains recommendations for determining if a year has complete data. In general, a site should have 50% data completeness in all quarters and meet a 75% completeness criteria for the full year. There should be no periods with more than 30 consecutive days without data. The guidance assumes that all IMPROVE sites will have at least 3 complete years of data in the base 2000-2004 period. The guidance also contains procedures for filling in missing data as part of the calculation of the 20% best and worst days. We recommend that these same procedures are followed for calculations in the modeled test.

There are several data completeness issues that may cause problems within the modeling analysis. First, a site may have less than 3 years of complete data. This will be a problem for the purpose of modeling and for tracking progress. States should work with their EPA Regional Office and Federal Land Managers to determine how to estimate baseline visibility for these area(s).

Another issue that is a more specific problem for modeling occurs when data is missing during the meteorological time period that was modeled. It is likely that most States will only be modeling a single year (or less). Therefore it is possible that some or all of the ambient data at one or more Class I areas is missing during that year. Without ambient data, it is impossible to identify the 20% best and worst days (used to calculate modeled RRFs).

---

[80]The f(rh) values in the Tracking Guidance were developed for the “old” IMPROVE algorithm and should not be used to calculate visibility using the “new” IMPROVE algorithm.

[81]Bering Sea Wilderness (Alaska) is the only Class I area that has no IMPROVE monitor and no representative IMPROVE site. On-site IMPROVE or representative IMPROVE data can be found for the other 155 sites.

[82]It may be informative to compare model response at the representative IMPROVE site versus the actual location of the Class I area. Large differences in the model response may indicate that the “representative” site may not adequately represent visibility at the Class I area.

Again, if this occurs, States should work with their Regional Office and FLM to determine the best way to calculate visibility on the best and worst days. Potential options are to use data from another nearby IMPROVE site, use nearby data from a different ambient data network, or interpolate ambient data to the site. Another alternative is to estimate the 20% best and worst days from the model outputs[83].

Using a constant sample of days to estimate baseline and future visibility. For a typical Class I area, there will be about 120 days per year having measurements needed to estimate $(dv)_{baseline}$ with Equation (6.1). 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 6.4 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 controls 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.

Because the uniform rate of progress analysis relies on mean visibility values 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 analysis. Further, because of the shape of the distribution of extinction coefficients, the mean of 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

---

[83]This is only recommended if the model performance for all species is good (at sites with data). While it is expected that the air quality models will do an adequate job of predicting the $PM_{2.5}$ components which are responsible for the largest visibility degradation, the models may not perform as well in estimating total visibility impairment. Accurately estimating the 20% best and worst days depends on accurately estimating all of the PM components. In particular, overprediction of certain components (e.g nitrates in the winter) could lead to identification of worst or best days in the wrong season and/or for the wrong reasons (e.g. wrong mix of species on the best or worst days).

days are likely to be relatively low, differences in the computed mean visibility for "best" days are likely to be small. Further, any resulting difference in the progress analysis 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 analysis (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 response 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 in section 3.3 for a grid whose cells are 12 km on a side. Similar to the $PM_{2.5}$ NAAQS attainment test, nearby grid cells should be averaged. 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.

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 14, this may make it advisable to simulate (at least) a full year so that relative response 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 "worst" days in that year. The same procedure is followed to derive RRFs over the ~24 "best" days in the 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 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 14.1.4 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 recommend calculation of 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 $\geq$ ~ 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.

Alternative RRF calculations. The default glidepath analysis is relatively simple in that a single mean RRF is calculated for each PM component (separate RRFs on worst and best days). A series of tests with more complicated methods has shown that 1) the difference between various versions of the test are usually small and 2) each of the alternative tests has limitations in its applicability (Environ, 2005). Possible variations include the use of day specific RRFs, season (or quarter) specific RRFs, or climatological based RRFs. In some cases, these more complicated techniques may provide different answers, but sometimes not. There are specific limitations noted with each of these alternatives. We have chosen to keep the single mean RRF test as the default recommendation. States are encouraged to explore other methods for estimating RRFs if it is felt that the default recommendation is too simplistic to accurately capture the change in future visibility at any particular Class I area. The SIP demonstration should use the most appropriate method of projecting future concentrations for the characteristics of each Class I area[84]. Alternative methods should be discussed in the modeling protocol and discussed with the appropriate EPA Regional Office and FLMs.

[84]In particular, issues may arise when dealing with significant visibility contributions from fires, coarse mass and fine soil (mostly wind-blown dust), and international transport (and possibly other issues). Each of these issues should be addressed in the modeling protocol and solutions should be discussed with the appropriate EPA Regional Office(s) on a case by case basis.

7.0 How Can Additional Analyses Can Be Used to Support the Attainment Demonstration?

By definition, models are simplistic approximations of complex phenomena. The modeling analyses used to assess whether various emission reduction measures will bring an individual area into attainment for the NAAQS contain many elements that are uncertain (e.g., emission projections, meteorological inputs, model response). These uncertain aspects of the analyses can sometimes prevent definitive assessments of future attainment status. The confidence in the accuracy of the quantitative results from a modeled attainment test should be a function of the degree to which the uncertainties in the analysis were minimized. In general, by following the recommendations contained within this guidance document, EPA expects that the attainment demonstrations will mitigate the uncertainty as much as is possible given the current state of modeling inputs, procedures, and science. However, while Eulerian air quality models represent the best tools for integrating emissions and meteorological information with atmospheric chemistry and no single additional analysis can match the expected reliability of these models' results, EPA believes that all attainment demonstrations will be strengthened by additional analyses that can supplement the modeling analysis to enhance the assessment of whether the planned emissions reductions will result in attainment.

Supplemental evidence should accompany all model attainment demonstrations. Generally, those modeling analyses that show that attainment will be reached in the future with some margin of safety (e.g., estimated concentrations below 82 ppb for ozone, 14.5 ug/m3 for annual $PM_{2.5}$, and 62 ug/m3 for 24-hour $PM_{2.5}$) will need more limited supporting material. For other attainment cases, in which the projected future design value is closer to the NAAQS, more rigorous supporting analyses should be completed. As noted earlier (see section 2.2), there may be some areas that can expect to achieve timely attainment despite failing the model attainment test, and vice versa. This section of the guidance will discuss some specific additional analyses that can be used to supplement the model projections. Of particular interest are analyses that help determine whether the model likely overpredicts, underpredicts, or accurately predicts the **air quality improvement** projected to occur by the attainment date. States should review these supplemental analyses, in combination with the modeling analysis, in a "weight of evidence" assessment of whether each area is likely to achieve timely attainment. Additional examples of possible weight of evidence determinations are provided in existing EPA guidance (U.S. EPA, 1996a). Again, it should be noted that no single supplemental analysis can serve as an adequate substitute for the air quality model; however, in aggregate, supplemental analyses may provide information which may indicate a different outcome than the modeled test.

In Section 7.1, we identify several broad types of analyses which can be used to corroborate one another as part of a set of supplemental analyses. 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 7.2, we more specifically address the submittal of a weight of evidence demonstration and recommend a framework for compiling and submitting a weight of evidence demonstration.

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, supplemental analyses differ in some respects for visibility-related applications. Section 7.3 describes how to use supplemental analyses in regional haze analyses.

7.1 What Types of Additional Analyses Should Be Completed as Part of the Attainment Demonstration?

There are three basic types of analyses that are recommended to supplement the primary modeling analysis. They are:

1) Additional modeling

2) Analyses of trends in ambient air quality and emissions

3) Observational models and diagnostic analyses

These broad groups of supplemental analyses are discussed in more detail below:

Modeling: The relative attainment tests described in sections 3, 4, and 5 are the primary modeling tools used in an attainment demonstration. The application of a photochemical grid model, developed on a regional or local scale, is the best tool available to judge the impacts of changes in future year emissions. In addition to this "primary" modeling analysis, there are various other models, applications, and tools that can be used to supplement the results of the modeled attainment test. These include, but are not limited to:

- Use of available regional or national scale modeling applications that are suitable[85] for the local area,

- Use of other appropriate local modeling attainment demonstrations that include the nonattainment area of interest,

- Use of photochemical source apportionment and/or process analysis modeling tools to help explain why attainment is (or is not) demonstrated,

[85] The resolution, emissions, meteorology, and other model inputs should be evaluated for applicability to the local nonattainment area. Additionally, model performance of the regional modeling for the local nonattainment area should be examined before determining whether the regional model results are suitable for use in the local attainment demonstration.

- Use of multiple air quality models / model input data sets (e.g., multiple meteorological data sets, alternative chemical mechanisms or emissions inventories, etc.). For results to be most relevant to the way we recommend models be applied in attainment demonstrations, it is preferable that such procedures focus on the sensitivity of estimated relative response factors (RRF) and resulting projected design values to the variations in inputs or model formulations.

- Use of dispersion models to address primary $PM_{2.5}$ contributions to monitors. In areas with large spatial gradients of primary $PM_{2.5}$, dispersion models are best suited to characterizing the change in primary $PM_{2.5}$ in the future. Areas that are relying on local primary PM controls to reach attainment should submit a local area analysis as part of the primary attainment demonstration. In other areas, a local area analysis may be useful as a supplemental analysis.

- Application of the attainment test with alternative procedures compared to the default recommendations in Sections 3, 4, and 5 of this guidance. Any alternate approaches should be accompanied with a technical justification as to why the approach is appropriate for the area in question and should be discussed with the appropriate EPA regional office.

As discussed in Section 2, EPA has determined that the best approach to using models to demonstrate attainment of the NAAQS is to use a model in a relative mode. However, for some model applications there may be strong evidence from the performance evaluation that the model is able to reproduce detailed observed data bases with relatively little error or bias. Particularly for cases such as these, some types of "absolute" modeling results may be used to assess general progress towards attainment from the baseline inventory to the projected future inventory[86]. There are several metrics that can be considered as part of this type of additional analysis:

- Percent change in total amount of ozone or $PM_{2.5}$ >= NAAQS[87] within the nonattainment area
- Percent change in number of grid cells >= NAAQS within the nonattainment area
- Percent change in grid cell-hours (days) >= NAAQS within the nonattainment area
- Percent change in maximum modeled 8-hour ozone within the nonattainment area

---

[86] Care should be taken in interpreting absolute metrics if the model evaluation shows a large underprediction or overprediction of ozone or $PM_{2.5}$ concentrations. An underprediction of observed concentrations will make it artificially easy to show progress towards absolute attainment levels and an overprediction will make it artificially difficult to show progress towards attainment.

[87]For each of these metrics, the appropriate comparison to the level of the NAAQS is 85 ppb for 8-hour ozone; 65 ug/m3 for 24-hour $PM_{2.5}$; and 15 ug/m3 for annual $PM_{2.5}$.

While these metrics can be used to estimate the magnitude, frequency, and relative amount of ozone or $PM_{2.5}$ reductions from any given future emissions scenario, there are no threshold quantities of these metrics that can directly translate to an attainment determination. Generally, a large reduction in the frequency, magnitude, and relative amount of 8-hour ozone nonattainment (i.e., >= 85 ppb) or $PM_{2.5}$ nonattainment (24-hour and/or annual) is consistent with a conclusion that a proposed strategy would meet the NAAQS. In the context of a weight of evidence determination, these metrics could be used to suggest that a particular location may be "stiff" or relatively unresponsive to emissions controls, while the rest of the modeling domain/nonattainment area is projected to experience widespread reductions. If a sound technical argument can be made for why atypically high RRFs at any particular location are not reasonable, then these types of supplemental modeling metrics would suggest that attainment is more likely to be achieved than the modeling analysis alone would indicate.

As discussed in section 3.4, an unmonitored area analysis may provide evidence that the area may not achieve timely attainment, even if modeling suggests that attainment will occur at all monitoring locations. In such cases, assessment of metrics concerning the frequency, magnitude, and relative amount of nonattainment may help supplement the information from the unmonitored area analysis. If application of the unmonitored area test indicates that most (if not all) of the unmonitored areas will be in attainment, then that information would be evidence that future attainment may be likely.

Uncertainty estimates associated with the spatial interpolation technique can also be considered when reviewing and interpreting the results of an unmonitored area analysis. When making a decision on whether attainment is likely to occur, areas with very high uncertainty estimates for interpolated design values should be given less weight than areas with low uncertainty estimates[88].

The overall modeling analyses can also be evaluated to determine how appropriate the modeling systems are for making regulatory decisions. Roth (2005) has proposed an "idealized evaluation framework" for judging the quality of modeling applications. The paper lists a series of twenty questions which can be used to judge the overall model application. These questions provide an objective way to compare the quality, and identify deficiencies in modeling applications. The absence of major deficiencies may provide a strong basis for acceptance of model results.

Trends in Ambient Air Quality and Emissions: Generally, air quality models are regarded as the most appropriate tools for assessing the expected impacts of a change in emissions. However, it may also be possible to extrapolate future trends in ozone or $PM_{2.5}$ based on measured historical trends of air quality and emissions. There are several elements to this analysis that are difficult to quantify. First, in most cases, the ambient data trends are best

[88]This is particularly true for $PM_{2.5}$ in areas with little or no ambient speciation data.

assessed by normalizing to account for year-to-year meteorological variations. Second, one must have an accurate accounting of the year-to-year changes in actual emissions (NOx, VOC, and/or SO2 and NH3) for the given area and any surrounding areas whose emissions may impact local concentrations. Third, one must have a solid conceptual model of how ozone or $PM_{2.5}$ is formed in the local area (e.g., NOx-limited, ammonia limited, transport-influenced, etc.). Assuming all of these prerequisites can be met, then it may be possible to develop a curve that relates past emissions changes to historical and current air quality. Once the relationship between past/present emissions and air quality is established, this "response factor" can be applied to the expected emissions reductions from a particular control strategy.

If available, meteorologically adjusted ozone and $PM_{2.5}$ concentrations can be used to establish air quality trends. There are several techniques that have been used to examine the influence of meteorology on air quality. Among them are (a) statistical modeling (U.S. EPA, 2005c); (b) filtering techniques (Rao and Zurbenko, 1995, Flaum, et al., 1996, Milanchus, et al., 1998, Hogrefe, et al., 2000), (c) using a probability distribution of meteorological severity based on climatological data (Cox and Chu, 1993, 1996), (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). Most of this work has examined the relationship between ozone and meteorology. Only recently have analyses examined the relationship between meteorology and $PM_{2.5}$. Additionally, compared to $PM_{2.5}$, the established relationship between ozone and meteorological variables is stronger (higher r-square values). In the case of $PM_{2.5}$, the relationship between concentration and meteorology is complicated by the fact that $PM_{2.5}$ components experience high concentrations at different times of the year and for different reasons. This makes it more difficult to meteorologically adjust $PM_{2.5}$ concentrations.

If a meteorologically adjusted trend in ozone or $PM_{2.5}$ can be estimated, then the information can be used to establish a link between emissions and air quality trends. This is not always straightforward due to the multitude of emissions precursors that may lead to high ozone and $PM_{2.5}$ concentrations. A careful analysis of (meteorologically adjusted) air quality trends and emissions trends of each of the ozone and PM precursors (as well as primary PM) is needed to fully establish relationships. Detailed emissions information as well as a solid understanding of the conceptual model of ozone or $PM_{2.5}$ formation is needed. If a trend can be established based on past emissions changes and air quality changes, then future year predicted emissions levels can be used to extrapolate future air quality.

A simpler (and more uncertain) way to qualitatively assess progress toward attainment is to examine recently observed air quality and emissions trends. Downward trends in observed air quality and in emissions (past and projected) are consistent with progress towards attainment. Strength of the evidence produced by emissions and air quality trends is increased if an extensive monitoring network exists and if there is a good correlation between past emissions reductions and current trends in ozone or $PM_{2.5}$. EPA recently prepared a report that analyzed statistically significant trends in ozone (U.S. EPA, 2004a and U.S. EPA, 2005c) and ozone precursor

emissions as well as a report which examined trends in $PM_{2.5}$ concentrations and precursors (U.S. EPA, 2004b).

Weight given to trend analyses depends on several factors. Analyses that use more air quality data and apply a greater variety of trend parameters provide more credible results. More weight can be placed on the results 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 the extrapolation does not extend very far into the future. Finally, trend analysis is most credible if the contemplated strategy is similar to a past strategy (e.g., both strategies focus on reducing sulfates for PM or NOx for ozone). 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 past emissions.

Observational Models In some cases ambient data can be used to corroborate the effects of a control strategy (e.g., Blanchard et al, 1999; Croes et al, 2003; Koerber and Kenski, 2005). Observational models take advantage of monitored data to draw conclusions about the relative importance of different types of emissions and precursors as factors contributing to observed $PM_{2.5}$ and ozone, as well as inferences which might be drawn about the effectiveness of various strategies to reduce concentrations. 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 general representativeness of the responses simulated with the air quality model for a limited sample of days. However, their ability to estimate *how much* control is needed is limited. Thus, observational approaches are suitable to corroborate results from more quantitative techniques, like air quality models. Additionally, observational models are limited to analyses based on current or past ambient conditions. A change in the relative mix of precursor emissions in the future (compared to current conditions) may lead to a different set of future control strategies compared to what the observational model may indicate based on current data. There are at least two types of observational models: source apportionment (i.e., "receptor") models and indicator species approaches.

Receptor models A large body of literature describes the theory and use of receptor models to identify and/or apportion sources which may be contributing to monitored air quality. Seigneur, (1997), summarized in Seigneur, (1999), provides a review of receptor models, 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, although, to a lesser degree, they can also be used to infer contributions from ozone and secondary PM precursors.

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, 2004d). 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. 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.

The second type of receptor models are multi-variate statistical models. Unlike with CMB, a priori assumptions about contributing source categories are not needed. Another distinction from the CMB approach is that multi-variate models look at day to day 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. Common multi-variate statistical models include positive matrix factorization (PMF) (Paatero and Tapper, 1994), and UNMIX (Maykut, 2003; Poirot et. al, 2001) These models may be useful when confirming whether a strategy is reducing the right sorts of sources.

Indicator Species Indicator species approaches are based on the predicted sensitivity of a secondary pollutant's concentration to changes in different precursors for that pollutant. 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. Measurement of "indicator species" is a potentially useful means for assessing which precursor category (e.g., VOC or NOx) limits further production of ozone or secondary $PM_{2.5}$ at a monitor's location at various times of day and under various sets of meteorological conditions. Several indicator ratios have been developed to examine the sensitivity of ozone to changes in NOx and VOC, and the sensitivity of particulate nitrate to changes in NOX, VOC, and ammonia.

Sillman (1998, 2002) and Blanchard, (1997, 1999, 2000, 2001) identify several sets of indicator species which can be compared to suggest whether ozone is limited by availability of VOC or NOx. Comparisons are done by looking at ratios of these species. States/Tribes should consult the Sillman (1998, 2002) and Blanchard, (1997, 1999, 2000, 2001) references for further details on measurement requirements and interpretation of observed indicator ratios (also see Section 18.5.1 for more details).

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, VOC, or by ammonia emissions using mechanisms which incorporate reactions dealing with secondary particulate matter. If a model accurately predicts observed ratios of indicator

species, then one can conclude with additional confidence that the predicted change in ozone or PM may be accurate.

The 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. Evidence produced by observational models is more compelling if several techniques are used which complement one another and produce results for which plausible physical/chemical explanations 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.

7.2 What Is Entailed In A Weight Of Evidence Determination?

As discussed in Section 2, augmenting a modeled attainment test with supplemental analyses may yield a conclusion differing from that indicated by the modeled attainment test results alone. Past modeling analyses have shown that future design value uncertainties of 2-4 ppb for ozone[89], can result from use of alternate, yet equally appropriate, emissions inputs, chemical mechanisms, and meteorological inputs (Jones, 2005; Sistla, 2004). Because of this uncertainty, EPA believes that weight of evidence determinations can be used in some cases to demonstrate attainment conclusions that differ from the conclusions of the model attainment test.

As part of their recommendations to transform the SIP process into one that is more performance-oriented, the Clean Air Act Advisory Committee (CAAAC) recommended increased use of weight of evidence within State/Local attainment demonstrations (AQMWG, 2005). One of the workgroup's recommendations to EPA was that "EPA, in conjunction with affected stakeholders, should modify its guidance to promote weight-of-evidence (WOE) demonstrations for both planning and implementation efforts. In particular, these demonstrations should reduce reliance on modeling data as the centerpiece for SIP/TIP planning, and should increase use of monitoring data and analyses of monitoring data, especially for tracking progress. Enhanced tracking and ambient monitoring data is a better use of available resources than intensive local modeling."

A weight of evidence (WOE) determination examines results from a diverse set of analyses, including the outcome of the primary attainment test, and attempts to summarize the results into an aggregate conclusion with respect to whether a chosen set of control strategies will result in an area attaining the NAAQS by the appropriate year. The supplemental analyses discussed above are intended to be part of a WOE determination, although the level of detail

---

[89]Few studies have been done to examine similar uncertainties for $PM_{2.5}$. Based on recent modeling analyses, a similar range of +- 2-4% of the NAAQS seems appropriate for $PM_{2.5}$. That translates to roughly 0.3-0.6 ug/m3 for the annual $PM_{2.5}$ standard. Consequently, the recommended weight of evidence range for $PM_{2.5}$ is nominally +- 0.5 ug/m3.

required in a WOE submittal will vary as a function of many elements of the model application (e.g., model performance, degree of residual nonattainment in the modeled attainment test, amount of uncertainty in the model and its inputs, etc.). Each weight of evidence determination will be subject to area-specific conditions and data availability. Area-specific factors may also affect the types of analyses which are feasible for a nonattainment area, as well as the significance of each. Thus, decisions concerning which analyses to perform and how much credence to give each need to be made on a case by case basis by those implementing the modeling/analysis protocol. States/Tribes are encouraged to consult with their EPA Regional office in advance of initiating supplemental analyses to determine which additional analyses may be most appropriate for their particular area.

The most useful supplemental analyses are those providing the best evidence as to how much air quality improvement can be expected as compared to the improvement projected by the air quality modeling analysis. Each analysis is weighed qualitatively, depending on: 1) the capacity of the analysis to address the adequacy of a strategy and 2) the technical credibility of the analysis. If the overall weight of evidence produced by the combination of the primary modeling analysis and the various supplemental analyses supports the attainment hypothesis, then attainment of the NAAQS 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/Tribe believes that the evidence, viewed as a whole, supports a conclusion that the area will, or will not, attain the NAAQS despite a model-predicted DVF concluding otherwise. In conclusion, the basic criteria required for an attainment demonstration based on weight of evidence are as follows:

1) A fully-evaluated, high-quality modeling analysis that projects future values that are close to the NAAQS.

2) Multiple supplemental analyses in each of the three various categories discussed above (modeling, air quality/emissions trends analyses, observational models).

3) A qualitative weighting for each separate analysis based on its ability to quantitatively assess the ability of the proposed control measures to yield attainment.

4) A description of each of the individual supplemental analyses and their results. Analyses that utilize well-established analytical procedures and are grounded with sufficient data should be weighted accordingly higher.

5) Where applicable, a written description as to why the full set of evidence leads to a conclusive determination regarding the future attainment status of the area that differs from the results of the modeled attainment test alone.

7.3 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 when examining visibility trends. Unlike the NAAQS, 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 supplemental analyses that may be used to examine visibility trends. We then identify several refinements to our recommended modeled test for uniform rate of progress analyses 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.

Additional air quality modeling. Sensitivity tests can 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 future emissions and/or differing growth rate assumptions. This may be a particular concern for regional haze analyses because the emissions projection period is generally longer than for most ozone and $PM_{2.5}$ attainment demonstrations. Uncertainty in emissions and growth rates become more important as the projection period is lengthened.

If trends in visibility for "worst" and/or "best" days are similar using sensitivity tests, alternative models and/or alternative modeling approaches, this finding supports conclusions reached in the uniform rate of progress analysis.

Review of trends. A review of trends generally involves a comparison, sometimes qualitative, between past trends in reconstructed visibility and estimated changes in emissions (e.g., early '90's to mid '00's). The IMPROVE dataset has a longer record than the corresponding urban FRM and speciation sites. Some IMPROVE sites have data going back to 1989. 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 PM components are becoming increasingly important sources of light extinction.

Observational models. Receptor 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. Trajectory 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. Grid model based techniques such as tagging,

DDM, or source apportionment may also be useful in identifying areas and emissions sources most responsible for visibility impairment on the worst or best days.

Refinements to the recommended uniform rate of progress analysis. If a strategy for meeting the glidepath appears generally successful, but the glidepath is not met in a limited number of Class I areas, States may consider refining the recommended uniform rate of progress analysis 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 assessment 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.

- Use an alternative light extinction equation such as the revised IMPROVE equation or an area-specific version.

-Available speciated data and other information may be reviewed to see whether the outcome of the test is being influenced by including one or more days with extraordinary events (e.g., a nearby major forest fire lasting a number of days or transported dust events). 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.

-Daily component specific RRFs can be examined to determine if one or more days are responding in a different way compared to the majority of the best or worst days. Determining why days may be more or less responsive to emissions controls may lead to conclusions regarding the suitability of particular days to be represented in the mean response. It may be appropriate, in some cases, to re-calculate mean RRFs with suspect days removed.

- Re-rank 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).

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). Utility 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 18 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 Equations (6.1) or (6.2). If it does, and diagnostic tests (also described in Section 18) are unable to resolve a performance problem, a State may need to address the goal for "best" visibility days in the particular Class I area(s) without using results from a grid model.

8.0 What Additional Analyses and Data Collection Can Be Completed to Assess Progress towards Attainment

The purpose of an attainment demonstration is to provide a best estimate as to whether the control measures included in a State Implementation Plan will result in attainment of the NAAQS by a specific date in the future. In most cases, it will be desirable to periodically track the air quality improvements resulting from the SIP to ensure that the plan is going to result in attainment by the appropriate dates. One possible tracking approach is a mid-course review (MCR). Also, the regional haze rule requires periodic tracking of progress towards visibility goals (every 5 years). In this section, we identify measurements and activities which will provide better support for mid course reviews, future modeling exercises and other supplemental analyses designed to determine the progress toward attainment of the NAAQS or reasonable progress goals. Improved data bases will increase the reliability of reviews and enable identification of reasons for attainment or non-attainment of the NAAQS.

Deploying additional air quality monitors. One type of additional monitoring which should be considered has already been mentioned in Section 3. Additional ozone and/or $PM_{2.5}$ monitors should be deployed in unmonitored locations where future design values are predicted to exceed the NAAQS via the unmonitored area test. This would allow a better future assessment of whether the NAAQS is being met at unmonitored locations. Also, additional PM speciation monitors should be deployed in areas that are expected to continue to exceed the NAAQS in the future (particularly if there is not a speciation monitor at the highest FRM sites in a nonattainment area). Additional ambient data can provide an increased understanding of the nature of the problem, which can help lead to the most cost efficient solution.

Collecting Measurements for Indicator Species Measurement of "indicator species" is a potentially useful means for assessing which precursor category limits further production of ozone or secondary $PM_{2.5}$ at a monitor's location at various times of day and under various sets of meteorological conditions. Many of the gas phase species needed to calculate indicator ratios are not routinely measured. Some of the species needed to calculate ozone indicator ratios are NOy and HNO3. Some of the species needed to calculate the gas ratio and excess ammonia (for nitrate analyses) include NH3 and HNO3. Thus, realizing the potential of the "indicator species method" as a tool for model performance evaluation and for diagnosing why observed concentrations do or do not meet previous expectations may depend on deploying additional monitors and/or measurements.

Making measurements aloft. Almost all measured ambient air quality and meteorological data are collected within 20 meters of the earth's surface. However, the modeling domain extends many kilometers above the surface. Further, during certain times of day (e.g., at night) surface measurements are not always representative of air quality or meteorological conditions aloft. Measurements aloft can be made by aircraft (usually during special studies) or on tall buildings or towers. Concentrations aloft can have marked effects when they are mixed with ground-level emissions during daytime. Thus, the weight given to

modeling results can be increased if good agreement is shown with air quality measurements aloft. The most important of these measurements are ozone, NOy, NO, NO_2, as well as several relatively stable species like CO and selected VOC species. Measurements of SO_2 may also be helpful for identifying presence of plumes from large combustion sources.

Measurements of altitude, temperature, water vapor, winds and pressure are also useful. Continuous wind measurements, made aloft in several locations, are especially important. They provide additional data to "nudge" meteorological model fields, but more importantly also allow for construction of more detailed conceptual models of local ozone formation (Stehr, 2004). For example, measurements of aloft winds and temperatures from lower atmosphere radar profilers can detect low level jets and be used to infer mixing depths. This information can be used to evaluate meteorological and air quality model outputs. This provides greater assurance that the air quality model correctly reflects the configuration of sources contributing to ozone formation.

Special Studies. Over the last 20 years, many States have embarked upon short term special studies to examine both ozone and PM issues. Data collected from special studies can be used to make improvements in the conceptual model of ozone and/or PM formation and to improve the inputs to models. The results of current modeling can be examined to determine the largest sources of uncertainty. This information can be used to design special studies in an effort to collect data that might allow future improvements in emissions, meteorological, and air quality modeling. Examples of special studies include the Central California Ozone Study (CCOS) (http://www.arb.ca.gov/airways/CCOS/CCOS.htm), California Regional Particulate Air Quality Study (CRPAQS) (http://www.arb.ca.gov/airways/crpaqs/publications.htm), Southern California Ozone Study (SCOS) (http://www.arb.ca.gov/research/scos/scos.htm), Texas Air Quality Study (TexAQS) and TexAQS II (http://www.tceq.state.tx.us/implementation/air/airmod/texaqs-files/TexAQS_II.html), Northern Front Range Air Quality Study (NFRAQS) (http://www.nfraqs.colostate.edu/) , and various NARSTO sponsored field studies (http://www.narsto.org/section.src?SID=9).

Extending measurements year round. Many States do not collect ozone and VOC data (and possibly other species) outside of the officially designated ozone season. Because $PM_{2.5}$ is a year round issue (especially for the annual NAAQS), PM precursor data is needed both during and outside of the ozone season to adequately evaluate secondary $PM_{2.5}$ performance and sensitivity to emissions controls. This is particularly relevant in areas with high wintertime nitrate concentrations. Collection of ozone, VOC, and precursor data on a year round basis may help improve performance evaluations and in particular, diagnostic analyses.

Collecting locally applicable speciated emissions data. While the U.S. EPA maintains a library of default VOC and PM emissions species profiles (U.S. EPA, 1993); (U.S. EPA, 2006a), some of these may be dated or may not properly reflect local sources. Use of speciated emissions data is a critical input to air quality models. For example, the accurate representation of the VOC speciation of current and future gasoline emissions may have an important impact on future ozone concentrations. Efforts to improve speciation profiles for local sources (especially

for large sources of primary $PM_{2.5}$) should enhance credibility of the modeling as well as several of the procedures recommended for use in supplemental analyses and the weight of evidence determinations.

9.0 What Documentation Do I Need To Support My Attainment Demonstration?

States/Tribes should follow the guidance on reporting requirements for attainment demonstrations provided in U.S. EPA (1994b). The first seven subjects in Table 9.1 are similar to those in the 1994 guidance. The 1994 guidance envisions an air quality model as the sole means for demonstrating attainment. However, the current guidance (i.e., this document) identifies supplemental analyses as well as a possible weight of evidence determination as a means for corroborating/refuting the modeled attainment test in an attainment demonstration. In addition, feedback received since the earlier guidance has emphasized the need for technical review of procedures used to identify a sufficient control strategy. Thus, we have added two additional subject areas which should be included in the documentation accompanying an attainment demonstration. These are a description of the supplemental analyses and/or weight of evidence determination, and identification of reviews to which analyses used in the attainment demonstration have been subject. In the end, the documentation submitted by the States/Tribes as part of their attainment demonstration should contain a summary section which addresses the issues shown in Table 9.1.

Table 9.1 Recommended Documentation for Modeled Attainment Demonstrations and Regional Haze Analyses

Subject Area	Purpose of Documentation	Issues Included
Conceptual Description	**Characterization (qualitative and quantitative) of the area's nonattainment problem; used to guide the development of the modeling analysis.**	**Emissions and air quality assessment;** **Measurements used;** **Analyses performed;** **Processes, conditions, and influences for ozone, PM, and/or regional haze formation.**

Subject Area	Purpose of Documentation	Issues Included
Modeling/Analysis Protocol	**Communicate scope of the analysis and document stakeholder involvement.**	**Names of organizations participating in preparing and implementing the protocol;** **Types of analyses performed; Steps followed in each type of analyses;** **Rationale for choice of the modeling system and model configurations.**
Emissions Preparations and Results	**Assurance of valid, consistent emissions data base. 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;** **VOC, NOx, SO2, NH3, $PM_{2.5}$, PM10, and CO (as appropriate) total emissions by State/County and for major source categories.**

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**	**Description of data base and procedures used to derive and quality assure inputs for modeling;** **Departures from guidance and their underlying rationale.** **Performance of meteorological model 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 on the days selected for analysis for each nonattainment area and appropriate sub-regions.**	**Summary of observational data base available for comparison;** **Identification of performance tests used and their results (including diagnostic analyses);** **Ability to reproduce observed temporal and spatial patterns;** **Overall assessment of what the performance evaluation implies.**

Subject Area	Purpose of Documentation	Issues Included
Description of the Strategy Demonstrating Attainment	**Provide the EPA and the public an overview of the plan selected in the attainment demonstration.**	**Qualitative description of the attainment strategy;** **Reductions in VOC, NOx, SO2, NH3, and/or CO emissions from each major source category for each State/county/Tribal land from current (identify) emission levels;** **Clean Air Act mandated reductions and other reductions;** **Show predicted relative response factors for ozone and/or each component of $PM_{2.5}$ and regional haze(as applicable)** **Show predicted future design values for the selected control scenario and identify any location(s) which fails the unmonitored area analysis described in Section 3;** **Identification of authority for implementing emission reductions in the attainment strategy.** **Evidence that emissions remain at or below projected levels throughout the 3-year period used to determine future attainment for $PM_{2.5}$ and ozone and/or 5-year period for uniform rate of progress assessments.**

Subject Area	Purpose of Documentation	Issues Included
Data Access	**Enable the EPA or other interested parties to replicate model performance and attainment simulation results, 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.**

Subject Area	Purpose of Documentation	Issues Included
Supplemental Analyses/Weight of Evidence Determination	**Assure the EPA and the public that the strategy meets applicable attainment tests and is likely to produce attainment of the NAAQS and/or uniform rate of progress by the required time.**	**Description of the modeled test and observational data base used;** **Identification of air quality model 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.**
Review Procedures Used	**Provide assurance to the EPA and the public that analyses performed in the attainment 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.**

This page is intentionally left blank

Part II. How Should I Apply Air Quality Models To Produce Results Needed To Help Demonstrate Attainment?

10.0 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 or meet the uniform rate of progress goals within a required time frame. 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 attainment demonstration. The procedure we recommend consists of nine steps:

1. Formulate a conceptual description of an area's nonattainment problem;
2. Develop a modeling/analysis protocol;
3. Select an appropriate air quality model to use;
4. Select appropriate meteorological episodes to model;
5. Choose a modeling domain with appropriate horizontal and vertical resolution and establish the initial and boundary conditions to be used;
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, as necessary.
9. Perform future year modeling (including additional control strategies, if necessary) and apply the attainment test

In this section, we briefly describe each of these steps to better illustrate how they are inter-related. Because many of these steps require considerable effort to execute, States/Tribes should 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. The steps outlined in this section are described in greater depth in Sections 11 - 18.

1. Formulate a conceptual description of an area's nonattainment problem. A State/Tribe needs to have an understanding of the nature of an area's nonattainment problem before it can proceed with a modeled attainment demonstration. For example, it would be difficult to identify appropriate stakeholders and develop a modeling protocol without knowing whether resolution of the problem may require close coordination and cooperation with other nearby States.

The State/Tribe containing the designated nonattainment area or the RPO containing the Class I area is expected to initially characterize the problem. This characterization 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? 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 precursor emissions? Are there ambient measurements suggesting which precursors and sources are important to further reduce ozone, $PM_{2.5}$, and/or regional haze? 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 Sections 7 and 8 (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 demonstration or uniform rate of progress analysis. The content of the protocol and identification of participating stakeholders are influenced by the previously developed conceptual description of the problem. The protocol outlines methods and procedures which will be used to perform the subsequent steps needed to generate the modeling results and then apply the modeled attainment test and unmonitored area analysis as well as other corroborating analyses. This is accomplished by: a) identifying those responsible for implementing the modeling, b) outlining the specific steps needed to complete the attainment demonstration, c) identifying those who will review each step as it occurs, and d) identifying procedures to be used to consider input/suggestions from those potentially affected by the outcome (i.e., "stakeholders"). In short, the protocol defines the "game plan" and the "rules of the game".

3. Select an appropriate model for use. This step includes reviewing non-proprietary, grid-based photochemical models to select the model that is most appropriate for the application in terms of (a) state-of-the science algorithms to represent the chemical and physical processes associated with ozone and/or PM formation, transport, and removal, (b) peer review, (c) model performance in prior applications, and (d) ease of use. Identifying the 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. For $PM_{2.5}$ nonattainment applications where large spatial gradients of primary $PM_{2.5}$ exist, a dispersion model may be needed in addition to a photochemical model.

4. Select appropriate meteorological time periods to model. Like the preceding step, this step requires review of available air quality and meteorological data. It also requires a thorough understanding of the form of the ozone and $PM_{2.5}$ NAAQS, the regional haze goals, and the modeled attainment tests described in Sections 3-6. Finally, it requires a review of meteorological conditions which represent annual $PM_{2.5}$ concentrations, both good and bad visibility days, and time periods which have been observed to accompany monitored exceedances of the 8-hour ozone or 24-hour $PM_{2.5}$ NAAQS. The object of these reviews is to select periods which a) include days with observed concentrations exceeding site-specific design values (8-hour ozone or 24-hour $PM_{2.5}$ NAAQS), b) select a representative mix of days for each quarter in applications dealing with the annual NAAQS and/or c) select days which are representative of those corresponding with good and poor visibility.

Due to increased computer speeds, it is now prudent to recommend modeling relatively long time periods. For 8-hour ozone and 24-hour $PM_{2.5}$, at a minimum, modeling episodes which

cover full synoptic cycles is desirable. Depending on the area and the time of year, a synoptic cycle may be anywhere from 5-15 days. Modeling even longer time periods of up to a full season may simplify the episode selection process and provide a rich database with which to apply the modeled attainment test. For annual average PM and regional haze applications, we recommend modeling (at least) a full year.

5. Choose a modeling domain with appropriate horizontal and vertical resolution and establish the initial and boundary conditions. Nested grid models will typically be used to support the modeled attainment test. In order to provide reasonable boundary conditions for the local nonattainment area, in many cases it is important to model a large regional domain with relatively coarse resolution, and a smaller sub-regional domain with relatively fine horizontal resolution. Meteorological and air quality data corresponding to the time periods that will be modeled need to be reviewed prior to choosing size of the area modeled. Appropriate domain size is influenced by the air quality goal being addressed, whether the model is being applied to address ozone and/or primary and/or secondary particulate matter, and the choice of days modeled.

The 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 size of individual grid cells and the number of required vertical layers for that portion of the modeling grid. Another factor affecting the choice of grid cell size is the available spatial detail in the emissions data used as input to an emissions model. Finally, factors which cannot be ignored in choosing size of a domain and its grid cells include the feasibility of managing large data bases and the resources needed to estimate meteorological inputs and air quality in many grid cells.

6. Generate meteorological inputs to the air quality simulation model. Prognostic meteorological models will ordinarily be used to generate the meteorological inputs used in the attainment demonstration modeling. The application of meteorological models and the choice of model grid resolution in the preceding step are closely related. Meteorological conditions near the area which is the focus of an attainment 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 cost/feasibility of running air quality and meteorological models and resolution at which it might be most desirable to treat dispersion of nearby emissions.

7. Generate emissions inputs to the air quality simulation model. Emissions are the central focus in a modeled attainment demonstration because they are the only input which is altered between the present and future case scenarios and represent the model input to which control strategies are applied. Emissions inputs to an 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 are affected by meteorological conditions. This requires an interface between meteorological inputs and emissions. The development of emissions data must also take into

account the horizontal/vertical model resolution of the model configuration and the size of the area to be 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. Emission inputs may be needed for number of scenarios including; (1) a base case corresponding to that of the selected meteorological time periods, (2) a baseline corresponding to that represented by the baseline monitored design value, (3) a future base case when attainment of the NAAQS or regional haze progress needs to be demonstrated, and (4) control scenarios in which additional emissions controls are applied to emissions in the future base case.

8. Evaluate performance of the air quality simulation model and perform diagnostic tests. The credibility of a modeled attainment test and other model outputs is affected by how well the model replicates observed historical air quality. Evaluating model performance and conducting diagnostic tests depend on the prior definition of the modeling exercise and specification of model inputs. Hence, this is generally the last step prior to using the model to support an attainment demonstration or glidepath assessment. In the past, the performance evaluation for ozone has relied almost exclusively on numerical tests comparing predicted and observed ozone, or visual inspection of predictions and observations. These are still important tools. However, photochemical grid models have many inputs, and it is possible to get similar predicted ozone and/or $PM_{2.5}$ concentrations with different combinations of inputs. There is no guarantee that ozone and/or secondary $PM_{2.5}$ will respond the same way to controls with these different combinations of inputs. Thus, we place greater emphasis on additional kinds of tests than was the case in past guidance. These include use of precursor observations, indicator species, and corroborative analyses with observational models. Diagnostic tests are separate simulations which are performed to determine the sensitivity of a model's predictions to various inputs to the model. This can be done for a variety of purposes, including selection of effective control strategies, prioritizing inputs needing greatest quality assurance and assessing uncertainty associated with model predictions. In performing such tests, States/Tribes should remember how model results are used in the modeled attainment tests recommended in Sections 3-6. Model results are used in a relative rather than absolute sense. Thus, diagnostic tests should be used to consider how relative, as well as absolute predictions, are affected by changes to model inputs.

9. Perform future year modeling (including additional control strategies, if necessary) and apply the attainment test. The base case model runs for performance evaluations should contain emissions inventories on a highly resolved basis which can best simulate the ozone and/or $PM_{2.5}$ concentrations that were measured. For some sources, it may not be appropriate to project day specific emissions to the future because they may not be representative of typical base case conditions. This is commonly the case for wildfire and continuous emissions monitor (CEM) based utility emissions. If needed, a separate baseline model run should be completed for the purpose of establishing a base to compare against future year model outputs (for calculating relative response factors).

The next step is to run the future year base case model run. The inventory should contain all known emissions controls expected to be in place in the future year, as well as

projected growth of emissions to the future. The attainment test should be performed using the future base case and the (current year) baseline. If attainment cannot be shown, then model runs which contain additional control measures are needed. Multiple future year control strategy runs may need to be completed until the attainment test is passed.

11.0 How Do I Get Started?- A "Conceptual Description"

A State/Tribe should start developing information to support a modeled attainment demonstration by assembling and reviewing available air quality, emissions and meteorological data. Baseline design values should be calculated at each monitoring site, as described in Section 3. For PM applications, 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/Tribe 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 the choice of initial diagnostic tests to identify potentially effective control strategies. In general, a conceptual description is useful for helping a State/Tribe identify priorities and allocate resources in performing a modeled demonstration.

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.

11.1 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. Examples are listed below. There are 3 different examples. One each for ozone, $PM_{2.5}$, and regional haze. The examples are not comprehensive. There could be other features of an area's problem which are important in particular cases. For purposes of illustration later in the discussion, we have answered each of the questions posed below. Our responses appear in parentheses.

11.1.1 8-Hour Ozone NAAQS

1. Is the nonattainment problem primarily a local one, or are regional factors important?

(Surface measurements suggest transport of ozone close to 84 ppb is likely. There are some other nonattainment areas not too far distant.)

2. Are ozone and/or precursor concentrations aloft also high?

(There are no such measurements.)

3. Do violations of the NAAQS occur at several monitoring sites throughout the nonattainment area, or are they confined to one or a small number of sites in proximity to one another?

(Violations occur at a limited number of sites, located throughout the area.)

4. Do observed 8-hour daily maximum ozone concentrations exceed 84 ppb frequently or just on a few occasions?

(This varies among the monitors from 4 times up to 12 times per year.)

5. When 8-hour daily maxima in excess of 84 ppb occur, is there an accompanying characteristic spatial pattern, or is there a variety of spatial patterns?

(A variety of patterns is seen.)

6. 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.)

7. Have there been any recent major changes in emissions of VOC or NOx in or near the nonattainment area? If so, what changes have occurred?

(Yes, several local measures [include a list] believed to result in major reductions in VOC [quantify in tons per summer day] have been implemented in the last five years. Additionally, the area has seen large regional NOx reductions from the NOx SIP call.)

8. Are there discernible trends in design values or other air quality indicators which have accompanied a change in emissions?

(Yes, design values have decreased by about 10% at four sites over the past [x] years. Smaller or no reductions are seen at three other sites.)

9. Is there any apparent spatial pattern to the trends in design values?

(No.)

10. Have ambient precursor concentrations or measured VOC species profiles changed?

(There are no measurements.)

11. What past modeling has been performed and what do the results suggest?

(A regional modeling analysis has been performed. Two emission scenarios were modeled: current emissions and a substantial reduction in NOx emissions throughout the regional domain. Reduced NOx emissions led to substantial predicted reductions in 8-hour daily maximum ozone in most locations, but changes near the most populated 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 8-hour daily maxima greater than 84 ppb?

(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 $\geq$ 85 F on these days.)

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 and 11 suggest there is a significant regional component to the area's nonattainment problem. Second, responses to questions 3, 4, 7, 8, and 11 indicate there is an important local component to the area's nonattainment problem. The responses to questions 4, 5 and 12 indicate that high ozone concentrations may be observed under several sets of meteorological conditions. The responses to questions 7, 8, and 11 suggest that ozone in and near the nonattainment area may be responsive to both VOC and NOx controls and that the extent of this response may vary spatially. The response to question 6 suggests that it may be appropriate to develop a strategy using a model with 12 km grid cells.

The preceding conceptual description implies that the State/Tribe containing the nonattainment area in this example will need to involve stakeholders from other, nearby States/Tribes to develop and implement a modeling/analysis protocol. It also suggests that a nested regional modeling analysis will be needed to address the problem. Further, it may be necessary to model at least several distinctive types of episodes and additional analyses will be needed to select episodes. Finally, sensitivity (i.e., diagnostic) tests, or other modeling probing tools, will be needed to assess the effects of reducing VOC and NOx emissions separately and at the same time.

11.1.2 Annual $PM_{2.5}$ 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 $PM_{2.5}$. There are higher concentrations of primary $PM_{2.5}$ in the core urban area compared to the suburbs and more rural areas.)

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 other primary particulate matter).

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 other primary $PM_{2.5}$ (OPP) appears less than what might be inferred from the inventory).

5. Do there appear to be any areas with large gradients of primary $PM_{2.5}$ in monitored or unmonitored areas?

(Cannot really tell for sources of primary $PM_{2.5}$ material until we resolve the preceding inventory/monitoring discrepancy. There are no other obvious major sources of primary particulate matter).

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, measures believed to result in major reductions in VOC and NOx have been implemented in the last 5 years. Reductions in power plant NOx have resulted from the NOx SIP call and SO_2 emissions reductions 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 trend appears to be downward, but the most recent air quality data has been higher. Overall, 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 and $PM_{2.5}$. Two emission scenarios were modeled: current emissions and a substantial reduction in NOx and SO2 emissions throughout a regional domain. Reduced NOx emissions led to substantial predicted reductions in 8-hour daily maximum ozone in most locations. Modeled SO2 reductions from the CAIR rule had a strong impact on sulfate concentrations)

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 often $\geq$ 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 $PM_{2.5}$ 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 and national PM control measures before simulating additional control measures. The responses to questions 4 and 5 suggest that it is premature to determine whether a "local area analysis" 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.

11.1.3 Example Regional Haze 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?

(Yes, but not always).

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 (6.1)) 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 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 12.0.

Questions like those posed in Section 11.1 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. These analyses are intended to channel resources available to support modeled attainment demonstrations onto the most productive paths possible. They will also provide other pieces of information which can be used to reinforce conclusions reached with an air quality model, or cause a reassessment of assumptions made previously in applying the model. As noted in Section 7, corroboratory analyses should be used to help assess whether a simulated control strategy is sufficient to meet the NAAQS.

12.0 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. The protocol should detail and formalize the procedures for conducting all phases of the modeling study, such as describing the background and objectives for the study, creating a schedule and organizational structure for the study, developing the input data, conducting model performance evaluations, interpreting modeling results, describing procedures for using the model to demonstrate whether proposed strategies are sufficient to attain the NAAQS and/or regional haze goals, and producing documentation to be submitted for EPA Regional Office review and approval. Much of the information in U.S. EPA (1991a) regarding modeling protocols remains applicable. States/Tribes 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 modeling as well as other supplemental analyses.

12.1 What Is The Protocol's Function?

As noted above, the most important function of a protocol is to serve as a means for planning and communicating up front how a modeled attainment demonstration will be performed. The protocol is the means by which States/Tribes, U.S. EPA, and other stakeholders can assess the applicability of default recommendations and develop alternatives. A good protocol should lead to extensive participation by stakeholders in developing the demonstration. It should also reduce the risk of spending time and resources on efforts which are unproductive or inconsistent with EPA policy.

The protocol also serves several important, more specific functions. First, it should identify who will help the State/Tribe or local air quality agency (generally the lead agency) undertake and evaluate the analyses needed to support a defensible demonstration (i.e., the stakeholders). Second, it should identify how communication will occur among States/Tribes and stakeholders to develop consensus on various issues. Third, the protocol should describe the review process applied to key steps in the demonstration. Finally, it should also describe how changes in methods and procedures or in the protocol itself will be agreed upon and communicated with stakeholders and the appropriate U.S. EPA Regional Office(s). Major steps to implement the protocol should be discussed with the appropriate U.S. EPA Regional Office(s) as they are being decided. States/Tribes may choose to update the protocol as major decisions are made concerning forthcoming analyses.

12.2 What Subjects Should Be Addressed In The Protocol?

At a minimum, States/Tribes should address the following topics in their modeling/analysis protocol:

1. Overview of Modeling/Analysis Project

a. Management structure

b. Technical committees or other communication procedures to be used

c. Participating organizations

d. Schedule for completion of attainment demonstration or uniform rate of progress analyses

e. Description of the conceptual model for the nonattainment area (or Class I area(s))

2. Model and Modeling Inputs

a. Rationale for the selection of air quality, meteorological, and emissions models

b. Modeling domain

c. Horizontal and vertical resolution

d. Specification of initial and boundary conditions

e. Episode selection

f. Description of meteorological model setup[90]

g. Development of emissions inputs

h. Geographic area identified for application of the attainment test(s)

i. Methods used to quality assure emissions, meteorological, and other model inputs

3. Model Performance Evaluation

a. Describe ambient data base

b. List evaluation procedures

c. Identify possible diagnostic testing that could be used to improve model performance

4. Supplemental Analyses

a. List additional analyses to be completed to corroborate the model attainment test

b. Outline plans for conducting a weight of evidence determination, should it be necessary

5. Procedural Requirements

a. Identify how modeling and other analyses will be archived and documented

b. Identify specific deliverables to EPA Regional Office

---

[90]States may want to document detailed meteorological modeling decisions in a separate document. At least a brief description of the meteorological model and model setup should be contained in the air quality modeling protocol.

13.0 What Should I Consider In Choosing An Air Quality Model?

Photochemical grid models are, in reality, *modeling systems* in which an emissions model, a meteorological model and an air chemistry/deposition model are applied. In this guidance, we use the term "air quality model" to mean a gridded photochemical 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, the 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 components differing from the modeling system's default combination). Thus, choosing an appropriate air quality model is among the earliest decisions to be made by those implementing the protocol. In this section, we identify a set of general requirements which an air quality model should meet in order to qualify for use in an attainment demonstration or regional haze application. We then identify several factors which will help in choosing among qualifying air quality models for a specific application. We conclude this section by identifying several air quality models which are available for use in attainment demonstrations or regional haze assessments. Meteorological and emissions models are discussed in Sections 16 and 17, respectively.

13.1 Types of Air Quality Models

States should use a photochemical grid model to simulate the effects of strategies to reduce ozone and the 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 are needed to perform uniform rate of progress assessments. Based on its conceptual description of a $PM_{2.5}$ nonattainment problem, a State could conclude that the $PM_{2.5}$ problem can be addressed solely by reducing primary components of measured $PM_{2.5.}$[91]

Greater flexibility is possible in choosing a modeling approach to address primary components of $PM_{2.5}$ (i.e., OPP, EC and the primary portion 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 dispersion) model may be used. In general, modeling primary PM components with a grid model is acceptable, but further refined dispersion modeling may be necessary in areas with large spatial gradients of primary PM. If a dispersion model is used to estimate RRF's for

[91]If this is the case, the State may not need to 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.

primary particulate matter, these estimates should be made at the monitoring location (see Section 5.3). Thus, a State may use a regional photochemical grid model to address both primary and secondary components of particulate matter or they may use a photochemical model to address secondary PM and an inert model applied over a more limited domain 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.

13.2 What Prerequisites Should An Air Quality Model Meet To Qualify For Use In An Attainment Demonstration?

A model should meet several general criteria for it to be a candidate for consideration in an attainment demonstration or uniform rate of progress assessment. These general criteria are consistent with requirements in 40 CFR 51.112 and 40 CFR part 51, Appendix W (U.S. EPA, 2005d). Note that, unlike in previous guidance (U.S. EPA, 1991a), we are not recommending a specific model for use in attainment demonstration or uniform rate of progress assessment. At present, there is no single model which has been extensively tested and shown to be clearly superior than its alternatives. Thus, 40 CFR Part 51 Appendix W does not identify a "preferred model" for use in attainment demonstrations of the NAAQS for ozone or $PM_{2.5}$ or uniform rate of progress assessments for regional haze[92]. Based on the language in 40CFR Part 51 Appendix W, models used for these purposes should meet requirements for "alternative models".

States/Tribes should use a non-proprietary model, which is a model whose source code is available for free (or for a "reasonable" cost). Furthermore, the user must be able to revise the code[93] to perform diagnostic analyses and/or to improve the model's ability to describe observations in a credible manner. Several additional prerequisites should be met for a model to be used to support an attainment demonstration or uniform rate of progress assessment.

---

[92]Appendix W recommends specific dispersion models for primary pollutants, but does not recommend specific photochemical models. Recommended dispersion models should be used for primary $PM_{2.5}$ analyses if dispersion models are used for a "local area analysis"(see Section 5.3).

[93]Air quality models are generally identified by a version number. The version of the model that is used in SIP applications should be identified. Code revisions to standard versions of models should be noted and documented.

(1) It should have received and been revised in response to a scientific peer review.

(2) It should be appropriate for the specific application on a theoretical basis.

(3) It should be used with a data base which is adequate to support its application.

(4) It should be shown to have performed well in past ozone or PM modeling applications. (If the application is the first for a particular model, then the State should note why it believes the new model is expected to perform sufficiently.)

(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 each of the major components of the modeling system (i.e., air chemistry/deposition, meteorological and emissions models) has been described and tested, and the results have been 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/Tribe to document that a "scientific peer review" has occurred. States/Tribes should then reference this documentation to gain acceptance of an air quality model for use in a modeled attainment demonstration.

13.3 What Factors Affect My Choice of A Model For A Specific Application?

States/Tribes should consider several factors as criteria for choosing a qualifying air quality model to support an attainment demonstration or uniform rate of progress assessment. These factors are: (1) documentation and past track record of candidate models in similar applications; (2) advanced science and technical features (e.g., probing tools) available in the model and/or modeling system; (3) experience of staff and available contractors; (4) required time and resources versus available time and resources; and (5) in the case of regional applications, consistency with regional models applied in adjacent regions. Finally, before the results of a selected model can be used in an attainment demonstration, the model should be shown to perform satisfactorily using the data base available for the specific application.

Documentation and Past Track Record of Candidate Models. For a model to be used in an attainment demonstration, evidence should be presented that it has been found acceptable for estimating hourly and eight-hourly ozone concentrations and/or hourly and 24-hour average $PM_{2.5}$ and PM component concentrations. 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.

Advanced Technical Features. Models are often differentiated by their available advanced science features and tools. For example, some models include advanced probing tools that allow tracking of downwind impacts from upwind emissions sources. Availability of

probing tools and/or science algorithms is a legitimate reason to choose one equally capable model over another.

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 13.2, a State/Tribe may use the factors described in this section (Section 13.3) to show that it is appropriate for use in a specific application. The selection of an "alternative model" needs to be reviewed and approved by the appropriate U.S. EPA Regional Office.

Satisfactory Model Performance in the Specific Application. Prior to use of a selected model's results in an attainment demonstration, the model should be shown to perform adequately for the specific application. The approach for evaluating model performance are discussed in Section 18.

13.4 What Are Some Examples Of Air Quality Models Which May Be Considered?

Air quality models continue to evolve and have their own strengths and weaknesses (Russell, 2000). Table 13.1 lists several current generation air quality models which have been used to simulate ambient ozone, PM, and regional haze concentrations. Table 13.2 lists several air quality models which have been used for various ozone and PM applications over the past decade, but are not widely used at this time. Table 13.3 lists several current dispersion models that have used to model primary PM. Table 13.4 lists several receptor models that have been used to identify sources of PM. The list is not intended to be comprehensive. Exclusion of a model from the list does not necessarily imply that it cannot be used to support a modeled attainment demonstration or uniform rate of progress assessment. In the same way, inclusion on the list does not necessarily imply that a model may be used for a particular application. States/Tribes should follow the guidance in Sections 13.1 and 13.2 in selecting an air quality model for a specific application.

Table 13.1 Current Air Quality Models Used To Model Ozone and PM

Air Quality Model	References
CAMx	Environ (2006a)
CMAQ	U.S. EPA (1998a)
UAM-V	Systems Applications International (1996)

Table 13.2 Other Air Quality Models Used to Model Ozone and PM

Air Quality Model	References
CALGRID	Scire, et al. (1989)
MAQSIP	MCNC (1999) Odman, et al. (1996)
SAQM	Chang, et al., (1997) CARB (1996)
URM	Kumar, et al., (1996)

Table 13.3 Current Dispersion Models Used to Model Primary PM

Air Quality Model	References
AERMOD	U.S. EPA (2004f) U.S EPA (2004g)
CALPUFF	Scire (2000)
ISC3	U.S. EPA (1995a) U.S EPA (1995b)

Table 13.4 Current Receptor Models

Air Quality Model	References
PMF	Eberly (2005) Hopke (2001)
UNMIX	Lewis (2003)
CMB	U.S. EPA (2004d)

14.0 How are the Meteorological Time Periods (Episodes) Selected?

Historically, attainment demonstrations have been based on a limited number of episodes consisting of several days each. In the past, the number of days modeled has been limited by the speed of computers and the ability to store the model output files. With the advancement in computer technology over the past decade, computer speed and storage issues are no longer an impediment to modeling long time periods. In fact, many groups have recently modeled entire summers and/or full years for ozone, $PM_{2.5}$, and regional haze (Baker, 2004a) (U.S. EPA, 2005b).

Ozone based research has shown that model performance evaluations and the response to emissions controls need to consider modeling results from relatively long time periods, in particular, full synoptic cycles or even full ozone seasons (Hogrefe, 2000). In order to examine the response to ozone control strategies, it may not be necessary to model a full ozone season (or seasons), but, at a minimum. we recommend modeling "longer" episodes that encompass full synoptic cycles. Time periods which include a ramp-up to a high ozone period and a ramp-down to cleaner conditions allow for a more complete evaluation of model performance under a variety of meteorological conditions.

More limited research has been conducted related to $PM_{2.5}$ modeling. Most model applications for the annual $PM_{2.5}$ NAAQS have modeled a full year (CAIR, RPOs, etc). This is a logical goal since every day of the year is included in the calculation of the annual NAAQS. Several tests have also been made using less than a full year of meteorological data. One study (Environ, 2004) found that modeling 14 days from each quarter provided results with a reasonable amount of uncertainty compared to modeling the full year. The following sections contain further recommendations for choosing appropriate time periods to model for attainment demonstrations and glidepath assessments.

At a minimum, four criteria should be used to select time periods which are appropriate to model:

1) Simulate a variety of meteorological conditions:
   - a) 8-Hour Ozone- Choose time periods which reflect a variety of meteorological conditions which frequently correspond with observed 8-hour daily maxima > 84 ppb at multiple monitoring sites.
   - b) 24-Hour $PM_{2.5}$- Choose time periods which reflect a variety of meteorological conditions which frequently correspond with observed 24-hour averages > 65 ug/m3 at violating monitoring sites.
   - c) Annual $PM_{2.5}$- Choose time periods from each quarter which reflect the variety of meteorological conditions which represent average concentrations for that quarter and year
   - d) Regional Haze- .Choose time periods which reflect the variety of meteorological conditions which represent visibility impairment on the 20% best and 20% worst days in the Class I areas being modeled.

2) Model time periods in which observed concentrations are close to the appropriate baseline design value or visibility impairment.

3) Model periods for which extensive air quality/meteorological data bases exist.

4) Model a sufficient number of days so that the modeled attainment test applied at each monitor violating the NAAQS is based on multiple days (see section 14.1.1).

These four criteria may sometimes conflict with one another. For example, there may only be a limited number of days with intensive data bases, and these may not cover all of the meteorological conditions which correspond with monitored concentrations close to site-specific design values during the base period. Thus, tradeoffs among the four primary criteria may be necessary in specific applications.

Those implementing the modeling/analysis protocol may use secondary episode selection criteria on a case by case basis. For example, prior experience modeling an episode or year, may result in its being chosen over an alternative. Another consideration should be to choose time periods occurring during the 5-year period which serves as the basis for the baseline design value (DVB). If observed ozone or 24-hour $PM_{2.5}$ exceedances occur on weekends, weekend days should be included within some of the selected time periods. If it has been determined that there is a need to model several nonattainment areas simultaneously (e.g., with a nested regional scale model application), a fourth secondary criterion is to choose time periods containing days of common interest to different nonattainment areas.

In this section, we first discuss each of the four identified primary criteria for choosing meteorological episodes to model. We then discuss the secondary criteria, which may be important in specific applications.

14.1 What Are The Most Important Criteria For Choosing Episodes?

In the following section, we identify how the four selection criteria apply to ozone. We then discuss issues related to time period selection for $PM_{2.5}$ and regional haze. We also recommend minimum time periods to model for 8-hour ozone, 24-hour $PM_{2.5}$, annual $PM_{2.5}$, and regional haze. As a matter of convenience, section 14.1 is organized by pollutant, but it could also be organized by the nature of the ambient data. For example, 8-hour ozone and 24-hour $PM_{2.5}$ episode selection criteria are similar because both standards are based on short term peak concentration periods. Regional haze calculations are based on an average of 20 or more days per year (20% best or worst days). Therefore regional haze episode selection will likely include more days throughout the year. The annual $PM_{2.5}$ NAAQS is unique because each and every ambient observation is included in the average. It is likely that a full year (or more) will be modeled to represent the annual $PM_{2.5}$ standard.

14.1.1 8-Hour Ozone NAAQS

Simulate 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 leading to elevated ozone concentrations. The time periods chosen should reflect the variety of conditions that lead to 8-hour ozone concentrations > 85 ppb.

Those implementing the modeling/analysis protocol should describe the rationale for distinguishing among episodes which are modeled. The selection may reflect a number of area-specific considerations. Qualitative procedures such as reviewing surface and aloft weather maps, and observed or modeled wind patterns may suffice for distinguishing episodes with distinctively different meteorological conditions. More quantitative procedures, such as a Classification and Regression Tree (CART) analysis or a principal component analysis (PCA), to identify distinctive groupings of meteorological/air quality parameters corresponding with high daily maxima or averages, may sometimes be desirable. An example of a CART analysis applied to select episodes is described by Deuel (1998). LADCO used CART to rank historical years for Midwestern cities by their conduciveness to ozone formation (Kenski, 2004). A PCA may also be used to characterize predominant meteorological conditions and relate those conditions to ozone concentrations (Battelle, 2004). This information can be used to quantify the relative "ozone forming potential" of different days, regimes, and years.

The interpretation of results of a wind rose analysis or a statistical analysis such as PCA or CART should focus on episodic time periods, rather than individual days. The winds may be blowing from different directions on consecutive days, but that does not necessarily mean that those days represent different meteorological regimes. Preference should be given to modeling episodic cycles.

Additionally, statistical analyses such as PCA normally limit the number of identified meteorological regimes to a relatively small number of generalized patterns. The analysis may indicate that only one or two of these patterns are responsible for most or all of the exceedance days in an area. But no two days and no two episodes are exactly the same. Further analysis should be performed on potential episode periods to differentiate subtle, but often important, differences between episodes. For this reason, it may be beneficial to model more than one episode from the most frequently occurring meteorological regimes which lead to ozone exceedances. Modeling a continuous time period which encompasses several episodes or a full ozone season will make it easier to adequately account for all of the potential meteorological conditions which correspond to high measured concentrations.

Choose episodes having days with monitored 8-hour daily maxima close to observed average 4th high daily maximum ozone concentrations.

We want to use episodes whose severity is comparable to that implied by the form of the NAAQS (i.e., an episode whose severity is exceeded, on average, about 3 times/year at the time of the selected episode). The objective is to choose episodes with days which are approximately

as severe as the average 4th high 8-hour daily maximum concentration specified in the NAAQS. As such, even if a full summer is being modeled, it is important to analyze the ambient data to ensure that an adequate number of ozone conducive days are modeled.

Air quality measurements recorded during the baseline/current period can be used to characterize episode severity. This is done by selecting a 5-year period which "straddles" a modeled episode. For example, if an episode from 2002 were modeled, we recommend looking at measured 8-hour daily maxima at each site in the nonattainment area during 2000-2004. Using this information it should be possible to assess the relative severity of the days chosen for modeling at each site. Limiting this characterization to the five years straddling an episode avoids problems posed by long term trends in emissions in assessing episode severity. However, it leaves unanswered the question of whether the 5-year period selected to assess severity of a modeled day is typical or atypical. If there is an underlying long term trend in ambient ozone attributable to meteorological cycles or emissions changes, it may not be appropriate to compare different periods with one another using air quality observations. Thus, if one uses a 10-year old episode with an exceptional data base, there is greater uncertainty in ranking its severity relative to the current period of interest than if the episode were drawn from the current period.

Note that if the episode is drawn from a recent time period (especially the three years upon which the nonattainment designation is based), days which are chosen are likely to have monitored observations very close to the baseline design value. In the absence of such information, we suggest "$\pm$ 10 ppb" as a default recommendation for purposes of prioritizing choice of episodes[94]. If the base and baseline/current periods do not coincide, "close to" is within $\pm$ 10 ppb of the design value during the base period straddling the episode. If it is not feasible to meet this default criterion for all monitoring sites, meeting it at sites with *baseline/current* design values $\geq$ 85 ppb should receive greatest priority.

Choose days with intensive data bases.

Preference should be given to days with measurements aloft, available measurements of indicator species (see Section 18) and/or precursor measurements. These preferences result from a desire to incorporate a rigorous model performance evaluation as a part of the attainment demonstration. This reduces the likelihood of "getting the right answer for the wrong reason". Thus, the likelihood of mischaracterizing ozone/precursor sensitivity is reduced.

---

[94]The analysis in section 4.1 showed that relative model response may differ on "low" ozone days compared to days that are at or above the level of the NAAQS. Therefore, ambient (and modeled) concentrations that are more than 10 ppb above the design value are preferable to episodes with ambient concentrations that are more than 10 ppb below the design value.

Choose a sufficient number of days to enable the monitored attainment test to be based on multiple days at each monitoring site violating the NAAQS.

Figure 4.1 indicates that the relative response factor computed at any given site appears to be affected by the minimum threshold value. Based on an analysis of modeled data, the recommended minimum (baseline) threshold value is 85 ppb. The minimum threshold value analysis (detailed in section 4.1) was also used to examine how the number of days contained in the mean RRF calculation influences the mean RRF. It was found that, on average, a minimum of 10 modeled days (in the mean RRF calculation) produces mean RRFs that are relatively robust.

The analysis cited earlier in the guidance (Timin, 2005b) was used to help determine the minimum number of days to use in a mean RRF calculation. The dataset consisted of 206 monitoring sites which had at least 10 days with predicted 8-hour daily baseline maximum ozone concentrations > 85 ppb. In the analysis we assumed that a mean RRF calculated from a "large" set of days is more stable than an RRF calculated from a small set of days. Using information on the variability of the model response on individual days, we are able to measure the variability of the mean RRF on any subset of days. The analysis used datasets of 25, 50, and 100 days[95]. The standard deviation of the daily RRFs was used to create the datasets and measure the variability of the RRFs.

Figure 14.1 shows an example of the variability of the mean RRF as a function of the number of days in the mean RRF calculation. The example plot is for a monitoring site in Harford County, MD. The mean RRF for a 50 day sample size is 0.90 (10% ozone reduction). The standard deviation of the daily RRFs was 0.034 (3.4%)[96]. The plot shows the range of the mean RRFs calculated using a sample size ranging from 3 to 25 days (a subset of the 50 days). Each subset (3 days, 4 days, 5 days, etc.) was sampled 1000 times. As can be seen in the plot, the range of mean RRFs varies widely for a small sample size (3 days) and is relatively stable for a large sample size (25 days). As the number of days increases, the variability of the mean RRF decreases. A similar conclusion was reached in a different study (Hogrefe, 2000) which found that the RRF is more variable when based on a small number of days.

---

[95]The 25, 50, and 100 day datasets were created by calculating the standard deviation of the daily RRFs for the monitoring sites with at least 10 days > 85 ppb. The distribution of the RRF was calculated from the standard deviation. The original dataset had an actual maximum number of 30 days.

[96]The standard deviation is in "RRF units". For example, an RRF of 0.90 is equal to a10% ozone reduction. A standard deviation of 3.4% is a measure of the variability such that ±3.4% is equal to a range in mean RRF of 0.866-0.934.

The ability to accurately capture a mean RRF with a small number of days is dependent on the variability of the daily RRFs (as measure by the standard deviation). Sites with a small standard deviation of the daily RRFs will be able to replicate the large dataset mean RRF with relatively few days.

Using the available information, we were able to calculate, for each monitoring site, the number of days needed to provide a mean RRF calculation that is within ±1% and ±2% of the "large dataset" mean, with a 95% confidence interval. The number of days needed to produce a robust mean RRF is dependent on the variability of the daily RRFs (as measured by the standard deviation). Therefore, more days are needed to produce a stable RRF if the standard deviation of the daily RRFs is high.

Table 14.1 summarizes the results for the 25th, 50th (median) and 75th percentile of the standard deviation for the 206 monitoring sites. The table presents results for a range of standard deviations, a range of large datasets (25, 50, and 100 days), and both ±1% and ±2% accuracy. The table shows that for the median standard deviation of the monitoring sites (2.4%), a minimum number of 10-16 days is needed to replicate the mean RRF to within ±1% (95% of the time) and a minimum number of 5-6 days is needed to replicate the mean RRF to within ±2% (95% of the time). The table also shows that a smaller standard deviation requires fewer days and a larger standard deviation requires more days.

Value (206 sites)	**Standard Deviation**	**± 1% (25/50/100 days)**	**± 2% (25/50/100 days)**
25th Percentile	1.9%	9/10/12	3/4/4
Median	2.4%	10/13/16	5/5/6
75th Percentile	3.1%	12/17/23	6/8/9

Table 14.1- Number of days needed to replicate the 25/50/100 day dataset mean RRF to within ±1% and ±2%, with a 95% confidence interval.

Based on these results, we recommend a minimum number of 10 days to be included in the mean RRF calculation for each monitoring site. This will ensure a relatively robust mean RRF value that is within ±1% (on average) of the large dataset mean. If relatively few ozone days are being modeled or certain monitors have relatively few exceedances (above 85 ppb), then we recommend using an absolute minimum number of 5 days in the calculation.

The minimum number of days recommendations can be combined with the minimum threshold recommendation to create a hierarchy of number of days/threshold combinations that

can address any situation. The recommended minimum concentration threshold identified in section 4.1 is 85 ppb. But similar to the minimum number of days, there may be situations where there are relatively few "high" modeled ozone days at certain monitors. Therefore, when possible, we recommend using the 85 ppb threshold[97], but it is acceptable to use a threshold as low as 70 ppb.

Therefore, the following criteria should be applied to determine the number of days and the minimum threshold at each ozone monitor:

- If there are 10 or more days with daily maximum 8-hour average baseline modeled ozone > 85 ppb then use an 85 ppb threshold.
- If there are less than 10 days with daily maximum 8-hour average baseline modeled ozone > 85 ppb then reduce the threshold down to as low as 70 ppb until there are 10 days in the mean RRF calculation.
- If there are less than 10 days with daily maximum 8-hour average modeled ozone > 70 ppb then use all days > 70 ppb.
- Don't calculate an RRF for sites with less than 5 days > 70 ppb[98].

The following table illustrates several examples of the recommended hierarchy of choosing the number of days vs. the minimum threshold.

Number of Days > 70 ppb	**Number of Days > 85 ppb**	**Number of Days in Mean RRF**	**Threshold < 85 ppb?**
50	15	**15**	No
20	12	**12**	No
12	7	**10**	Yes
11	3	**10**	Yes
9	6	**9**	Yes
6	1	**6**	Yes
3	0	**N/A**	N/A

Table 14.2- Examples of the recommended hierarchy in choosing the number of days in the mean RRF calculation vs. the minimum threshold. The "number of days" refers to the number of days (at each monitor) when the daily modeled 8-hour ozone maximum is > 70 or 85 ppb.

---

[97]As noted in section 4.1, for areas with a very high design value (>110 ppb), a higher threshold might be appropriate.

[98]Any situation where there are less than 5 days available for RRF calculations at monitoring sites with relatively high concentrations, (above the NAAQS and/or close to the area-wide design value) should be discussed with the appropriate U.S. EPA Regional office(s).

In summary, States should try to model enough episode days so that the mean RRF calculation at each monitor contains a minimum of 10 days with a modeled concentration > 85 ppb. If there are less than 10 days > 85 ppb, then the threshold should be lowered until 10 days are included in the calculation. The threshold should not go below 70 ppb and the number of days should always be at least 5. In trying to meet these recommendations, the greatest priority should be given to identifying episode days with appropriate ozone concentrations at the monitoring sites with the highest design values. Sites with design values below the NAAQS should be given a low priority.

14.1.2 Recommendations for the 24-Hour $PM_{2.5}$ NAAQS

The episode selection criteria for the 24-hour $PM_{2.5}$ NAAQS are similar to the ozone criteria. The complicating factor for the 24-hour NAAQS is that exceedances of the standard can occur under a variety of meteorological conditions at different times of the year and can be due to a variety of different PM components. Therefore the universe of episodes could be large and much more varied. The form of the standard is a 3 year average of the 98th percentile concentration from each year. Therefore, modeling 3 full years of meteorological data would best capture the full range of conditions that lead to 24-hour design values. But the standard is only influenced by the very high end of the distribution of $PM_{2.5}$ days (top 2% from each year). This argues for modeling a smaller set of days on an episodic basis. We recommend one of two possible approaches:

1) Model every day for a full year (or multiple years). This is recommended for both dispersion modeling of primary $PM_{2.5}$ components and photochemical modeling of secondary and primary components. Many areas that violate the 24-hour $PM_{2.5}$ NAAQS will also violate the annual $PM_{2.5}$ NAAQS. Therefore, full year modeling may already exist or is being planned for the annual NAAQS attainment test. States that have nonattainment areas that violate both $PM_{2.5}$ standards should take advantage of planning opportunities to address both standards at the same time. Modeling at least a full year will also help ensure that a sufficient number of days are included in the RRF calculations.

2) Model episodes when high $PM_{2.5}$ concentrations occur. Similar to ozone, episodes should be selected where $PM_{2.5}$ concentrations are greater than the NAAQS (> 65 ug/m3) and are close to the baseline design value. Similar to ozone, data analyses can be completed to help select a variety of meteorological episodes which lead to high $PM_{2.5}$ concentrations. In some cases, there may be very limited conditions which lead to high 24-hour average $PM_{2.5}$ concentrations, and in other cases there may be a wide variety of cases. The specific situation in each nonattainment area will determine the number of episodes and the time periods which need to be modeled. For example, if exceedance level $PM_{2.5}$ concentrations in an area only occur in the winter, then a limited number of winter days can be modeled. In other areas, exceedance days may occur in all seasons.

The 98th percentile $PM_{2.5}$ values may be highly variable from year to year. Therefore, it is important to model a year or time periods when high $PM_{2.5}$ episodes occur in the area. The modeling should include a year and/or episodes that are representative of the 24-hour design values.

Calculation of RRFs- Regardless of the time periods modeled, component specific RRFs should be calculated for all four quarters or for all quarters in which high 24-hour $PM_{2.5}$ concentrations occur (following the recommendations in section 5.2). We do not have specific analyses that have analyzed an appropriate minimum threshold concentration for 24-hour average $PM_{2.5}$. If there are very few days that exceed the 65 ug/m3 standard, then a secondary recommendation is to include all days with concentrations > 50 ug/m3, or calculate RRFs based on the highest modeled $PM_{2.5}$ days from each quarter. A reasonable estimate of the high end of the distribution might be between the top 10% of days to the top 25% of days. We recommend that RRF calculations contain a similar number of days as for ozone. This means that each quarterly RRF calculation should contain no less than 5 days[99].

EPA will be re-examining these recommendations as they relate to the new 35 ug/m3 24-hour NAAQS as more information becomes available regarding the typical features of areas violating the revised standard. Additional guidance will be issued at a later date.

14.1.3 Annual $PM_{2.5}$

Similar to episode selection for the daily standards, the most important consideration for the annual $PM_{2.5}$ NAAQS is to model a variety of meteorological conditions throughout the year. The easiest way to accomplish this is to model a full year. Since all days count toward the annual average, each season should be represented in the modeling. If less than a full year is modeled then a minimum number of days should be modeled from each quarter.

One study (Environ, 2004) examined how well a subset of modeled days in each quarter represented modeled quarterly and annual averages. The study calculated the uncertainty associated with absolute model predictions and RRFs that were based on 7, 14, or 28 days from each quarter (the periods were contiguous). They found that the relative uncertainty of the annual average concentrations was in the range of 40-70% for SO4, 30-50% for anthropogenic organics, and 60-90% for nitrates (the low percentage is for 28 days and the high percentage is for 7 days). The uncertainty was lowered by about half when considering the uncertainty in RRFs for the 7-28 day periods vs. the annual average. This shows the advantage of using the models in a relative sense. But it also shows that modeling less than a full year introduces

---

[99]A careful analysis is needed for the days that exceed the 24-hour NAAQS. It may be necessary to divide days by species as well as by quarter. For example, exceedence days with high nitrates may respond differently than exceedence days with high organics (in the same quarter). Averaging the response from many dis-similar days may give an unrealistic mean response.

additional uncertainty into the model results (compared to modeling a full year). The conclusion from the study was that at least 14 days are needed to represent each quarter. This assumes that the model results are being used in a relative sense. The study also found that nitrate is the most difficult of the species to adequately represent with a reduced number of modeling days.

Based on current information, for annual average $PM_{2.5}$, we recommend modeling a full year (or more). Reduced modeling of 15 or more days per quarter[100] may be acceptable in some cases, but areas where nitrate is an important component of $PM_{2.5}$ should consider modeling a full year (or at least the time periods when nitrates are high). Modeling less than a full year (e.g. one month per quarter) can be useful for initial strategy development and sensitivity analyses. In this way, information can be learned from partial year runs and can be supplemented with full year model runs later in the process. This may save resources and allow more model runs to be completed (relatively) quickly.

14.1.4 Regional Haze

The goals for regional haze focus on the 20% of days with best and worst visibility. Sampling in Class I areas occurs 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 uniform rate of progress assessment. Thus, the preferred approach for regional haze-related model applications is to simulate an entire, representative year (i.e., one whose mean derived deciview values for "20% worst" and "20% best" days approximates mean values for deciviews averaged over 5 years for the best and worst days). States can then base the RRF values on the best and worst days in each Class I area for the modeled year (the ~24 best and worst days from the modeling year). For areas included in modeling for the annual $PM_{2.5}$ NAAQS, particularly if that modeling simulates an entire representative year (or more), the modeling for $PM_{2.5}$ should also provide adequate results to analyze for regional haze.

In some cases, the meteorology for a particular year may be representative of the five year period, but due to large emissions changes during the period, a single year of ambient visibility is not representative of the five year mean. One solution is to compile the emissions inventory using five year average emissions for the base period. Because the base year and future year visibility represents a five year period, it is appropriate to average both base and future year emissions in a similar manner.

---

[100]The recommendation is for 15 or more total days per quarter, not 15 or more days with monitored data.

If it is not feasible to model an entire year, then RRFs are likely to be relatively stable if at least 10 worst days (and 10 best days) are modeled (similar to the ozone recommendations). If the worst days occur throughout the year, then more time periods may need to be modeled. Worst days that occur in the same season are more likely to have similar PM components and are therefore likely to respond in a similar way.

14.2 What Additional, Secondary Criteria May Be Useful For Selecting Time Periods to Model?

In Section 14.0, we noted that there may often be conflicts among the four primary criteria recommended as the basis for choosing episodes to model. Several additional, secondary selection criteria may be helpful for resolving these conflicts.

Choose time periods which have already been modeled. That is, of course, provided that past model performance evaluation for such a time period was successful in showing that the model worked well in replicating observations. Given that the four primary criteria are met approximately as well by such time periods they are by other candidate time periods, a State/Tribe could likely save a substantial amount of work in evaluating model performance. However, large changes in emissions and precursor levels or ratios may make the use of older time periods undesirable.

Choose time periods which are drawn from the period upon which the baseline design value is based. As we note in Section 3, fewer emission estimates and fewer air quality model simulations may be needed if the base case period used to evaluate model performance, and the baseline period used in the recommended modeled attainment test are the same. Following this criterion could also make the second primary criterion more straightforward.

Choose episodes having observed concentrations "close to" the NAAQS on as many days and at as many sites as possible. This criterion mainly applies to the standards with short term averaging times (8-hour ozone and 24-hour $PM_{2.5}$). It is related to the modeled attainment test and to the fourth primary criterion for episode selection. The more days and sites for which it is reasonable to apply the test, the greater the confidence in the outcome of the modeled attainment test.

It is desirable to include weekend days among those chosen, especially if concentrations greater than the NAAQS are observed on weekends. Weekend days often reflect a different mix of emissions than occurs on weekdays[101]. This could also lead to different spatial patterns of 8-hour ozone and/or 24-hour $PM_{2.5}$ concentrations. Thus, for increased confidence that a control strategy is effective it needs to be tested on weekends as well as on weekdays. If emissions and spatial patterns of high ozone or $PM_{2.5}$ do differ on weekends versus

[101] http://www.arb.ca.gov/aqd/weekendeffect/weekendeffect.htm

weekdays, including weekend days in the choice of episodes will provide a mechanism for evaluating the accuracy of a model's response to changes in emissions.

If it has been determined that there is a need to model several nonattainment areas simultaneously, choose time periods which meet the primary and secondary criteria in as many of these nonattainment areas as possible. As discussed in Section 15, a State/Tribe or group of States/Tribes may decide to apply a model on a regional or a nested regional scale to demonstrate attainment or assess uniform rate of progress in several nonattainment areas or Class I areas at the same time. Time and resources needed for this effort could be reduced by choosing time periods which meet the primary and secondary criteria in several nonattainment areas (or Class I areas) which are modeled. Several organizations are modeling much of the Eastern U.S. for an entire year. This type of application should allow for application of the attainment tests at many locations using the appropriate set of days for each area.

15.0 What Should Be Considered When Selecting The Size And Horizontal/Vertical Resolution Of The Modeling Domain?

A modeling domain identifies the geographical bounds of the area to be modeled. The appropriate domain size depends on the nature of the strategies believed necessary to meet the air quality goal. This, in turn, depends on the degree to which air quality observations suggest that a significant part of an observed exceedance is attributable to regional concentrations which approach or exceed levels specified in the NAAQS. The choice of domain size is also affected by data base management considerations. Generally, these are less demanding for smaller domains.

Horizontal resolution is the geographic size of individual grid cells within the modeling domain. Vertical resolution is the number of grid cells (i.e., layers) considered in the vertical direction. The choice of suitable horizontal and vertical resolution depends on spatial variability in emissions, spatial precision of available emissions data, temporal and spatial variation in mixing heights, the likelihood that mesoscale or smaller scale meteorological phenomena will have a pronounced effect on precursor/pollutant relationships, data base management constraints, and any computer/cost constraints.

We begin this section by discussing factors States/Tribes should consider in choosing domain size. Next, we address the selection of horizontal grid cell size and the number of vertical layers. We conclude by discussing factors affecting the decision on the size and resolution of coarse scale and fine scale grids within a nested model.

15.1 How is the Size of the Modeling Domain Chosen?

Historically (until ~1995), ozone attainment demonstrations used urban scale modeling domains which were typically several hundred kilometers (or less) on a side. With the advent of nested grid models, most model applications began to use either relatively fine regional grids, or urban-scale inner grids nested within relatively coarse regional-scale outer grids. Recent PM modeling has generally used nationwide coarse grid domains with regional scale nested domains. We expect that most urban scale ozone and $PM_{2.5}$ attainment demonstrations will utilize a national or regional nested grid modeling approach.

The principal determinants of model domain size are the nature of the ozone and/or $PM_{2.5}$ problem and the scale of the emissions which impact the nonattainment area. Isolated nonattainment areas that are not impacted by regional transport of ozone and/or PM and its precursors may be able to use a relatively small domain[102]. Some areas of the western U.S. may fall into this category. Most nonattainment areas in the eastern U.S. have been shown to be impacted by transported ozone and/or PM and ozone and/or PM precursors from hundreds of

[102]Due to relatively long lifetimes of $PM_{2.5}$ aerosols (compared to ozone), PM modeling domains may generally need to be larger than ozone domains.

miles or more upwind of the receptor area (U.S EPA, 1998b) (U.S. EPA, 2005b). The modeling domain should be designed so that all major upwind source areas that influence the downwind nonattainment area are included in the modeling domain. The influence of boundary conditions should be minimized to the extent possible.

The inner domain of a nested model application should include the nonattainment area and surrounding counties and/or States. The size of the inner domain depends on several factors. Among them are:

1) The size of the nonattainment area.

2) Proximity to other large source areas and/or nonattainment areas.
   -Relatively isolated areas may be able to use a smaller fine grid domain.
   -Nearby source areas should be included in the fine grid.

3) Proximity of topographical features which appear to affect observed air quality.

4) Whether the model application is intended to cover multiple nonattainment areas.

5) Typical wind speeds and re-circulation patterns.
   -Very light wind speeds and re-circulation patterns may obviate the need for a large fine grid domain.

6) Whether the photochemical model utilizes one-way or two-way nested grids.
   -The fine grid domain of a model with one-way nested grids may need to be larger (compared to a model with two-way nested grids). The grid needs to be large enough to capture re-circulation due to shifting wind directions. A two-way nested grid model allows for continuous feedback from the fine grid to the coarse grid and vice versa.

7) Computer and time resource issues.

15.2 How are the Initial and Boundary Conditions Specified?

Air quality models require specification of initial conditions for model species in each grid cell in the model domain (in all layers) and boundary conditions for all grid cells along each of the boundaries (in all layers). Generation of initial and boundary conditions for individual model species include gas-phase mechanism species, aerosols, non-reactive species and tracer species. There is no satisfactory way to specify initial conditions in every grid cell. Thus, we recommend using a “ramp-up” period by beginning a simulation prior to a period of interest to

diminish the importance of initial conditions[103]. We recommend a ramp-up period of at least 2-3 days for ozone modeling. Due to longer lifetimes of fine particulate matter, we recommend a ramp-up period of at least 5-10 days for $PM_{2.5}$. In this way, the choice of initial conditions are not likely to influence the modeling results. For nested model applications, initial conditions can be specified using model predictions from the outer grid if the nested grids are started after the beginning of the simulation for the outer grid.

Boundary conditions can be specified in several ways. One option is to nest the area of interest within a much larger domain using nested regional models, as described previously. As noted above in Section 15.1, use of a large regional domain acts to diminish the importance of boundary conditions. Alternatively, initial and boundary conditions can be derived from another regional nested modeling application or from a global model[104]. Another option is to use default initial and boundary concentration profiles representing relatively clean conditions which have been formulated from available measurements or results obtained from prior modeling studies, (e.g. prescribing 35 ppb for annual ozone). Another approach to consider is using the model's simulated pollutant values (generated for emissions) averaged over one or more of the upper layers to specify a value for the lateral and top boundary conditions. This approach allows for better representation of spatial and temporal variation of boundary conditions (i.e., diurnal/seasonal/vertical profiles). This method also avoids arbitrarily guessing at the future-year boundary conditions.

If there is no larger regional model application available, then it is recommended that background boundary conditions be used to specify initial and boundary concentrations for the attainment demonstration modeling. However, concentration fields derived from a larger domain regional or global chemistry model (i.e. nesting approach) is considered more credible than the assumption of static concentrations, since the pollutant concentration fields reflect simulated atmospheric chemical and physical processes driven by assimilated meteorological observations. Therefore, we recommend using boundary conditions derived from a regional or global scale model, whenever possible.

Diagnostic testing which indicates a large impact on the model results from initial or boundary conditions may indicate that the domain is not large enough or the ramp-up period is too short. In either case, it should generally be assumed that initial and boundary conditions do

---

[103]Sensitivity simulations can be completed to determine the necessary length of the ramp-up period. A longer ramp-up period may be needed for very large domains where the characterization of long range transport is important.

[104]One atmosphere model applications for ozone and PM may commonly use global models to specify boundary conditions. This is especially important for PM species due to their long lifetimes. A number of recent studies show that long-range, intercontinental transport of pollutants is important for simulating seasonal/annual ozone and particulate matter (Jacob, 1999; Jaffe, 2003; Fiore, 2003).

not change in the future. The use of altered initial or boundary conditions in the future year should be documented and justified.

15.3 What Horizontal Grid Cell Size Is Necessary?

As we discuss in Section 16, most applications will use a prognostic meteorological model to provide meteorological inputs needed to make air quality estimates. Typically, these models are set up to produce meteorological fields for nested grids with a 3:1 ratio. In past modeling applications, the most commonly used grid cell sizes have been 108, 36, 12 and 4 km cells. In this section we provide recommendations for choosing the grid size to use as an upper limit for regional and urban scale models or for fine portions of nested regional grids.

This section only applies to the use of numerical grid models. If a Lagrangian dispersion model is used, estimates at (or near) the monitor locations should be used to calculate RRF values for primary PM.

In past guidance, we have recommended using horizontal grid cell sizes of 2-5 km in urban scale ozone modeling analyses (U.S. EPA, 1991a). Sensitivity tests performed by Kumar (1994) in the South Coast Air Basin compared hourly base case predictions obtained with 5 km versus 10 km versus 20 km grid cells. Results indicate that use of finer grid cells tends to accentuate higher hourly ozone predictions and increase localized effects of NOx titration during any 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, 1996). As in Los Angeles, the use of smaller (i.e., 4 km) grid cells again leads to higher domain wide maximum hourly ozone concentrations. However, when reviewing concentrations at specific sites, Haney (1996) found that for some hours concentrations obtained with the 12 km grid cells were higher than those obtained with the 4 km cells. Other studies have shown that model performance does not necessarily improve with finer resolution modeling. In one study, the model performance at 12 km resolution was equal to or better than performance at 4 km resolution (Gego, 2005).

Another important aspect in choosing the horizontal grid cell size is the relative response of the model at various grid cell resolutions. Recent sensitivity tests comparing relative response factors in predicted 8-hour daily maxima ozone near sites in the eastern United States indicate relatively small unbiased differences ($\leq$ .04, in 95% of the comparisons) using a grid with 12 km vs. 4 km grid cells (LADCO, 1999; Arunachalam, 2006). The largest differences in the relative response of models at varying resolution is likely to occur in oxidant limited areas. The horizontal resolution may have a large impact on the spatial distribution and magnitude of NOx "disbenefits" (i.e., ozone increases in oxidant limited areas when NOx emissions are reduced).

A recent study of the Atlanta area (Cohan, 2006) used the Decoupled Direct Method (DDM) to examine the sensitivity of ozone production efficiency to horizontal resolution. The efficiency of ozone production should be closely related to model response to emisions controls. The analysis found that 12km resolution is sufficient (in Atlanta) to predict ozone production efficiency. However, 4km resolution is needed to identify small areas of local NOx inhibition. They also found that model performance for ozone at 4km and 12km resolution was similar.

Horizontal resolution will also have an impact on primary particulate matter. Because there is no chemistry, we expect spatial gradients in concentrations to be much higher for primary PM components compared to secondary components. The model's ability to capture these gradients will depend in large part upon the horizontal resolution of the model. Therefore, emissions density is an important factor in considering grid resolution. Additionally, terrain features (mountains, water) should be considered when choosing grid cell size.

Intuitively, one would expect to get more accurate results in urban applications with smaller grid cells (e.g., 4 km) provided the spatial details in the emissions and meteorological inputs support making such predictions. Thus, using 4 km grid cells for urban or fine portions of nested regional grids and 12 km cells in coarse portions of regional grids are desirable goals. However, model performance at 4 km resolution may not be any better than at 12 km resolution. Additionally, extensive use of urban grids with 4 vs. 12 km grid cells and regional grids with 12 vs. 36 km grid cells greatly increases computer costs, run times and data base management needs. Further, elsewhere in this guidance we identify needs to model large domains, many days, and several emission control scenarios. We also identify a number of diagnostic tests which would be desirable and suggest using more vertical layers than has commonly been done in the past. Also, there may be ways of dealing with potential problems posed by using larger than desired grid cells. For example, use of plume-in-grid algorithms for large point sources of NOx should be considered as an alternative with coarser than desired grid cells. And dispersion modeling of primary PM can be used to offset the need of fine spatial resolution in a grid model.

The relative importance of using a domain with grid cells as small as 4 km should be weighed on a case by case basis by those implementing the modeling/analysis protocol. The most important factor to consider is model response to emissions controls. Analysis of ambient data, sensitivity modeling, and past modeling results can be used to evaluate the expected response to emissions controls at various horizontal resolutions for both ozone and $PM_{2.5}$. If model response is expected to be different (and more accurate) at higher resolution, then higher resolution modeling should be considered. If model response is expected to be similar at both high and lowe(er) resolution, then high resolution modeling may not be necessary.

In this guidance, we identify upper limits for horizontal grid cell size which may be larger than desired for some applications. This is intended to provide flexibility for considering competing factors (e.g., number of modeled days versus grid cell size) in performing a modeling analysis within the limits of time and resources.

15.3.1 Horizontal Resolution Recommendations for Ozone

For coarse portions of regional grids, we recommend a grid cell size of 12 km if feasible, but not larger than 36 km. For urban and fine scale portions of nested regional grids, it may be desirable to use grid cells about 4 km, but not larger than 12 km. States/Tribes should examine past model applications and data analyses for their area when choosing the fine grid resolution. Past model applications and data analyses may help determine whether a grid cell size as small as 4 km (or smaller) is necessary for a particular area. Model performance and the relative response to emissions controls should be considered in the decision. States/Tribes should consider diagnostic tests to assess the difference in model performance and response from varying model grid resolution, particularly in oxidant-limited areas.

All ozone monitor locations within a nonattainment area should ordinarily be placed within the fine scale portion of a nested regional grid if nested models are used. States/Tribes choosing an urban grid or fine portion of a nested grid with cells 12 km or larger should consider applying plume-in-grid algorithms to major point sources of NOx. The use of plume-in-grid should be discussed with the appropriate EPA Regional Office.

15.3.2 Horizontal Resolution Recommendations for the $PM_{2.5}$ NAAQS

The formation of secondary particulate matter shares the same basic photochemical processes as ozone. In fact, they are closely interconnected in many respects. Since there appears to be little difference in RRF estimates for 8-hour ozone with 12 km vs. 4 km grid cells, we would expect this finding to hold for secondary components of particulate matter. But what about primary components of particulate matter? Because there is no chemistry, we expect spatial gradients in 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. Although it is clear that spatial resolution of primary PM will impact the predicted concentrations, it is not clear how it will impact the relative change in concentrations due to controls (RRFs). Areas without strong gradients in primary PM will likely have little benefit from fine scale resolution. Areas that have large gradients in primary PM may need to use finer resolution grid models or may need to supplement the grid modeling with dispersion modeling (see Section 5.3) . This particularly true if violating monitors are strongly impacted by one or more large sources of primary PM.

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, in some cases, cells as small as 4 km (or possibly smaller) 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 decreases the grid cell size from 12 km to 4km. Alternatively, large sources of primary PM can be

modeled with a dispersion model or a combination of grid and dispersion models (see Section 5.3).

We expect that modeling analyses for nonattainment areas will use 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. For the regional outer nest of the domain, grid cells as large as 36 km may be used.

15.3.3 Horizontal Resolution Recommendations for Regional Haze

The glidepath analysis for regional haze 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 efficiency of light scattering by secondary particles. All these attributes indicate that it will be necessary to model a regional scale domain for regional haze related applications. 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 wish to perform diagnostic tests using plume-in- grid analyses, as well as finer horizontal resolution to determine if results may differ using more finely resolved emissions and meteorology. Guidance on vertical resolution presented in Section 15..4 is also applicable for regional haze-related applications.

15.4 How Should the Vertical Layers Be Selected?

As described in Section 16, the preferred approach for generating meteorological data fields for input to air quality simulation models is to use a prognostic meteorological model with four dimensional data assimilation (FDDA). Such models normally use more than 30 vertical 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 as many as 30 vertical layers may not be feasible or cost effective. In this section we identify factors to consider in choosing the number of vertical layers chosen for the air quality model applications.

In the past, short ozone episodes of only several days usually encompassed periods of mostly clear skies with very little precipitation. As such, ozone models often did not explicitly model clouds or precipitation. However, we are recommending modeling longer episodes (or even a full ozone season) for ozone and a full year for $PM_{2.5}$ and regional haze applications. Therefore the photochemical model needs to account for cloud processes and a full range of precipitation types. In order to adequately parameterize these processes, the top of the modeling domain should typically be set at the100 millibar level (~16,000 meters). In turn, this means that many more vertical layers will be needed to capture the meteorological processes both below and above the boundary layer, up to the top of the model.

The accuracy of predicted base case ozone and PM concentrations will be affected by how well the model is able to characterize dilution of ozone, PM, and precursors. This, in turn, depends in part on how precisely the model can estimate mixing heights (i.e., the PBL). The precision of mixing height estimates is affected by the thickness of the model's vertical layers aloft which are near the anticipated mixing height (Dolwick, 1999). Ozone concentrations are most heavily influenced by the rate of rise in the morning mixing height and the maximum afternoon mixing height. Since PM is measured as a 24-hour average concentration, PM is also influenced by the strength of nighttime mixing and the presence of a low level inversion. Because 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 vertical layers considered by the model.

Placement of vertical layers within the planetary boundary layer is also an important issue. For practical reasons, it is best to have an air quality model's vertical layers align with the interface between layers in the meteorological model. In view of the importance of carefully specifying the temporal variation in mixing height, we recommend high precision below and near the anticipated maximum afternoon mixing height. In addition, specifying the vertical extent of mixing overnight during stable conditions is also an important consideration in determining the vertical layer structure. In this regard, we recommend that the lowest layer in the air quality model be no more than 50 meters thick. In general, layers below the daytime mixing height should not be too thick, or large unrealistic step increases in mixing may occur. Layers above the boundary layer are important for characterizing clouds and precipitation, but are less important to the daily mixing processes of pollutants. Therefore, vertical resolution above the boundary layer is typically much coarser.

There is no correct minimum number of vertical layers needed in an attainment demonstration. The vertical resolution will vary depending on the application. Recent applications of one atmosphere models (with model tops at 100mb) have used anywhere from 12 to 21 vertical layers with 8-15 layers approximately within the boundary layer (below 2500m) and 4-6 layers above the PBL (Baker, 2004b), (Hu, 2004).

There are also ozone model applications which may not need to consider the full set of meteorological data through the tropopause. These applications typically use vertical domains which extend up to 4 or 5 km. These applications are most appropriate for short ozone episodes that occur under high pressure conditions (little cloud cover or precipitation). In these cases, fewer vertical layers are needed to represent the atmosphere up to the top of the domain (4-5 km). However, where appropriate, EPA encourages the use of full-scale one-atmosphere models which account for all atmospheric processes up to ~100 mb.

16.0 How are the Meteorological Inputs Prepared for Air Quality Modeling?

In order to solve for the change in pollutant concentrations over time and space, air quality models require certain meteorological inputs that help simulate the formation, transport, and removal of pollutant material. The required meteorological inputs can vary by air quality model, but consistently involve parameters such as wind, vertical mixing, temperature, moisture, and solar radiation. While model inputs can be derived strictly from ambient measurements, a more credible technical approach is to use meteorological grid models to provide the necessary inputs. When these models are applied retrospectively (i.e., for historical time periods) they are able to blend ambient data with model predictions via four-dimensional data assimilation (FDDA), thereby yielding temporally and spatially complete data sets that are grounded by actual observations.

This section provides recommendations for generating, or otherwise acquiring, the meteorological data sets sufficient for air quality modeling purposes. Additional suggestions are provided to assist in the configuration of standard meteorological modeling analyses. The last section outlines procedures for evaluating whether the meteorological input is of sufficient quality for input into the air quality model. In general, it is recommended that States/Tribes spend appropriate effort in accurately characterizing the meteorological fields in view of several sensitivity runs which show that relatively small perturbations in meteorological inputs can have large impacts on resultant air quality modeling results (Dolwick, 2002).

16.1 What Issues are Involved in the Generation and Acquisition of Meteorological Modeling Data?

The recommended approach for generating the meteorological data needed to conduct the attainment demonstration is to apply dynamic meteorological models with FDDA. These models use the fundamental equations of momentum, thermodynamics, and moisture to determine the evolution of specific meteorological variables from a given initial state. When modeling past events, the use of data assimilation helps to "nudge" solutions so that they do not diverge greatly from the actual observed meteorological fields. A major benefit of using dynamic meteorological models is that they provide a way of consistently characterizing meteorological conditions at times and locations where observations do not exist. Examples of frequently used meteorological models are listed below:

- The Penn State University / National Center for Atmospheric Research mesoscale model known as MM5 (Grell, 1994),
- The Regional Atmospheric Modeling System (RAMS) (Pielke, 1992), and
- The Weather Research and Forecasting Model (WRF) (Skamarock, 2005).

Recent advances in relatively low-cost computational power have resulted in widespread use of MM5 and similar models for air pollution applications over the past decade (Olerud, 2000; Doty, 2001; Johnson, 2003; Baker, 2004b). EPA expects that the large majority of future

attainment demonstration analyses will be based on meteorological data from these types of dynamic meteorological models. However, there are other acceptable means of generating meteorological data for an air quality modeling simulation. Over the next several years, EPA further expects that more meteorological input data sets will be developed from archived National Weather Service (NWS) model simulations. It is possible that data from archived model simulations (such as WRF) could be used to feed air quality simulations. Some of these prognostic meteorological models are already being used to drive real-time air quality forecasts (McQueen, 2004). Additionally, in some cases the dynamic meteorological model may not adequately capture key meteorological elements of an airshed's conceptual model (e.g., source-receptor transport vectors to key monitoring locations). In cases such as these, it may be appropriate to blend the dynamic model data with wind data from an objective analysis of observed wind fields.

A description of the methods used to generate the meteorological fields should be included in the modeling protocol. In cases in which standard meteorological modeling (e.g., MM5, RAMS, or WRF in a retrospective analysis mode) is not used, it is recommended that a detailed description of the technique that is used to generate the three-dimensional meteorological fields be shared with the appropriate EPA regional office(s) prior to conducting the air quality modeling analysis.

16.2 How Should the Meteorological Modeling Analysis be Configured?

As with other parts of the air quality modeling system, choices made regarding how to configure the meteorological modeling can affect the quality and suitability of the air quality model predictions. Decisions regarding the configuration of complex dynamic meteorological models can be particularly challenging because of the amount of flexibility available to the user. The following are recommendations on how to establish the configuration of a meteorological model for air quality analyses. The focal point in establishing the proper meteorological modeling configuration should be to get the best possible meteorological model performance.

Selecting a Model Domain: As noted in Section 15, it is expected that most attainment demonstrations will cover large areas and use nested grids. The outermost grid should capture all upwind areas that can reasonably be expected to influence local concentrations of ozone, and/or $PM_{2.5}$. In terms of selecting an appropriate meteorological modeling domain, one should extend the grid 3 to 6 cells beyond the domains of each air quality modeling grid to avoid boundary effects. For example, if 4 km grid cells are to be used in the fine portion of a nested regional air quality model, then the meteorological fields at this detail would need to extend 12-24 km beyond the bounds of the 4 km grid used for air quality predictions. In terms of grid resolution, EPA recommends that the dynamic meteorological models use the same grid resolution as desired for the air quality model applications. In some cases, however, this may not always be feasible. One possible reason for modeling with meteorology using a different grid resolution is in the case of unacceptable model performance from the meteorological model at the desired grid resolution. In other instances, the need for finer resolution may be

emissions-driven more than meteorologically-driven and the costs do not warrant the generation of additional resolution in the meteorological data. In these specific situations it is recommended that the air quality model application use available results from meteorological models on the next coarser scale (i.e., 36 km for a desired 12 km estimate, 12 km for a desired 4 km estimate). The coarse grid meteorological fields can be mapped to the more finely resolved air quality modeling domain.

Selecting Physics Options: Most meteorological models have a suite of "physics options" that allow users to select how a given feature will be simulated. For example, there may be several options for specifying the planetary boundary layer scheme or the cumulus parameterization. In many situations, the "optimal" configuration cannot be determined without performing an initial series of sensitivity tests which consider various combinations of physics options over specific time periods and regions. While these tests may not ultimately conclude that any one configuration is clearly superior at all times and in all areas, it is recommended that these sensitivity tests be completed, as they should lead to a modeling analysis that is best-suited for the domain and period being simulated. Examples of sensitivity analyses can be found in McNally (2002), Olerud, (2003), and Huang, (2005). Typically, the model configuration which yields predictions that provide the best statistical match with observed data over the most cases (episodes, regions, etc.) is the one that should be chosen, although other more qualitative information can also be considered. Additionally, model configurations should be designed to account for the pollutants and time periods that are of most interest. As an example, a wintertime PM simulation in the Midwest (with high measured nitrate concentrations) may need a meteorological model configuration that employs a land-surface model that properly handles snow cover fields and their effects on boundary layer humidities and temperatures.

Use of Data Assimilation: As noted above, the use of FDDA helps to keep the model predictions from widely diverging from what was actually observed to occur at a particular point in time/space. However, if used improperly, FDDA can significantly degrade overall model performance and introduce computational artifacts (Tesche and McNally, 2001). Inappropriately strong nudging coefficients can distort the magnitude of the physical terms in the underlying atmospheric thermodynamic equations and result in "patchwork" meteorological fields with strong gradients between near-site grid cells and the remainder of the grid. Additionally, if specific meteorological features are expected to be important for predicting the location and amount of pollution formed, based on an area's conceptual model, then the meteorological modeling should be set up to ensure that FDDA does not prevent the model from forming these features (e.g. nocturnal low-level wind jets). In general, analysis nudging strengths should be no greater than 1.0×10^{-4} for winds and temperatures and 1.0×10^{-5} for humidity. In the case of observation nudging (i.e., FDDA based on individual observations as opposed to analysis fields), it is recommended that the resultant meteorological fields be examined to ensure that the results over the entire domain are still consistent. Further, based on past experience, we recommend against using FDDA below the boundary layer for thermodynamic variables like temperature and humidity because of the potential for spurious convection. If the dynamic model is applied without FDDA, it is suggested that the simulation durations be shorter than 24 hours.

Conversion of Meteorological Outputs to Air Quality Model Inputs: Even before determining how the meteorological model is configured, careful thought should be given to the compatibility between candidate meteorological models and the air quality model(s) chosen for use. A variety of post-processors exist to convert the outputs from the meteorological models into the input formats of the air quality models. Some examples include: the Meteorology Chemistry Interface Processor (MCIP) (Otte, 2004), MM5CAMx (Environ, 2006a), and RAMSCAMx (Environ, 2006a). These meteorological preprocessors provide a complete set of meteorological data needed for the air quality simulation by accounting for issues related to: 1) data format translation, 2) conversion of parameter units, 3) extraction of data for appropriate window domains, 4) reconstruction of the meteorological data on different grid and layer structures, and 5) calculation of additional variables.

16.3 How Should the Performance of the Meteorological Modeling Be Evaluated?

While the air quality models used in attainment demonstrations have consistently been subjected to a rigorous performance assessment, in many cases the meteorological inputs to these models are accepted as is, even though this component of the modeling is arguably more complex and contains a higher quantity of potential errors that could affect the results of the analysis (Tesche, 2002). EPA recommends that States/Tribes devote appropriate effort to the process of evaluating the meteorological inputs to the air quality model as we believe good meteorological model performance will yield more confidence in predictions from the air quality model. One of the objectives of this evaluation should be to determine if the meteorological model output fields represent a reasonable approximation of the actual meteorology that occurred during the modeling period. Further, because it will never be possible to exactly simulate the actual meteorological fields at all points in space/time, a second objective of the evaluation should be to identify and quantify the existing biases and errors in the meteorological predictions in order to allow for an downstream assessment of how the air quality modeling results are affected by issues associated with the meteorological data. To address both objectives, it will be necessary to complete both an operational evaluation (i.e., quantitative, statistical, and graphical comparisons) as well as a more phenomenological assessment (i.e., generally qualitative comparisons of observed features vs. their depiction in the model data).

Operational Evaluation: The operational evaluation results should focus on the values and distributions of specific meteorological parameters as paired with and compared to observed data. It is recommended that the observation - model matching be paired as closely as possible in space and time. Typical statistical comparisons of the key meteorological parameters[105] will include: comparisons of the means, mean bias, mean normalized bias, mean absolute error, mean absolute normalized error, root mean square error (systematic and unsystematic), and an index of

[105] It is difficult to say which meteorological parameters will most affect any particular modeling exercise and it will vary by parameter and location. However, in general, it is expected that following key variables should be most closely evaluated: temperature, water vapor mixing ratio, wind speed, wind direction, clouds/radiation, precipitation, and the depth and evolution of vertical mixing over the modeling periods.

agreement. For modeling exercises over large domains and entire ozone seasons or years, it is recommended that the operational evaluation be broken down into individual segments such as geographic subregions and/or months/seasons to allow for a more comprehensive assessment of the meteorological strengths and weaknesses. Other useful ways of examining model performance include: aloft, surface, individual episodes (e.g., high ozone / $PM_{2.5}$ days), diurnal cycle, as a function of synoptic regimes, or combinations of the above. It is recommended that the ambient data used in these statistical comparisons be quality checked by doing standard range check and buddy analyses. To the extent that modelers can set aside a portion of the ambient data strictly for evaluation purposes (i.e., data not used in the FDDA), that is also encouraged. Figure 16.1 shows a sample operational evaluation summary. This graphic was developed using the Atmospheric Model Evaluation Tool (AMET) (Gilliam, 2005), though other evaluation tools are available.

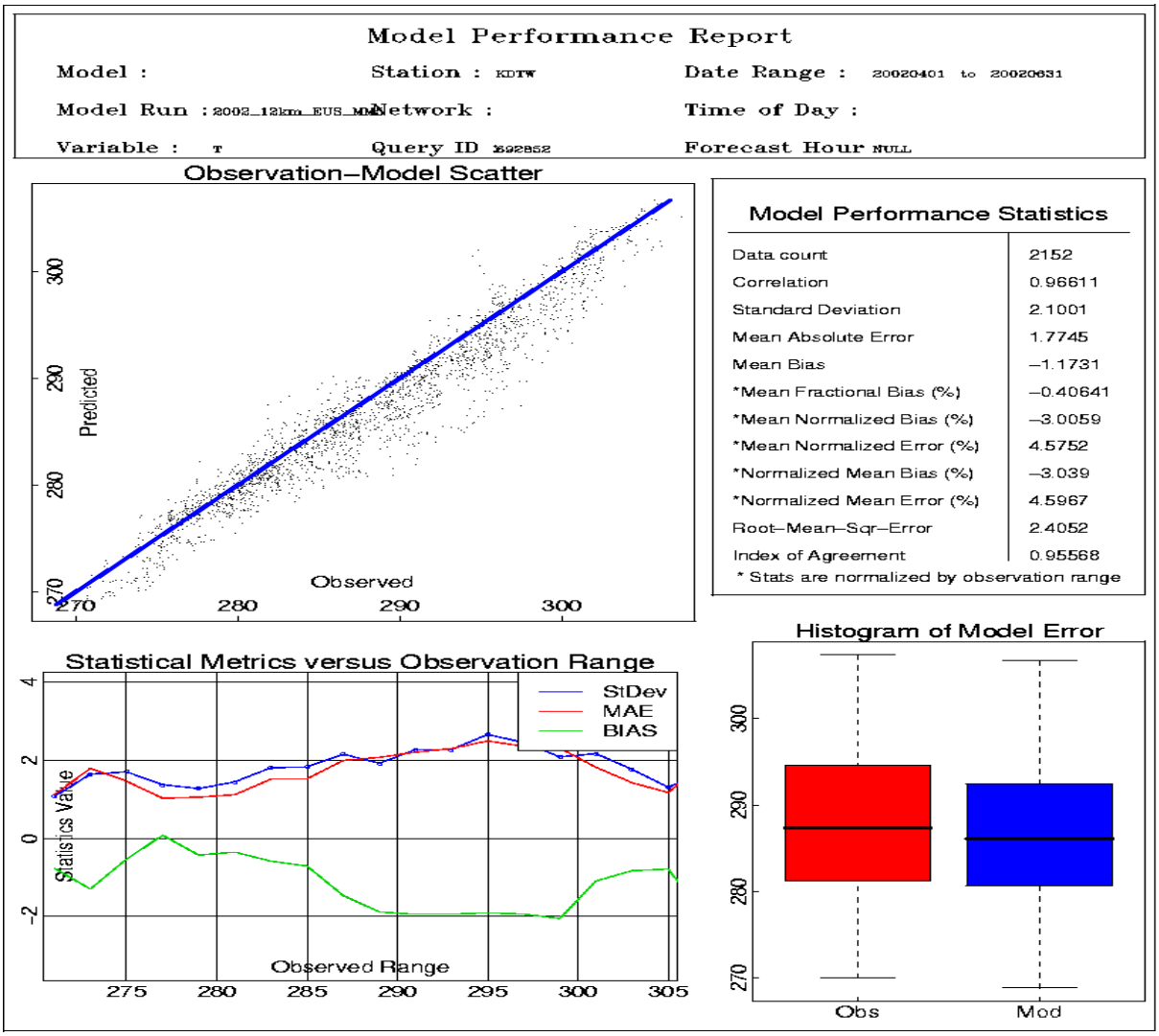


Figure 16.1 Sample operational evaluation summary for a 2002 MM5 case. This particular analysis looked at April through June temperature performance at a single site. The upper left quadrant shows a scatter plot of observed versus predicted temperatures, the lower left traces show how model performance varies as a function of the observation range, the upper right table lists some of the statistical outputs, and the lower right compares the spectrum of predictions against the observational spectrum.

It may be helpful when calculating domainwide and/or regional summary statistics to compare the results against previously generated meteorological model performance "benchmarks" (Emery et al., 2001). However, because of concerns about potentially misleading comparisons of model performance findings across different analyses with differing model configurations and FDDA strengths, EPA does not recommend using these benchmarks in a "pass/fail" mode, but only as a means of assessing general confidence in the meteorological model data. Statistical results that are outside the range of the compiled benchmarks may indicate an issue that should be given further examination. Figure 16.2 shows a sample analysis in this vein.

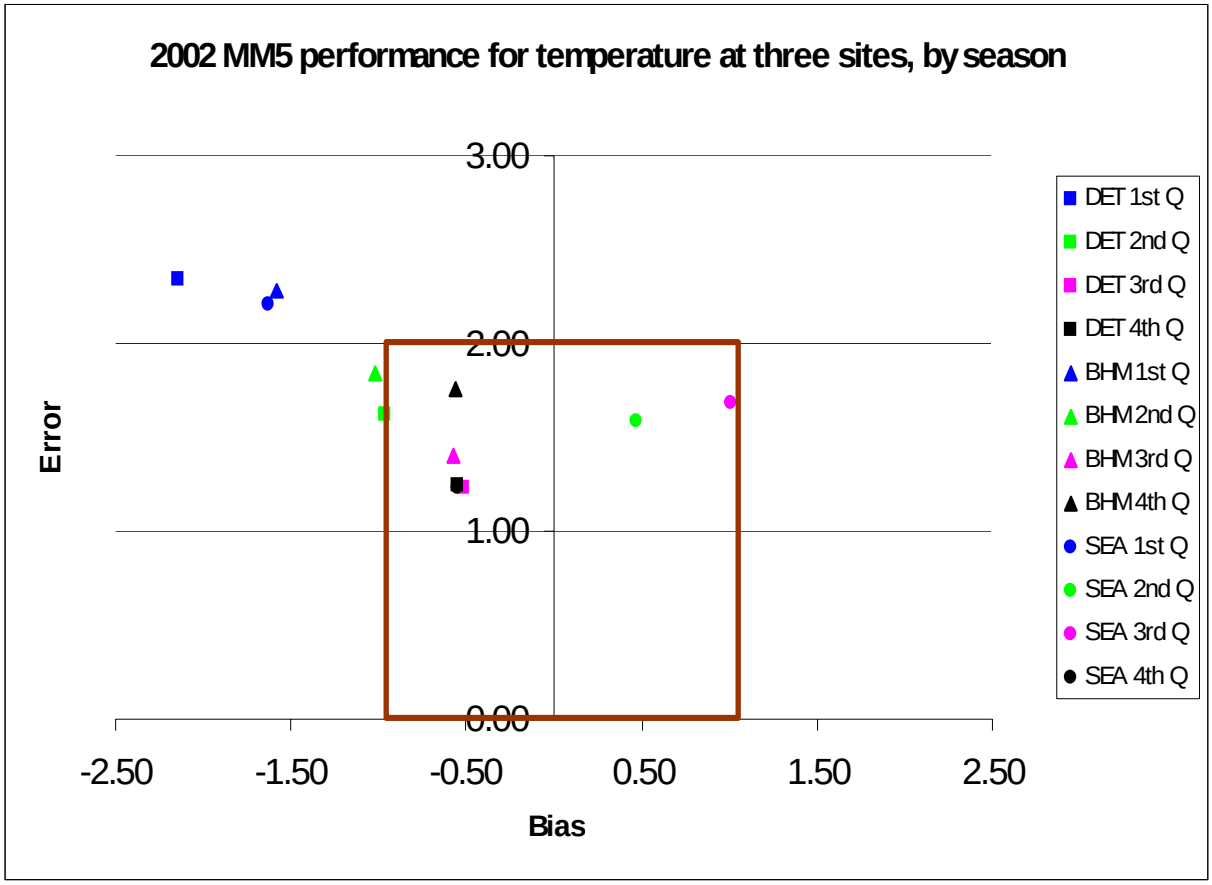


Figure 16.2: Sample "soccer plot" summary of temperatures from a 2002 MM5 case. This particular analysis looked at seasonal temperature performance at three sites. The benchmarks in this case were set at +/- 1.0 K for bias and 2 K for error. This analysis called attention to a relatively large cold bias in the winter.

In most cases the performance evaluation will be completed on the raw meteorological fields. However it is also important to compare the results before and after the meteorological post-processing to ensure that the meteorological fields going into the air quality model have not been adversely affected.

Phenomenological Evaluation: As discussed in Chapter 11, it is recommended that a conceptual description of the area's air quality problem be developed prior to the initiation of any air modeling study. Within the conceptual description of a particular modeling exercise, it is recommended that the specific meteorological parameters[106] that influence air quality be identified and qualitatively ranked in importance. When evaluating meteorological models or any other source of meteorological data, the focus of the phenomenological evaluation should be on those specific meteorological phenomena that are thought to strongly affect air pollution formation and transport within the scope of a specific analysis. It is expected that this event-oriented evaluation will need to summarize model performance in terms of statistical metrics such as probability of detection and false alarm rate. As an example of a potential phenomenological analysis, many regional air quality modeling exercises attempt to assess the effects of transport of emissions from one area to a downwind area with an intent to establish source-receptor relationships. For these types of modeling analyses, accurate transport wind trajectories are needed to properly establish these source-receptor linkages. In this type of model application, a useful event-based meteorological evaluation would be to compare model-derived trajectories versus those based on ambient data to determine what error distance can be associated with the model fields.

[106] Possible examples include: lake/sea breezes, low-level jets, or amount of convection on a given day.

17.0 How Are the Emission Inputs Developed?

Air quality modeling for 8-hour ozone, $PM_{2.5}$, and regional haze requires emission inputs for base case, baseline, and future modeling years. As explained in the EPA Emission Inventory Guidance (U.S. EPA, 2005e), 2002 is designated as a new base year for 8-hour ozone and $PM_{2.5}$ SIPs and regional haze plans; therefore, wherever possible, 2002 should be used for baseline modeling[36]. The use of an alternative baseline year should be discussed with the appropriate EPA regional office. The future-year depends on the nonattainment classification of the individual State or Tribe, as described in Section 3.6. Note that emissions data should be consistent with the data used in the modeled attainment test, described in Sections 3-6.

Preparation of emissions data for air quality models for the base and future years requires several steps. First, States and Tribes need to compile base-year inventories for their modeling region (e.g., the States and Tribes in the modeling grid). For ozone model applications, emission inventories should include a complete accounting of anthropogenic and biogenic VOC (speciated), NOx, and CO**.** For PM and regional haze model applications, emission inventories should also include emissions of SO2, NH3, $PM_{2.5}$ (speciated), and PM coarse (PMC)**.** Second, modelers must collect "ancillary data" associated with the inventories, which prescribes the spatial, temporal, and chemical speciation information about the emission inventory. Third, modelers use the ancillary data for "emissions modeling". Emissions models spatially, temporally, chemically, and vertically allocate emission inventories to the resolution needed by AQMs. Fourth, modelers must collect data on growth rates and existing control programs for use in projecting the base year emission inventories to the future year, and then use an emissions model to prepare that future year inventory data for input to the air quality model. Fifth, emissions inventories that reflect the emissions reductions needed for attainment will have to be prepared for air quality modeling.

Sections 17.1 and 17.2 summarize the issues for preparing emission inventories. Section 17.3 describes the needs for ancillary data. Section 17.4 summarizes the emissions modeling steps. Section 17.5 and Section 17.6 summarizes the issues associated with modeling of future year emissions data. Finally, Section 17.7 provides guidance on emissions to be used for local area primary $PM_{2.5}$ modeling.

17.1 What Emission Inventory Data are Needed to Support Air Quality Models?

Air quality models require hourly emissions for each grid cell and model layer in the modeling domain. Depending on the application, models may be run for episodic periods or for

---

[36] 2002 is the recommended inventory year for the baseline modeling (the starting point for future year projections). Other years may be modeled for the base case modeling (for performance evaluation) if episodes are chosen from years other than 2002. If all of the episodes are from 2002, then the base case and baseline inventories will be for the same year. See Section 14.1 for more details on selecting modeling time periods.

up to a full year or multiple years. The emissions need to be allocated to the chemical species required for the chemical mechanism and version of the model being used. The speciated emissions of VOC and $PM_{2.5}$ depend on the air quality model and chemical mechanism being used. For primary $PM_{2.5}$ emissions, the models typically divide the PM into (at a minimum) organic carbon, elemental carbon, primary sulfate, primary nitrate, and "all other" $PM_{2.5}$ emissions. Calculating the speciated emissions is part of the emission modeling step, described in Section 17.4 based on data described in Section 17.3.

Emission inventory data from five general categories are needed to support air quality modeling: stationary point-source emissions, stationary area-source emissions (also called non-point), mobile emissions for on-road sources, mobile emissions for nonroad sources (including aircraft, railroad, and marine vessels), and biogenic/geogenic emissions. The emission inventory development and emissions modeling steps can be different for each of these categories.

Point Sources- Point source inventories for modeling should be compiled at a minimum by country, State/Tribe, county, facility, stack, and source category code (SCC) but often are further subdivided by "point" and "segment"(see references below to point-source inventory development). The point source data must include information on the location of sources (e.g., latitude/longitude coordinates); stack parameters (stack diameter and height, exit gas temperature and velocity); and operating schedules (e.g., monthly, day-of-week, and diurnal).

Stationary Area Sources- Stationary-area source emissions data should be compiled by country, State/Tribe, county, and SCC. The most current stationary source emissions methods can be found at http://www.epa.gov/ttn/chief/.

On-Road Mobile- On-road mobile source emissions should be estimated using the most current version of the U.S. EPA MOBILE model (http://www.epa.gov/omswww/m6.htm), and for California, the most current version of EMFAC (http://www.arb.ca.gov/msei/msei.htm) in concert with vehicle miles traveled (VMT) data representative of the time periods modeled. The MOBILE model allows modelers to override default settings to obtain a local-specific on-road inventory, and modelers should consider using these options to improve their inventories. On-road emissions and VMT should be compiled at least at the country, State/Tribe, county, and SCC level, though modelers may optionally compile and use data for individual road segments (called "links"). The link approaches requires starting and ending coordinates for each road link. As noted in Section 17.3, link-based emissions can be based on Travel Demand Models (TDMs).

Nonroad Mobile- For nonroad mobile sources, the emissions should be compiled as country, state/tribe, county and SCC totals. For activities other than aircraft, locomotives, and marine, inventory developers should use the most current version of EPA's NONROAD model (http://www.epa.gov/oms/nonrdmdl.htm) or alternative model. The California OFFROAD model should be preferentially used for emissions in California, available at http://www.arb.ca.gov/msei/msei.htm. Since the NONROAD model does not include emissions of aircraft, locomotives, and commercial marine vessels, these emissions should be estimated using other approaches described as follows:

- Aircraft emissions should ideally be calculated based on airport-specific landing and takeoff activity data and the Federal Aviation Administrations (FAA) Emissions Data and Dispersion Model (EDMS), available at http://www.faa.gov/about/office_org/headquarters_offices/aep/models/edms_model/. Although aircraft emissions may be at the county and SCC level, point (latitude/longitude) or shape file spatial resolution may be desirable (but possibly more challenging to use in emissions models). County emissions of aircraft at airports can be allocated to point locations using emissions modeling features, such as the SMOKE area-to-point feature.
- Railroad emissions should ideally be calculated at the county and SCC level using fuel usage activity data distributed among train types (e.g., long haul, yard, passenger) and rail ton/miles data.
- Commercial marine vessel emissions should ideally be calculated at the county and SCC level by fuel/cargo activity, or by growth from previously determined inventory efforts. The emissions are distributed to ports and cargo lanes by activity data. County emissions of port-based emissions can be allocated to point locations using emissions modeling features, such as the SMOKE area-to-point feature.

The most current mobile source emissions methods can be found at http://www.epa.gov/ttn/chief/.

Biogenic Emissions - Biogenic emissions from plants and soil contribute VOC (including terpenes that are imported for PM and regional haze modeling), NOx, and CO emissions, which are best calculated as part of the emissions modeling step described in Section 17.4. These emissions require land use data, which is described as an ancillary dataset in Section 17.3. Geogenic emissions (such as volcanic emissions) are often not relevant for many areas, but if geogenic sources in the modeling domain are expected to contribute in a significant way to air quality problems, they should be included.

Other Considerations - For all sectors, emissions data must be compiled at a minimum as an annual total for the base modeling year. Emissions can also be compiled as monthly total emissions or (for ozone) an average summer day inventory. In any case, the temporal allocation step during emissions modeling (see Sections 17.3 and 17.4) must adjust the inventory resolution for the modeling time period. Additionally, we encourage the use of more temporally specific data where it is available and can be expected to improve model performance. For example, hour-specific Continuous Emissions Monitoring (CEM) data may be used as a source for hour-specific NOx and SO_2 emissions and exit gas flow rates[37], and the hourly heat-input CEM data

[37]Day specific emissions data, such as CEM data or wildfire data may be useful in the base case modeling to improve model performance. But in some cases, it may not be appropriate to project day-specific data to the future. For example, if a large power plant was shutdown for maintenance during the base case period, it would not be logical to project zero emissions for that source in the future (if it is assumed that the plant will be operating in the future year).

can be used to allocate emissions of other pollutants. CEM data are provided by EPA's Clean Air Markets Division in a format suitable for use in emissions processing at http://cfpub.epa.gov/gdm/index.cfm?fuseaction=emissions.prepacksmoke.

Additionally, inventory developers are encouraged to focus their emissions data collection, quality assurance, and emissions modeling efforts on the emissions sources that will have the greatest impact on their modeling results. In other words, inventory developers should devote the greatest attention to the sources that are most important: those in the immediate nonattainment areas, or large sources that may be responsible for transport into the nonattainment areas. Such a hierarchical approach to inventory development is described in (Smith,1998).

Inventories should be built using the most current, accurate, and practical methods available. Several references are available for guidance on building emission inventories. The first is the "Emissions Inventory Guidance for Implementation of Ozone and Particulate Matter NAAQS and Regional Haze Regulations" (U.S. EPA, 2005e). Additionally, modelers may want to consider EPA's approaches used for developing the 2002 NEI. Available NEI documentation may be used to help guide the development of the modeling inventory (http://www.epa.gov/ttn/chief/net/2002inventory.html). Lastly, 10 volumes (many with several chapters comprised of separate documents) have been issued by the Emission Inventory Improvement Program (EIIP) for use in inventory development. Nine of these volumes are relevant to creating emission inventories for ozone, PM, and regional haze modeling (U.S. EPA, 2004h):

- Volume I: Introduction
- Volume II: Point Sources
- Volume III: Area Sources and Area Source Method Abstracts
- Volume IV: Mobile Sources
- Volume V: Biogenics Sources
- Volume VI: Quality Assurance/Quality Control
- Volume VII: Data Management Procedures
- Volume IX: Particulate Emissions
- Volume X: Emission Projections

The EIIP documents are available electronically through the EPA website at http://www.epa.gov/ttn/chief/eiip/techreport/. The quality assurance procedures contain essential steps for inventory preparation, which help assure that the emission inventory is appropriate for SIP air quality modeling.

Therefore, before projecting the baseline to the future, it may be necessary to remove certain day-specific inventory information and replace it with average data. This issue is not a concern for day-specific mobile source or biogenic emissions data which are dependent on day specific (or even hourly) meteorological data for the time periods modeled.

17.2 Can The National Emission Inventory Be Used As a Starting Point?

We recommend that States and Tribes start with available inventories suitable for air quality modeling of the selected time period(s). If no such inventory is available, States/Tribes may derive an inventory suitable for use with models starting from the National Emission inventory (NEI), available from http://www.epa.gov/ttn/chief/net/. The 2002 NEI can be used as a starting for inventory development. However, the detail and accuracy of the emissions in the NEI may not be sufficient for use in attainment demonstration modeling. NEI data may be sufficient for estimating emissions in States that are hundred of kilometers downwind of the nonattainment area, but may not be sufficient for estimating local emissions that are critically important in estimating ozone or PM in the nonattainment area. Thus, States/Tribes should review the contents of the NEI for accuracy and completeness for regional and local scale modeling and amend the data where it is insufficient to meet the needs of the local air quality model application. While some benefits can be realized for states using the same inventory (such as the NEI) as a starting point for their work, the more important goal for SIP inventory development is to ensure that the data are representative and accurate enough for making good control strategy decisions.

17.3 What Other Data are Needed to Support Emissions Modeling?

The emission inventories must be converted (through emissions modeling) from their original resolution (e.g., database records) to input files for air quality models. These input files generally require emissions to be specified by model grid cell, hour, and model chemical species. This section describes the ancillary data that modelers should collect that allow emissions models to convert the emission inventory data.

Ancillary data for temporal allocation are necessary for stationary point, stationary area, and all mobile sources. To facilitate temporal allocation of the emissions, factors (called profiles) must be created to convert annual emissions to specific months (monthly profiles), average-day emissions to a specific day of the week (weekly profiles), and daily emissions to hours of the day (hourly profiles). Additionally, a cross-reference file is needed to assign the temporal profiles to the inventory records by SCC, facility, or some other inventory characteristics. Where available, the operating information that may be available from the point-source inventory should be used to create inventory-specific temporal factors. EPA provides a starting point for the temporal profiles and cross-reference files, available at: http://www.epa.gov/ttn/chief/emch/temporal/.

For point sources, hourly Continuous Emissions Monitoring data are recommended for use in model-evaluation runs. For future-year runs, we recommend creating an “average-year”or “typical year” temporal allocation approach that creates representative emissions for the “baseline inventory” but that also includes similar daily temporal variability as could be expected for any given year. Care should be taken to not reduce or increase day-to-day

variability in the averaging approach, with the exception of eliminating year-specific outages or other year-specific anomalies[38] within the years used for the model-attainment test.

The emissions models also need information about the chemical species of the VOC and $PM_{2.5}$ emissions for stationary point, stationary area, and all mobile sources. These data are used to disaggregate the total VOC and $PM_{2.5}$ emissions to the chemical species expected by the air quality model and are called speciation "factors" or "profiles". EPA provides a starting point for the VOC and $PM_{2.5}$ speciation data, which are available at: http://www.epa.gov/ttn/chief/emch/speciation/. For large or critical VOC and $PM_{2.5}$ sources in the modeling domain, States/Tribes should consider determining the individual chemical compounds contributing to the total VOC and $PM_{2.5}$. If collected, this information should then be used to compile improved speciation profiles for the critical facilities or source categories. These speciation profiles should be assigned to the inventory by a speciation cross-reference file, which also needs to be created or updated from the available defaults. The cross-reference files typically assign speciation profiles based on SCC code, though facility-specific assignments for point source code is also possible if plant-specific data are available.

For all source sectors that are compiled at a county resolution, the emissions models also need information about allocating the countywide emissions to individual grid cells that intersect the county. Such sectors include stationary area, nonroad mobile, and non-link on-road mobile sources. The spatial allocation process assigns fractions of county-total emissions to the model's grid cells intersecting the county based on a "surrogate" data type (e.g., population or housing data). The appropriate types of surrogate data to use for each SCC in the inventories should be identified for this processing step. Spatial surrogates can be created using Geographic Information Systems (GISs) to calculate the fraction of countywide emissions to allocate to each grid cell based on the surrogate type. These calculations can also be made using EPA's Surrogate Tool (http://www.cep.unc.edu/empd/projects/mims/spatial/srgtool/), which is based on the Multimedia Integrated Modeling System (MIMS) Spatial Allocator (http://www.cep.unc.edu/empd/projects/mims/spatial/). Additionally, all SCCs needing spatial surrogates should be assigned a surrogate in a cross-reference file. Point sources do not need spatial surrogates, since the emissions models assign the grid location based on the latitude and longitude of the point sources. Additionally, EPA provides spatial surrogates and cross-references for a limited set of modeling grids and surrogate types at: http://www.epa.gov/ttn/chief/emch/spatial/. Finally, for future-year modeling, emissions developers can choose to change their spatial surrogate data based on predicted changes in land

---

[38]The baseline inventory should contain emissions estimates that are consistent with emissions patterns which might be expected in the future year(s). If unplanned outages or shutdowns at point sources occurred in the base year, then it may not be appropriate to project those emissions to the future. Using a typical or average year baseline inventory provides an appropriate platform for comparisons between the base year and future years. However, certain emissions strategies may depend on capturing temporal emissions patterns on days with peak ambient concentrations. This should be considered when creating typical or average year inventories.

use patterns, population growth, and demographics, however, the impact and utility of such approaches is not well characterized[39].

For biogenic emissions modeling, the Biogenic Emission Inventory System, version 3 (BEIS3) (http://www.epa.gov/asmdnerl/biogen.html) model comes with all needed ancillary data, except the gridded land-use data and meteorology data for a specific air quality modeling domain and grid. Other biogenic models are acceptable as well, provided that the users have demonstrated their suitability for the modeling application. Land use and meteorology data that are compatible with BEIS3 are needed for the specific grid and grid-cell resolution that is being used. For BEIS3, land use data can be created with the MIMS Spatial Allocator (http://www.cep.unc.edu/empd/projects/mims/spatial/), using raw Biogenic Emissions Landcover Data (BELD), version 3 (http://www.epa.gov/ttn/chief/emch/biogenic/). For future-year modeling, emissions developers can choose to change their land cover data based on predicted changes in land use patterns, however, the impact and utility of such approaches is not well characterized. The impact and model characterization of biogenic terpene and sequiterpene on secondary aerosol format for PM modeling is currently evolving. Emissions modelers should consider the latest available information on this issue where biogenic sources are key sources in their modeling region.

On-road emissions for fine-scale model grids (e.g., 4-km grid cells or smaller) may be based on a link-based mentioned in Section 17.2. The VMT and speed data needed for a link-based approach can be provided using a Travel Demand Model (TDM). These models require their own sets of inputs, which depend on the specific TDM used. Details on using TDMs for link-based on-road mobile emissions are available from the EIIP document "Use of Locality-Specific Transportation Data for the Development of Source Emission Inventories" (http://www.epa.gov/ttn/chief/eiip/techreport/volume04/iv02.pdf).

Emissions models have other input files that must be created. For example, criteria may be needed for selecting elevated from non-elevated point sources. Each model has a large number of files and settings which work together in fairly complex ways; therefore, care must be taken to determine the files needed for the emissions model, and to prepare all needed input files in a way that will support using the emissions model for the specific air quality modeling episode.

17.4 How Are Inventory Data Converted Into Air Quality Model Input?

Emissions models are used to convert inventory data to inputs for air quality modeling. As described in Section 17.3, additional ancillary data is needed to complete the process. The emissions data from each of the five emissions sectors are temporally allocated, chemically speciated, and spatially allocated. The resulting hourly, gridded, and speciated emissions from

---

[39]At the time this document was written, tools to readily predict future-year land use patterns are not readily available in a form for use in emissions modeling.

all sectors are then combined before being used by an air quality model. In this section, we will provide information on several emissions models and summarize some additional issues that are key for emissions modeling.

Emissions models Several emissions models are available for use in SIPs. While no single model has been specifically created for all situations of SIP development, each model is generally capable of performing the temporal, chemical, and spatial allocation steps as well as various other steps. Users of these models are responsible for ensuring that the emissions processing steps are transforming the emission inventories as intended and are not changing the emissions in any unexpected way. The models each have different capabilities, limitations, and nuances. Therefore, when choosing an emissions model, it is worthwhile to discuss the choice with the developers of these systems and/or with EPA to establish which model is best for the specific application.

Currently there are three main emissions models being used to process emissions for input into photochemical grid models and a fourth model that is under construction at the time that this document was written. They are: Sparse Matrix Operator Kernel Emissions (SMOKE); Emissions Modeling System (EMS-2001); and Emissions Preprocessor System - Version 2.5 (EPS 2.5) and version 3.0 (EPS-3). The Consolidated Community Emissions Processing Tool (CONCEPT) is under development by the Regional Planning Organizations.

The Sparse Matrix Operator Kernel Emissions (SMOKE), software and User's Guide are available through the University of North Carolina, Carolina Environmental Program (http://www.cep.unc.edu/empd/products/smoke). SMOKE supports processing of criteria, mercury, and toxics inventories for stationary point, stationary area, mobile, and biogenic emissions. It can create input files for the CMAQ, CAM_X, UAM-V, and REMSAD air quality models. SMOKE was the basis for development of the BEIS3 system, so BEIS3 in its original form can be used easily with SMOKE. Applications of SMOKE have been presented at several of the International Emissions Inventory Workshops (Houyoux, 2000; Strum, 2003). SMOKE is available for UNIX and Linux operating systems and is not recommended for use on a PC. It does not require third party software. It does not include utilities for creating speciation profiles, biogenic land use, or spatial surrogates, though the latter two datasets can be built using the Multimedia Integrated Modeling System (MIMS) Spatial Allocator Tool (http://www.cep.unc.edu/empd/projects/mims/spatial/). Support for the SMOKE system is available through the Community Modeling and Analysis System (CMAS) help desk (http://www.cmascenter.org/html/help.html).

The Emissions Modeling System, (EMS 2001, http://www.ladco.org/tech/emis/ems_2001/) is a later version of EMS-95, which was used in the modeling underlying the U.S. EPA NOx SIP call rule to reduce regional NOx emissions (U.S. EPA 1998b), as well as in other applications of nested regional air quality models. It can create inputs for the CAM_X and UAM-V models. It includes the BIOME3 model, which provides access to similar calculations of biogenic emissions as are available in the BEIS3 system. EMS-2001 can be run on either Linux or Windows NT, and users must purchase a license for the SAS®

software to use it. It includes utilities for creating speciation profiles, biogenic land use, and spatial surrogates. An updated version has new spatial processors, which limit the need for GIS software.

The Emissions Preprocessor System - Version 2.5 (EPS-2.5), software and User's Guide are available through Systems Applications International/ICF Consulting (www.uamv.com). EPS-2.5 is a comprehensive emissions processing system that supports processing of stationary point, stationary area, and mobile emissions for the development of base and future-year modeling inventories for input to the CAM_X, UAM-V, and REMSAD models. EPS-2.5 consists of a set of stand-alone FORTRAN programs that do not require third-party software. The system is capable of preparing local, regional, and continental-scale emission inventories for ozone, particulate matter, mercury, and air toxics modeling applications. EPS 2.5 is available for UNIX and Linux operating systems. It includes utilities for creating source-specific temporal and chemical speciation profiles based on locally provided detailed information for episode specific emission inventories. It also includes utilities for preparing spatial surrogates. In addition, EPS-2.5 has the capability of creating modeling inventories required for the application of source apportionment techniques such as UAM-V's Ozone and Precursor Tagging Methodology (OPTM). A newer version of EPS is also available, called EPS-3.

The Consolidated Community Emissions Processing Tool (CONCEPT) is an open source model written primarily in PostgreSQL. Users are encouraged to make additions and enhancements to the model. The database structure of the model makes the system easy to understand, and the modeling codes themselves are documented to encourage user participation in customizing the system for specific modeling requirements. The CONCEPT model structure and implementation allows for multiple levels of QA analysis during every step of the emissions calculation process. Using the database structures, an emissions modeler can easily trace a process or facility and review the calculation procedures and assumptions for any emissions value. CONCEPT includes modules for the major emissions source categories: area source, point source, on-road motor vehicles, non-road motor vehicles and biogenic emissions, as well as a number of supporting modules, including spatial allocation factor development, speciation profile development, growth and control for point and area sources, and CEM point source emissions handling. Additional work by the emissions modeling community has begun development of CEM preprocessing software, graphical QA tools, and an interface to the traffic demand models for on-road motor vehicle emissions estimation.

Biogenic emissions Estimates for biogenic emissions can be made using the BEIS emissions model (Geron, et al., 1994). The BEIS3 model estimates CO, NOx, and VOC emissions from vegetation and soils in the gridded, hourly, and model-species forms needed for air quality modeling. Guenther, et al., (2000) contains the technical development and review of BEIS3. Vukovich, et al., (2002) summarizes new input data and algorithms as implemented within SMOKE. Arunachalam, et al., (2002) presents the impact of BEIS3 emissions on ozone. For more detailed local estimates, a State or Tribe should review the biogenic emissions on a county basis and update as needed the spatial patterns of land use data. BEIS3.13 was the latest version of BEIS3 in use by EPA at the time this document was written.

Other models for biogenic emissions include the BIOME model that is a part of EMS-2001, the Global Biosphere Emissions and Interactions System (GloBEIS), and the Model of Emissions of Gases and Aerosols from Nature (MEGAN). GloBEIS estimates emissions from natural (biogenic) sources and is designed for use in combination with photochemical modeling systems, such as CAMx (http://www.globeis.com/). The MEGAN model is available from http://bai.acd.ucar.edu/Megan/.

All of the biogenic models require a mix of land uses to be specified for each county or grid cell, as well as hourly temperature and in some cases other meteorological data. If a State or Tribe 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.

Agricultural Ammonia Emissions- Ammonia emissions from agricultural operations are typically estimated using standalone models. A commonly used model to estimate ammonia is the Carnegie Mellon University Ammonia Model (http://www.cmu.edu/ammonia/). A promising model that is under development is the CONCEPT Farm Model (http://www.conceptmodel.org/nh3/nh3_index.html).

17.5 Are there Other Emissions Modeling Issues?

In addition to the general emissions modeling steps and the biogenic emissions modeling, there are several other issues of which the air quality modeler should be aware. These are:

- Elevated point sources
- Advanced approaches for mobile source modeling
- Quality assurance

In the remainder of this section, we briefly address each of these issues.

Elevated Point Sources Point sources need to be assigned to an appropriate model layer[40] (the vertical modeling dimension). Depending on the air quality model that is being used, different emissions modeling steps can be taken. Models such as UAM-V and CAM_X, expect input emissions files separately for layer-1 emissions and elevated point-source emissions. Additionally, elevated point sources may be flagged for treatment with a plume-in-grid (PinG) approach. For these models, emissions modelers must supply a criteria for specifying which point sources will be treated as elevated and as PinG sources. In this case, the air quality model calculates the plume rise of the point source emissions when the model is run.

---

[40]Point sources generally comprise most of the elevated emissions (above layer 1), although other area sources, such as fires and aircraft, may also have emissions assigned to elevated layers.

Models such as CMAQ expect (1) a 3–D emissions input file, for which the elevated plume rise has already been calculated and (2) a separate optional PinG emissions file. Emissions modelers may optionally specify which sources to treat as elevated and PinG. Since the emissions model must calculate plume rise in advance, it must use meteorological data such as temperature, pressure and wind speeds. For the sake of consistency, the meteorological data that is used in calculating plume rise is usually the same as what is used by the air quality model.

Mobile Source Modeling Mobile source emissions modeling takes two primary approaches. The first approach is to compute emissions from VMT and emission factors from the MOBILE model prior to use by an emissions model. The second approach is to allow an emissions model, such as SMOKE, to drive the MOBILE model using gridded meteorology data. Many more assumptions must be made about using average temperatures in the first approach, since emissions are calculated on an annual total, monthly, or average-day basis and therefore do not include the day-to-day temperature variability that the second approach includes. It is widely assumed that the second approach is more robust for local-scale modeling, though we do not recommend one approach over the other. States/Tribes are encouraged to choose an approach that gives sufficient model performance for their attainment demonstration modeling.

Quality Assurance The third additional emissions modeling topic we have summarized here is emissions modeling quality assurance (QA). A brief synopsis of appropriate quality assurance (QA) approaches for emissions modeling is available in Section 2.19 of the SMOKE manual (http://cf.unc.edu/cep/empd/products/smoke/version2.1/html/ch02s19.html). The purpose of QA for emissions modeling is to ensure that the inventories are correctly processed using the information the modeler intended. (It is assumed here that the inventory itself has already been QA'd through inventory QA procedures, as referenced in Section 17.3.) Emissions modeling QA includes such activities as reviewing log files for errors and warnings and addressing problems; comparing emissions between each of the processing stages (e.g., data import, speciation, temporal allocation) to ensure mass is conserved; checking that the correct speciation, temporal allocation, and spatial allocation factors were applied; ensuring that emissions going into the air quality model are consistent with expected results; and checking modeling-specific parameters such as stack parameters. In addition, the process of running emission inventories through emissions models and air quality models often provides insights into the emission inventories. These insights can lead to inventory changes that improve the quality of inventories for additional modeling iterations.

In general, this guidance also encourages the use of graphical analysis and Geographic Information Systems (GIS) for improved QA of emissions data and processing. A commonly used analysis tool for model-ready emissions data is the Package for Analysis and Visualization of Environmental Data (PAVE), available at http://www.cep.unc.edu/empd/EDSS/pave_doc/index.shtml.

Lastly, at the time this document was written, the EPA is developing the Emissions Modeling Framework (EMF), which includes extensive data management and quality assurance

features intended to assist emissions modelers with the enhanced quality of their data. This tool will be available for use in 2007 for support of ozone, $PM_{2.5}$, and regional haze SIPs .

17.6 How Are Emissions Estimated for Future Years?

Emissions estimates for future years are called "emissions projections". These projections include both emissions growth (due to increased or decreased activities) and emissions controls (due to regulations that reduce emissions in specific ways in the future). The goal in making projections is to obtain a reasonable estimate of future-year emissions that accounts for the key variables that will affect future emissions. Each State/Tribe is encouraged to incorporate in its analysis the variables that have historically been shown to drive its economy and emissions, as well as the changes in growth patterns and regulations that are expected to take place between the time of their base year and future years. For details on which future year(s) should be modeled for attainment demonstrations, refer to Section 3.6.

A report entitled "Procedures For Preparing Emissions Projections" describes emissions projections issues and approaches (U.S. EPA, 1991b; also available at www.epa.gov/ttn/chief/eiip/techreport/volume10/x01.pdf). It is currently the most comprehensive resource available about emissions projections. In this section, we will address an overall approach for tackling emissions projection issues that incorporates use of this document and points out where other approaches or new data are available to amend the information included in that report.

Developers of projection-year emissions are advised to take the steps in the bulleted list below. Each of these steps corresponds to a subsequent subsection.

- Identify sectors of the inventory that require projections and sectors for which projections are not advisable, and prioritize these sectors based on their expected impact on the modeling region (Section 17.6.1).

- Collect the available data and models that can be used to project emissions for each of the sectors (Section 17.6.2).

- For key sectors, determine what information will impact the projection results the most, and ensure that the data values reflect conditions or expectations of the modeling region (Section 17.6.3)

- Create inputs needed for emissions models, create future year inventories, quality assure them, and create air quality model inputs from them (Section 17.6.4)

The remainder of this section provides additional details about these steps.

17.6.1 Identify and prioritize key sectors and regions for projections

Emissions modelers should evaluate their inventories for those sectors and regions that are most critical to their modeling results. The purpose of identifying these key sectors is to direct more resources and efforts to estimating future-year emissions for these sectors in the subsequent steps.

Sectors can be subsets of the larger groups of stationary area, on-road mobile, nonroad mobile, and point sources. For ozone modeling, in regions that are NO_X-limited, sectors with higher NOX emissions will be more important for ensuring correct projections, and for VOC-limited regions, sectors with higher VOC emissions will be more important. Mobile sources are typically large contributors to both VOC and NOx and are therefore usually a priority in estimation of future-year emissions, particularly for urban areas. For PM and regional haze modeling, the relative importance of primary $PM_{2.5}$ emissions and precursors will depend on the area of the country. Primary $PM_{2.5}$ emissions may be important in urban areas. In the East, SO2 emissions may be of primary importance. And NOx, VOC, and NH3 will vary in importance depending on the particular situation.

In some cases, there are sectors that are very difficult or highly uncertain to project into the future. Wildfire emissions is a common example of such a sector, since models are not readily available that estimate the magnitude and location of future-year emissions. In these cases, it may be advisable to create an “average year” inventory and temporal distribution of the emissions.

Emissions modelers can consider the proximity of the emissions and the expectation of long-range transport when reviewing projection assumptions. Those emissions that have an impact on the nonattainment status (whether transported from outside areas or locally generated) should be given highest priority in evaluating the projection assumptions. For example, less priority can be given to emissions from distant states for pollutants that do not transport from those states to the nonattainment area.

17.6.2 Collect available data and models for each sector needing projections

The EIIP projections document (referenced previously) provides a starting point for this step. The EIIP document provides the majority of information about sources of data on growth and controls. This section supplements that document.

For each sector needing projections (and in particular for the priority sectors), emissions modelers should collect and review growth and control information. As noted in the EIIP document, the sources of such information depend upon the sector of interest. New data and approaches beyond those in the EIIP document are available and provided here; likewise, new data and approaches will become available subsequent to the publication of this document. The information provided here is a snapshot of information that will grow and adapt over time, and emissions modelers should seek out such new information as part of this step.

For sectors and regions with lower priority for consideration (as described in Section 17.6.1), emissions modelers can rely on available default data. For sectors and regions that are important to modeling the nonattainment area, emissions modelers should spend additional effort obtaining improved information from other modeling applications (such as modeling done by RPOs or other States) or developing the improved information themselves.

In addition to the sources of control information about non-EGU point and stationary area sources listed in the EIIP projection document, EPA has provides control assumptions with its modeling data used for promulgating air quality rules. For example, emissions growth and control assumptions were provided with the Clean Air Interstate Rule (See http://www.epa.gov/air/interstateairquality/pdfs/finaltech01.pdf for additional information on obtaining actual data files) More current information is usually available than what is available from EPA, which is best obtained by working with state and local regulatory officials for the SIP modeling region. The remainder of this section focuses on sources of growth information.

Point sources.

Section 2 of the EIIP projections document gives details about projection of point-source emissions. There are two major subsets of point sources: electric generating utilities (EGUs) and non-EGUs. The Clean Air Markets Division (CAMD) of the U.S. EPA uses the Integrated Planning Model (IPM) to model emissions trading programs that now dominate the prediction of future-year emissions from EGUs. More information on IPM is available at (http://www.epa.gov/airmarkt/epa-ipm/). Additionally, IPM-based emissions are posted by CAMD on EPA's website (http://www.epa.gov/airmarkets/epa-ipm/iaqr.html). Other trading models may exist and could be used for estimation of future-year emissions.

To prevent double-counting of emissions sources, emissions modelers who use IPM should be careful to match the sources of emissions from IPM with the base year emissions sources. This helps to ensure that the EGU part of their point source inventory is separated from the non-EGU part based on the facilities included in IPM. The facilities included in IPM are defined using the National Electric Energy Database System (NEEDS). The NEEDS dataset should be compared to the point inventory using the Office of Regulatory Information Systems (ORIS) Plant ID field from both. In some cases (e.g., co-generation facilities), only some of the units at these facilities are included in IPM; therefore, the separation of EGUs from non-EGUs should be done by unit and not by facility.

Since base-year emissions of EGU point sources can be based on hour-specific emissions from the CEM program, emissions modelers must choose a temporal allocation approach for estimated future-year emissions. Emissions modelers should choose an approach that is representative of future expected behavior and not limited to any single year's closures and maintenance schedule. Ideally, types of EGUs that are run during high demand conditions would have temporal allocation approaches that reflected those peaks, and units that are run continuously would be temporally allocated in a more uniform way. Analysis of several years of CEM data could be helpful to determine these unit-by-unit trends and develop profiles accordingly.

Additional considerations for projecting point source emissions without the IPM model are provided in the EIIP projection document. However, several of the references in that document are now out of date, as follows.

- The latest SCC codes are available at http://www.epa.gov/ttn/chief/codes/ and not the website provided in the document.
- The AP-42 emission factor website is now http://www.epa.gov/ttn/chief/ap42/index.html.
- The website with further information about point source inventories is http://www.epa.gov/ttn/chief/eiip/techreport/
- The references in Table 2.1-1 web link is no longer applicable. Instead, the following information can be used to help provide similar information:
  - Energy consumption by industry report, 2002: http://www.eia.doe.gov/oiaf/analysispaper/industry/consumption.html
  - Annual Energy outlook report, 2005: http://www.eia.doe.gov/oiaf//aeo/pdf/0383(2005).pdf

The references in Table 2.1-2 are also out of date and should be updated as follows:

- US EPA Emissions trends reports. These reports can be used to provide the historical trends of emissions sectors based on the National Emission Inventory: http://www.epa.gov/air/airtrends/
- Integrated Planning Model: http://www.epa.gov/airmarkt/epa-ipm/
- Multiple Projections System - is no longer in use
- California Emission Forecasting System: http://www.arb.ca.gov/html/databases.htm

The tools and data sources described in the "Additional resources" section below should also be evaluated for use for projection of point source emissions.

Stationary area sources.

Section 3 of the EIIP projections document provides details about projecting area-source (a.k.a nonpoint emissions)**.** However, several of the references or statements in that document are now out of date, as follows.

- The web sites for SCCs, AP-42 emission factors, and more information about area-source inventories are out of date. These should be updated to the same sites as were listed in the point sources section above.

- The ASEM model mentioned in Section 3.2 of the EIIP projection document was not completed and is not available.

- The references in Table 3.1-2 are out of date and should be updated as follows:
  - US EPA Emissions trends reports. These reports can be used to provide the historical trends of emissions sectors based on the National Emission Inventory:

http://www.epa.gov/air/airtrends/
- Multiple Projections System - is no longer in use
- California Emission Forecasting System:
http://www.arb.ca.gov/html/databases.htm

The tools and data sources described in the "Additional resources" section below should also be evaluated for use for projection of point source emissions.

On-road mobile sources.

Section 3 of the EIIP projections document provides details about projecting on-road mobile-sources**.** It is still a very relevant snapshot of on-road mobile projection approaches; however, several of the references or statements in that document are now out of date, as follows. Additionally, several new capabilities are available, which also are included in the list below.

- The newer versions of the MOBILE model support 28 vehicle types instead of the 8 types listed in this document, however, the NEI and other inventories often group these 28 types to the same 8 described in the document.
- The web sites for SCCs, AP-42 emission factors, and more information about area-source inventories are out of date. These should be updated to the same sites as were listed in the point sources section above.
- The EPA Office of Mobile Sources is now called the Office of Transportation and Air Quality (OTAQ). Its website is http://www.epa.gov/otaq/index.htm.
- The new website for the MOBILE model is http://www.epa.gov/otaq/mobile.htm.
- The PART5 model is no longer used, because it has been superceded by the MOBILE model, starting with version 6 (MOBILE6).
- The increased popularity of sport utility vehicles has continued, but may be affecting more than only the percentage of light duty gasoline vehicles. Additionally, since MOBILE6 includes 28 vehicle types, the increases in larger and heavier vehicles can be more carefully accounted for using MOBILE6 than the EIIP projection document mentions.
- Updates to the SMOKE model and other emissions models permit the temperature information used in MOBILE6 to come from the prognostic meteorological models (e.g., MM5 and MCIP) that are used to prepare inputs to air quality models.
- OTAQ has created the National Mobile Inventory Model, which drives both the MOBILE6 model and the NONROAD model to create base-year and future-year inventories. Additional information on this tool is available at: http://www.epa.gov/otaq/nmim.htm.
- The references in Table 4.1-1 are out of date and should be updated as follows:
  - MOBILE6 - is an operational model and can be used
    See http://www.epa.gov/otaq/mobile.htm
  - The PART5 model is no longer needed because MOBILE6 replaces it
  - The California mobile sources emissions page is now:
    http://www.arb.ca.gov/msei/msei.htm
  - The EIIP mobile source document is now available at:
    http://www.epa.gov/ttn/chief/eiip/techreport/volume04/index.html

- Emission Inventory Guidance documentation is now available at:
http://www.epa.gov/ttn/chief/eidocs/eiguid/index.html

Additionally, the Economic Growth Analysis System (EGAS) contains VMT growth information and can be used to grow VMT by SCC. This system is further described in the "Additional resources" section below.

Nonroad mobile sources.

Section 4 of the EIIP projections document provides details about projecting nonroad mobile sources**.** Several of the references or statements in that document are now out of date, as follows. Additionally, several new capabilities are available, which also are included in the list below.

- As noted above for on-road sources, the Office of Mobile Sources has changed to the Office of Transportation and Air Quality and has a different web site.
- The EPA's NONROAD model has been completed and has been used in a variety of ways by EPA, states, and local agencies. The website for the NONROAD model is:
http://www.epa.gov/otaq/nonrdmdl.htm
- California's OFFROAD model website is now:
http://www.arb.ca.gov/msei/off-road/updates.htm
- OTAQ has created the National Mobile Inventory Model, which drives both the MOBILE6 model and the NONROAD model to create base-year and future-year inventories. Additional information on this tool is available at:
http://www.epa.gov/otaq/nmim.htm.
- For aircraft, locomotive, and marine emissions, the best source of data on the impact of national rules from EPA's Office of Transportation and Air Quality (OTAQ) is the Federal Register notices including these rules. Since each rule is different, there is no single guidance that can be made about how these should be incorporated. As EPA OAQPS incorporates these data into their work, it is possible in the future that it will be available on EPA's Emissions Modeling Clearinghouse at http://www.epa.gov/ttn/chief/emch/index.html.
- For aircraft emissions, the Federal Aviation Administration data can be used to assist with growth estimates. More information is provided in the "additional resources" section, below.

Additional resources.

In addition to sector-specific resources, there are other sources of information that can be used for more than a single sector as information to help determine the most applicable growth rates for the modeling region. These include the Economic Growth Analysis System (EGAS), Bureau of Economic Analysis data, Bureau of Labor Statistics "Employment Outlook", Census Bureau data, Trade Organizations and individual facilities, and the Chemical Economics Handbook. The following paragraphs provide a brief description and additional references for each of the sources of data just listed. Section 17.6.3 provides information about how to choose the best information for a given sector.

Multiple sources of information can be used to better inform emissions modelers. For example, if a state and/or industry has shown steady decreases in gross product over the past five years, that information could temper (or be used in place of) a projected 30% growth predicted from EGAS over the next 10 years. Such a discrepancy would simply highlight the fact that all models, including REMI or those used by DOE can produce results that are not consistent with other expectations. In addition, because of changes in technologies and fuel efficiency, economic growth is not necessarily a sufficient indicator of emissions growth. While this guidance does not provide a prescriptive approach for combining such information, the authors recognize that predicting the future of emissions is not an exact science and therefore multiple sources of information should be used to develop an informed approach.

EGAS. The latest EGAS model at this time is version 5 (http://www.epa.gov/ttn/ecas/egas5.htm) The default version of EGAS relies primarily on three sources of data: (1) state-specific economic data from the Regional Economic Model, Inc. (REMI) Policy Insight ® model (version 5.5) that includes population growth estimates, (2) Region-specific fuel-use projections from the U.S. Department of Energy (DOE) Annual Energy Outlook 2004, and (3) VMT projections included in the MOBILE6.0 model. EGAS outputs growth factors that do not include control information for a user-define base year and multiple future years through 2035. EGAS uses default cross-walks from these data sources to SIC, SCC and MACT codes. The DOE data are used to compute growth factors assigned by default to stationary source fuel combustion sectors, the VMT projections are used for on-road mobile, and the REMI data are used for the remaining sectors. Additionally, EGAS5 supports emissions modelers adding new data sources, changing assignments of data sources to SICs, SCCs, and MACT codes, and creating custom configurations of data that best represent the growth expected in their modeling region. The additional sources of information listed below can be input into EGAS to develop such custom scenarios. It should be noted that these sources have very little, if any, projection information. Instead, they provide historical data that could be extrapolated to produce projections or to derive projections using other methods.

EGAS results should be assigned to the inventory to give the best mapping of the raw data to the inventory. Generally, SIC is the preferred cross-walk for point sources and SCC is the preferred cross-walk for non-point sources. However, the DOE-based fuel consumption forecasts (assigned by SCC) can also be useful for some sectors that use fuel combustion.

Bureau of Economic Analysis. The Bureau of Economic Analysis (BEA) provides historical data about macroeconomic factors at the state level such as gross state product, personal income by industry, employment by industry, wages and earnings by industry, and population by state. The website to obtain state and/or state-industry data is http://www.bea.doc.gov/bea/regional/gsp/. Such information may be used to verify emissions growth forecasts in EGAS by comparing available historical BEA data (1997-2004) with the REMI economic activity data driving emissions in EGAS for the same set of years. A good use of the gross state product data is to verify growth in sectors/codes that use REMI output and value-added data. Additionally, employment and population statistics can be compared with sectors/codes using REMI employment and population data. The BEA also provides industry-

specific national information at http://www.bea.doc.gov/bea/dn2/home/annual_industry.htm. However, since these data are at the national level, the use of the data should be limited to comparisons with national simulations in EGAS.

Bureau of Labor Statistics. The Bureau of Labor Statistics (BLS) provides historical state-specific and national employment data. The BLS database is fairly large and provides more detail than needed for verifying the growth factors generated in EGAS using REMI employment data. However, BLS employment data is likely to be as accurate if not more accurate than employee data provided by BEA and US Census. The data are available at http://www.bls.gov/sae/home.htm**.**

Census Bureau. The US Census Bureau provides data on total employees, total payroll, number of establishments, and value of shipments by NAICS code (2 digit - 6 digit) for 1997 and 2002. This database is ideal for examining changes in the number of establishments from 1997 - 2002. However, it does not provide data from 1998-2001 so its trend information is limited to two "snapshots". Therefore, this data is useful primarily to evaluate the state/SCC and state/SIC combinations from EGAS that show zero emissions growth, which indicates an assumption by EGAS or its input data that those processes or industries do not exist in the state or region of interest. The census data are available at http://www.census.gov/http://www.census.gov/econ/census02/guide/geosumm.htm.

Federal Aviation Administration. The Federal Aviation Administration (FAA) provides aviation data users with summary historical and forecast statistics on passenger demand and aviation activity at U.S. airports. The summary level forecasts are based on individual airport projections. The Terminal Area Forecast (TAF) includes forecasts for active airports in the National Plan of Integrated Airport Systems (NPIAS). The TAF can be accessed at http://www.apo.data.faa.gov/. The TAF model allows users to create their own forecast scenarios. The TAF database, which contains a query data application, allows the public to access and print historical and forecast (up to 2025) aviation activity data by individual airport, state or FAA region.

Trade Organizations and individual facilities. Trade organizations can also be a helpful resource because they often have projections for future-year growth of their industry and may be aware of pending facility closures or new facility construction. This resources is most relevant for large industrial sources that usually are included in point source inventories. In addition, emissions modelers should consider contacting large sources of emissions in their modeling region for their expectations of emissions growth or reduction. In many cases, large industries may be willing to provide such information when presented with what will be used if no additional information can be provided. For these types of data sources, emissions modelers should be aware of possible conflicts of interest, ie. industry groups or facilities may feel an incentive to under-represent future-year emissions in hopes of avoiding additional requirements for reducing emissions; however, many industrial representations have been valuable stakeholders who are interested in providing accurate information.

Chemical Economics Handbook. The Chemical Economics Handbook, produced by Access International, Inc., is a series of reports on prices, production and consumption of hundreds of chemical industry products and commodities. Past and current information on chemical products and commodities is available, and projections of future prices, production and consumption are often available. Reports on specific industries are also available. Reports at an industry level can often be used to verify the efficacy of future industry modeling results. Each report is updated every 3 years. Projections, some up to 5 years from the current day, are often prepared using proprietary methods. Reports are available by subscription, and can be obtained as hard copy, CD, or through the Internet at http://www.sriconsulting.com/CEH/Public/index.html.

EPA base-year control assumptions. EPA OAQPS creates and uses data that quantify assumed base year controls on non-EGU point and stationary area sources. The most recent data published (at the time of this writing) are described with the Clean Air Interstate Rule (CAIR) Technical Support Document (TSD). The Emission Inventory chapter of this TSD contains a description of the base year controls assumed and how to get the data files from an FTP site; it is available online at http://www.epa.gov/air/interstateairquality/pdfs/finaltech01.pdf. Additionally, two possible future references for these types of data include (1) the Emissions Modeling Framework, which will include a Controls Programs Database and (2) the EPA OAQPS Emissions Modeling Clearinghouse, available at http://www.epa.gov/ttn/chief/emch/index.html.

17.6.3 Evaluate and refine data for projections in modeling region

For key sectors, emissions modelers should determine what information will impact the projection results the most, and ensure that the data values reflect conditions or expectations of the modeling region. The key information is identified based on a combination of base-year emissions data and growth and control rates. An iterative process is helpful to complete this step, as follows:

- Estimate future-year emissions by combining the base-year emissions with the projection information.

- Review future-year emissions data to identify large sectors of key pollutants (e.g. NOx, SO2, and VOC). Emissions summaries by state/SCC or state/SIC can be helpful for this step to consolidate the data to review.

- For the largest future-year emissions sectors, review the growth and control assumptions, compare among all available data sources, and/or collect additional information to refine the assumptions. A representative growth rate should be identified from the available data sources and all information known about the sources and sectors. Stakeholder review of the data can be helpful during this step; for example, an industrial facility with large projected emissions may be able to review the data and provide additional information for a more informed future-year emissions estimate.

- Additionally, emission modelers should also compare the future-year emissions to base-year emissions to identify overall growth rates for industrial sectors and reconsider excessively high growth rates, especially when associated with significant emissions.

- Using the new information, repeat step 1 and continue these steps until a satisfactory future-year projection across the entire inventory has been completed.

Emissions models (e.g., SMOKE, MOBILE6, NONROAD, etc.) provide the capability to create future-year inventories using base year inventories and projection information. Emissions modelers will need to convert the projection data into specific formats for input to these models. Prior to starting this step, emissions modelers should determine which emissions model will be used to perform the calculations and make sure that the type of information needed by the model is being collected.

17.6.4 Create future-year inventories and air quality model inputs

Using the final projection data determined in the previous step, create the final inputs to emissions models being used. Then, use the emissions model to create the future-year inventory. Other inputs to emissions models, such as spatial surrogates, speciation profiles, or temporal allocation factors may also need to be adjusted to reflect conditions expected in the future.

Once a future-year inventory and other data have been created, it must undergo the same steps as for the base-year modeling, such as temporal allocation, speciation, spatial allocation, elevated source selection, special on-road mobile processing, and quality assurance**.** Every attempt should be made to use consistent approaches between the future year and the base year for all of these modeling steps. Inconsistencies in approaches between the future-year modeling and the base-year modeling can lead to artificial differences in air quality modeling results that can affect conclusions. Therefore, it is critical to avoid such differences whenever possible.

17.7 Emissions Preparation for a Local Area Analysis

In section 5.3 we discuss modeling of local primary $PM_{2.5}$ sources with a dispersion model. The emissions for these sources should closely match those used in photochemical models. For example, every effort should be made to temporalize and speciate the primary emissions using available profiles. Information on input and formatting of the emissions for dispersion models can be found in the respective users' guides. Pre-processing of the primary emissions with an emissions model may be needed to derive the detailed information needed to temporalize and speciate the emissions.

In section 5.3 we recommend using "actual emissions" in the local area analysis, as compared to using "allowable emissions". 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. We define "actual emissions"as the best estimate of the actual emissions that occurred during the baseline period being modeled[41].

Modeling with allowable emissions is sometimes warranted. For example, for permit modeling, we generally compare the absolute predicted modeled concentrations against the NAAQS or the PSD increments. In these cases we want the modeling to be conservative. That is, we want to be sure that a maximum permitted level of emissions cannot cause of violation of the NAAQS or PSD increment. Therefore, in the case of permit modeling, it is sometimes appropriate to model with allowable emissions. But for a local area analysis, we are trying to determine the actual (or typical) contribution from sources to a monitor and we are using the model results in a relative sense. Therefore, it is only appropriate (and in fact conservative) to use actual emissions. Using actual emissions should lead to a reasonable estimate of of air quality improvements from reducing emissions at the flagged sources. Using allowable emissions might lead to an overestimate of benefits from emissions controls at flagged sources.

[41]In the emissions baseline modeling, the actual emissions may sometimes be replaced with "typical" emissions (e.g. EGU and fire emissions).

18.0 What are the Procedures for Evaluating Model Performance and What is the Role of Diagnostic Analyses?

The results of a model performance evaluation should be considered prior to using modeling to support an attainment demonstration or regional haze assessment. The performance of an air quality model can be evaluated in two ways: (1) how well is the model able to replicate observed concentrations of $PM_{2.5}$ components, ozone and/or precursors (surface and aloft), and (2) how accurate is the model in characterizing the sensitivity of ozone and/or $PM_{2.5}$ to changes in emissions? The first type of evaluation can be broadly classified as an "operational evaluation" while the second type of evaluation can be classified as a "diagnostic evaluation". The modeled attainment tests recommended in Sections 3-6 use models to predict the response of ozone and $PM_{2.5}$ to changes in emissions and then applies the resulting relative response factors to observed (rather than modeled) ambient data. Thus, while historically, most of the effort has focused on the operational evaluation, the relative attainment test makes the diagnostic evaluation even more important.

In addition to the model performance evaluation, diagnostic analyses are potentially useful to better understand whether or not the predictions are plausible. Diagnostic analyses may also be able to provide: (1) information which helps prioritize efforts to improve and refine model inputs, (2) insight into which control strategies may be the most effective for meeting the NAAQS, and (3) an indication of the "robustness" of a control strategy. That is, diagnostic tests may help determine whether the same conclusions would be reached regarding the adequacy of a strategy if alternative, plausible, assumptions were made in applying the model for the attainment test.

In this section, we first discuss the general concept of operational and diagnostic evaluations. We then identify and discuss issues related to the ambient measurements of ozone and $PM_{2.5}$. We then discuss methods which may be useful for evaluating model performance. It is recommended that performance be assessed by considering a variety of methods. The section continues by identifying potentially useful diagnostic tests which States/Tribes should consider at various stages of the modeling analysis to increase the confidence in the model predictions of future ozone and/or $PM_{2.5}$ levels. The section concludes with a discussion on interpreting the results of the operational and diagnostic evaluations.

18.1 How Do I Assess Model Performance And Make Use of Operational and Diagnostic Evaluations?

As noted above, model performance can be assessed in one of two broad ways: how accurately does the model predict observed concentrations for specific cases, and how accurately does the model predict *responses* of predicted air quality to changes in inputs (e.g. relative response factors)? Given existing data bases, nearly all analyses have addressed the first type of performance evaluation. 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. Typically, this type of operational evaluation is comprised principally of statistical assessments of model versus observed pairs. Operational evaluations are generally accompanied by graphical and other qualitative descriptions of the model's ability to replicate historical air quality patterns. The robustness of an operational evaluation is directly proportional to the amount and quality of the ambient data available for comparison.

The second type of model performance assessment, a diagnostic evaluation, can be made in several ways. One way to evaluate the response of the model is to examine predicted and observed ratios of "indicator species". Indicator species techniques have been developed for both ozone and secondary PM species (in particular nitrate) (Sillman, 1995 and 1998; Ansari and Pandis, 1998; Blanchard et al., 2000). If ratios of observed indicator species are very high or very low, they provide a sense of whether further ozone or secondary $PM_{2.5}$ production at the monitored location is likely to be limited by availability of NOx or VOC (or NH3). Agreement between paired observed and predicted ratios suggests a model may correctly predict the sensitivity of ozone or secondary $PM_{2.5}$ at the monitored locations to emission control strategies. Thus, the use of indicator species has the potential to evaluate models in a way which is most closely related to how they will be used in attainment demonstrations. A second way for assessing a model's performance in predicting the sensitivity of ozone or $PM_{2.5}$ species to changes in emissions is to perform a retrospective analysis. This involves comparing model predicted historical trends with observed trends. Retrospective 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 an attainment demonstration. More types of diagnostic analyses are provided in Section 18.5. We recommend that diagnostic analyses be performed during the initial phase of the model application and during any mid-course review.

18.2 What Pollutants Are of Interest for Model Performance?

There are a number of gas phase species that can be used to evaluate both ozone and secondary $PM_{2.5}$. Their presence may affect the response of ozone or secondary PM components to emissions strategies. Also, the model's performance in predicting certain gaseous species may provide diagnostic clues which help explain poor performance in predicting ozone or secondary $PM_{2.5}$. Gaseous pollutants of interest include:

- Ozone (O3)
- Nitric acid (HNO3)
- Nitric oxide (NO)
- Nitrogen dioxide (NO2)
- Peroxyacetyl nitrate (PAN)
- Volatile Organic Compounds (VOCs)
- Ammonia (NH3)
- NOy (sum of NOx and other oxidized compounds)
- Sulfur dioxide (SO2)

- Carbon monoxide (CO)
- Hydrogen peroxide (H2O2)

Some of the species listed above are currently available from either existing monitoring network systems and/or special studies. However, it is important to note that many of the species listed above are not routinely measured, (e.g. H2O2, HNO3, and NH3). 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 performed. 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.

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 chemical components that constitute $PM_{2.5}$. Components of $PM_{2.5}$ of interest include:

- Sulfate ion (SO4) and/or Sulfur (S)
- Nitrate ion (NO3)
- Ammonium ion (NH4)
- Elemental Carbon (EC)
- Organic Carbon (OC) and/or Organic Carbon Mass[42]
- Crustal (weighted average of the most abundant trace elements in ambient air)
- Mass of "other" primary particulate matter[43]
- Mass of individual constituents of inorganic PM

18.3 What Are the $PM_{2.5}$ Measurement versus Model Concentration Issues?

Ambient measurements are needed to perform a model evaluation. Historically, ozone model evaluations have used routinely measured ambient data (e.g. ozone) collected on a continuous basis. The measurements tend to be of high frequency and quality. Conversely, $PM_{2.5}$ and $PM_{2.5}$ component measurements are (mostly) filter based measurements which have relatively long sampling times (usually 24 hours) and aren't collected every day (usually once every three or

---

[42] For predicted/observed comparisons, organic carbon is preferred over organic carbon mass to allow for transparency between model predictions and observed measurements. Organic carbon mass may include hydrogen, oxygen, and other components.

[43] Other inorganic particulate matter can contribute up to 10% of $PM_{2.5}$ mass in certain locations. Unspeciated inorganic PM (sometimes referred to as "other PM") is often used interchangeably with "crustal" or "soil" PM. The term "soil" usually refers to the specific definition of soil from the IMPROVE program. "Other primary (inorganic) PM" includes heavy metals and other inorganic species in addition to those commonly found in soil.

six days). Speciated data play an especially important role, as the deterministic models predict exact chemical components which can be compared to some of the corresponding measured analytes. In addition, there are known positive and negative sampling artifacts as well as analysis artifacts associated with FRM and speciation monitors. For example, FRM monitors do not measure all of the $PM_{2.5}$ in the air, and the speciation samplers don't measure $PM_{2.5}$ in the same way as the FRM monitors (Frank, 2006). Due to the complex nature of the measurements, a basic understanding of the various $PM_{2.5}$ and speciation monitoring technology and measurements are needed before undertaking a model evaluation. This section attempts to address the range of measurements and measurement issues associated with $PM_{2.5}$ and its components.

Table 18.1 summarizes many of the $PM_{2.5}$ species measurements and networks and lists sampling time and measurement issues. This data will commonly be used as part of an operational evaluation.

Table 18.1 $PM_{2.5}$ Species Measurements by Monitoring Network and Related Sampling Time Issues.[44]

Network	Measurement	Notes	Sampling Time
FRM	$PM_{2.5}$	- Negative artifacts- nitrate, organic carbon - Positive artifacts- organic carbon, water	24-hour average 1/3 days or 1/6 days (some daily sites)
STN	$PM_{2.5}$	-STN total measured mass may not be reliable (cutpoint issues)	24-hour average 1/3 days or 1/6 days
	SO4		
	NO3	- Differences among sampler measurements across network. Different sampler types use different measurement techniques.	
	NH4	- Negative artifacts may occur on the Nylon filters due to volatilization of ammonium from the filter because the alkaline filter may not retain NH3 released in the decomposition of ammonium nitrate.	

[44] Most of the issues identified in Table 18.1 are discussed in (Frank, 2006).

Network	Measurement	Notes	Sampling Time
	Organic Carbon	- Blank correction is needed to account for positive artifact.[45] - OC/EC split is operationally defined using TOT method (NIOSH).	
	Elemental Carbon	- OC/EC split is operationally defined using Thermal Optical Transmittance (TOT) method (NIOSH)	
CASTNet	SO4		weekly average
	NO3	Negative artifact	
	NH4		
	SO2		
	HNO3	Split between NO3 and HNO3 may not be reliable in warm weather due to conversion from NO3 to HNO3	
IMPROVE	$PM_{2.5}$		24-hour average 1/3 days
	SO4 and S	- Sulfur measurements are usually used to estimate sulfate	
	NO3		
	NH4	- Only measured at 3 Eastern sites.	
	Organic Carbon	- OC/EC split is operationally defined using Thermal Optical Reflectance (TOR) analysis. - Blank corrections are applied to all data before they are reported.	
	Elemental Carbon	- OC/EC split is operationally defined using Thermal Optical Reflectance (TOR) analysis.	

In general, the speciation data can be directly compared to the model outputs. Of note, care should be taken in examining issues with blank corrections for organic carbon (especially when measured OC is low) as well as the EC/OC split. The STN and IMPROVE networks use different analysis techniques when estimating EC and OC. This will generally have a larger impact on EC concentrations. It may not be appropriate to compare EC concentrations between networks. The Clean Air Scientific Advisory Committee (CASAC) has recommended to transition the STN network to the IMPROVE carbon protocol (TOR) in order to provide data

[45] Network average blank corrections for many of the STN samplers are large (up to 1.53 ug/m3).

consistency among the two networks (U.S. EPA, 2005f). Over the next two to three years (2007-2009), EPA plans to convert all of the STN sites (approximately 200 trends and supplemental sites) to IMPROVE carbon sampling and analysis protocols.

Due to the previously stated positive and negative artifacts, the FRM data may not be directly comparable with model outputs without some adjustments. The models predict PM species as they might be measured in the ambient air. But the FRM measurements may not always measure what is in the air (due to known positive and negative artifacts). As part of the $PM_{2.5}$ attainment test in Section 5 we recommended default methodologies to adjust speciation (STN and IMPROVE) data to better approximate the FRM data. A similar adjustment could be made to the model output data to allow the direct comparison to FRM measurements. But due to the relatively limited information to be gained by evaluating daily average total measured $PM_{2.5}$ mass (as compared to $PM_{2.5}$ components), it may not be worth the effort. More meaningful information may be gleaned from continuous $PM_{2.5}$ ambient data. Comparisons between model output and continuous data can be made on an hourly basis. This may provide important information with regards to the daily temporal patterns predicted by the model.

Ambient Measurement Networks

Provided below is an overview of some of the various ambient air monitoring networks currently available. Network methods and procedures are subject to change annually due systematic review and/or updates to the current monitoring network/program. Please note, there are other available monitoring networks which are not mentioned here and more details on the networks and measurements should be obtained from other sources.

AQS The **A**ir **Q**uality **S**ystem (AQS) is not an air quality monitoring network. However it is a repository of ambient air pollution data and related meteorological data collected by EPA, state, local and tribal air pollution control agencies from tens of thousands of monitors. The monitoring data in AQS are the result of the various Clean Air Act requirements to provide a national database of ambient air pollution data. This information management system contains over 1000 pollutants from 1957 through the present day. AQS contains all the routine hourly gaseous pollutant data collected from **S**tate and **L**ocal **A**ir **M**onitoring **S**tations (SLAMS) and **N**ational **A**ir **M**onitoring **S**tations (NAMS) sites. SLAMS is a dynamic network of monitors for state and local directed monitoring objectives (e.g., control strategy development) which consists of approximately 4,000 monitoring stations. A subset of the SLAMS network, the NAMS has an emphasis on urban and multi-source areas (i.e, areas of maximum concentrations and high population density).

The AQS database includes criteria pollutant data (SO2, NO2, O3, CO, PM10, $PM_{2.5}$ and Pb) and speciation data of particulate matter (SO4, NO3, NH4, EC, OC, and crustal Material), air toxic data, photochemical assessment data, and meteorological data. The data are measured and reported on an hourly or daily average basis. The AQS system continues to expand as regulations consider to include more ambient air pollutants (U.S. EPA, 2006b). An overview of the AQS can be found at http://www.epa.gov/ttn/airs/airsaqs/index.htm. For querying the database or viewing

the data one can use EPA's Air Explorer tool (http://www.epa.gov/mxplorer/aboutthedata.htm) which is a collection of user-friendly visualization tools for air quality analysts. The tools generate maps, graphs, and data tables dynamically.

IMPROVE The Interagency **M**onitoring of **PRO**tected **V**isual **E**nvironments (IMPROVE) network began in 1985 as a cooperative visibility monitoring effort between EPA, federal land management agencies, and state air agencies (IMPROVE, 2000). Data are collected at Class I areas across the United States mostly at National Parks, National Wilderness Areas, and other protected pristine areas. Currently, there are approximately 160 IMPROVE rural/remote sites that have complete annual $PM_{2.5}$ mass and/or $PM_{2.5}$ species data. Of the total sites, 110 sites are for regional haze and visibility (and are located at or near Class I areas) and 50 are protocol sites (determined by the needs of state and local air pollution control agencies to meet their respective State Implementation Plan (SIP) requirements). The website to obtain IMPROVE documentation and/or data is http://vista.cira.colostate.edu/improve/.

CASTNet Established in 1987, the **C**lean **A**ir **S**tatus and **T**rends **N**etwork (CASTNet) is a dry deposition monitoring network where data are collected and reported as weekly average data (U.S. EPA, 2002b). CASTNet provides atmospheric data on the dry deposition component of total acid deposition, ground-level ozone and other forms of atmospheric pollution. The data are collected in filter packs that sample the ambient air continuously during the week. CASTNet now comprises over 70 monitoring stations across the United States. The longest data records are primarily at eastern U.S. sites. More information can be obtained through the CASTNet website at http://www.epa.gov/castnet/.

STN The **S**peciation **T**rends **N**etwork (STN) began operation in 1999 to provide nationally consistent speciated $PM_{2.5}$ data for the assessment of trends at representative sites in urban areas in the U.S. The STN was established by regulation and is a companion network to the mass-based Federal Reference Method (FRM) network implemented in support of the $PM_{2.5}$ NAAQS. As part of a routine monitoring program, the STN quantifies mass concentrations and $PM_{2.5}$ constituents, including numerous trace elements, ions (sulfate, nitrate, sodium, potassium, ammonium), elemental carbon, and organic carbon. There are currently 54 Trends sites nationally in operation. Trends sites are largely static urban monitoring stations with protocols for sampling methods; dedicated to characterizing aerosol mass components in urban areas of the U.S. to discern long-term trends and provide an accountability mechanism to access effectiveness of control programs. In addition, there are approximately 181 supplemental speciation sites which are part of the STN network and are SLAMS sites. The STN data at trends sites are collected 1 in every 3 days, whereas supplemental sites collect data either 1 in every 3 days or 1 in every 6 days. Comprehensive information on the STN and related speciation monitoring can be found at http://www.epa.gov/ttn/amtic/specgen.html and http://www.epa.gov/aqspubl1/select.html.

NADP Initiated in the late 1970s, the **N**ational **A**cid **D**eposition **P**rogram (NADP) monitoring network began as a cooperative program between federal and state agencies, universities, electric utilities, and other industries to determine geographical patterns and trends in precipitation

chemistry in the U.S. NADP collects and reports wet deposition measurements as weekly average data (NADP, 2002). The network is now known as NADP/NTN (National Trends Network) with nearly 200 sites in operation. The NADP analyzes the constituents important in precipitation chemistry, including those affecting rainfall acidity and those that may have ecological effects. The NTN measures sulfate, nitrate, hydrogen ion (measure of acidity), ammonia, chloride, and base cations (calcium, magnesium, potassium). Detailed information regarding the NADP/NTN monitoring network can be found at http://nadp.sws.uiuc.edu/.

SEARCH The **S**outh **E**astern **A**erosol **R**esearch and **CH**aracterization (SEARCH) monitoring network was established in 1998 and is a coordinated effort between the public and private sector to characterize the chemical and physical composition as well as the geographical distribution and long-term trends of $PM_{2.5}$ in the Southeastern U.S (Edgerton, 2005; Hansen, 2003; Zheng, 2002). SEARCH data are collected and reported on an hourly/daily basis. There are currently six SEARCH measurement sites: Birmingham, Alabama (urban), Centreville, Alabama (rural), Gulfport, Mississippi (urban), Jefferson Street, Atlanta, Georgia (urban), Oak Grove, Mississippi (rural), Yorkville, Georgia (rural), suburban Pensacola, Florida (suburban), and downtown Pensacola, Florida (urban). Background information regarding standard measurement techniques/protocols and data retrieval can be found at http://www.atmospheric-research.com/studies/SEARCH/index.html.

Supersites The Supersite (SS) monitoring program began in 2001 as an EPA OAQPS and ORD co-funded project which measures speciated $PM_{2.5}$ at eight air shed sites. Specifically, these eight monitoring supersites are located in Baltimore, Maryland; Pittsburgh, Pennsylvania; New York City, New York; Atlanta, Georgia; Houston, Texas; St. Louis, Missouri; Fresno, California and Los Angeles, California. These supersites were developed as special study platforms for research and measurement methods to address the scientific uncertainties associated with $PM_{2.5}$. The goals of the SS program include (1) fine particulate characterization to understand processes and source-receptor relationships and to support $PM_{2.5}$ and ozone SIPs, (2) testing of advanced sampling methods to enable a smooth transition from established routine techniques, and (3) development of monitoring data and samples shown in (1) to support health effects and exposure studies. PM Supersites information can be obtained at http://www.epa.gov/ttn/amtic/supersites.html.

18.4 How Should the Operational Evaluation of Performance Be Completed?

As noted above, an operational evaluation is used to assess how accurately the model predicts observed concentrations. Therefore, an operational evaluation can provide a benchmark for model performance and identify model limitations and uncertainties that require diagnostic evaluation for further model development/improvement. Computer graphics, ozone and PM metrics/statistics, and observational models are all potentially useful for evaluating a model's ability to predict observed air quality. This section describes the recommended statistical measures, graphical displays, and other analytical techniques which should be considered as part

of an operational evaluation of ozone and $PM_{2.5}$ model performance. Note that model predictions from the ramp-up days should be excluded from the analysis of model performance.

18.4.1 Ozone Operational Evaluation

Statistics: It is recommended that, at a minimum, the following three statistical measures be calculated for hourly ozone and 8-hourly maxima ozone over the episode days in an attainment demonstration.

- Mean Normalized Bias (MNB): This performance statistic averages the model/observation residual, paired in time, normalized by observation, over all monitor times/locations. A value of zero would indicate that the model over predictions and model under predictions exactly cancel each other out. The calculation of this measure is shown in Equation 18.1.

$$\text{MNB} = \frac{1}{N}\sum_{1}^{N}\left(\frac{(\text{Model}-\text{Obs})}{\text{Obs}}\right)\cdot 100\% \qquad (18.1)$$

- Mean Normalized Gross Error (MNGE): This performance statistic averages the absolute value of the model/observation residual, paired in time, normalized by observation, over all monitor times/locations. A value of zero would indicate that the model exactly matches the observed values at all points in space/time. The calculation of this measure is shown in Equation 18.2.

$$\text{MNGE} = \frac{1}{N}\sum_{1}^{N}\left(\frac{|\text{Model}-\text{Obs}|}{\text{Obs}}\right)\cdot 100\% \qquad (18.2)$$

- Average Peak Prediction Bias and Error: These are measures of model performance that assesses only the ability of the model to predict daily peak 1-hour and 8-hour ozone. They are calculated essentially the same as the mean normalized bias and error (Equation 18.1 and 18.2), except that they only consider daily maxima data (predicted versus observed) at each monitoring location. In the attainment test, models are used to calculate relative response factors near monitoring sites by taking the ratio of the average 8-hour daily maximum concentrations calculated for the future and current cases. Thus, the model's ability to predict observed mean 8-hour daily maxima is an important indicator of model performance.

Appendix A provides other statistics such as mean bias, mean error, mean fractional bias, mean fractional error, root mean square error, correlation coefficients, etc. which should also be

calculated to the extent that they provide meaningful information. Wherever possible, these types of performance measures should also be calculated for ozone precursors and related gas-phase oxidants (NOx, NOy, CO, HNO3, H2O2, VOCs and VOC species, etc.) and ozone (and precursors) aloft.

Thresholds: EPA recommends that the three metrics above be calculated two ways: 1) for pairs in which the 1-hour or 8-hour observed concentrations are greater than 60 ppb[46], and 2) for all pairs (no threshold)[47]. This will help to focus the evaluation on the models ability to predict NAAQS-relevant ozone and minimize the effects of the normalization. In terms of pairing model predictions with monitored observations, EPA recommends that the grid cell value in which the monitor resides be used for the calculations. It would also be acceptable to consider bi-linear interpolation of model predictions to specific monitoring locations[48]. States/Tribes should recognize that, even in the case of perfect model performance, model-observed residuals are unlikely to result in exact matches due to differences between the model predictions which are volume averages and the observations which are point values.

Averaging: The statistics should initially be calculated for individual days (averaged over all sites) and individual sites (averaged over all days). As appropriate, States/Tribes should then aggregate the raw statistical results into meaningful groups of subregions or subperiods. For example, examination of model performance within and near the non-attainment area(s) of interest may be more important than examination of performance in other parts of the modeling domain. Similarly, priority may be placed on examination of the days that are potentially used in the attainment test (base period days with 8-hour ozone > 70 or 85 ppb). That is not to say that model performance evaluations should ignore performance on lower ozone days or in areas outside of the nonattainment areas. Simply, additional factors should be considered when prioritizing more detailed analyses.

---

[46] Past ozone modeling applications have used a minimum cutoff of either 40 ppb or 60 ppb. Due to the interest in predicted ozone concentrations at or above the 8-hour standard (85 ppb), the higher cut off (60 ppb) is recommended.

[47] The use of a 0 ppb threshold can add valuable information about the ability of the model to simulate a wide range of conditions. Because of the tendency of the MNB and MNGE metrics to inflate the importance of biases at the lowest observed values (which are in the denominator), it is recommended that the alternate metrics of normalized mean bias (NMB) and normalized mean gross error (NMGE) be used as substitutes for evaluations with no minimum threshold.

[48] In certain instances, States/Tribes may also want to conduct performance evaluations using the “near the monitor” grid cell arrays. A “near the monitor” analysis may be useful when strong ozone gradients are observed, such as in the presence of a sea breeze or in strongly oxidant limited conditions. Furthermore, a “near the monitor” performance evaluation is consistent with the RRF methodology.

Plots/Graphics: Along with the statistical measures, EPA recommends that the following four sets of graphical displays be prepared and included as part of the performance analysis.

- Time series plots of model and predicted hourly ozone for each monitoring location in the nonattainment area, as well as key sites outside of the nonattainment area. These plots can indicate if there are particular times of day or days of the week when the model performs especially poorly. Figure 18.1 shows a sample time series plot. This graphic was developed using the air quality analysis module of the Atmospheric Model Evaluation Tool (AMET) (Appel, 2005), though other evaluation tools exist and can be used.

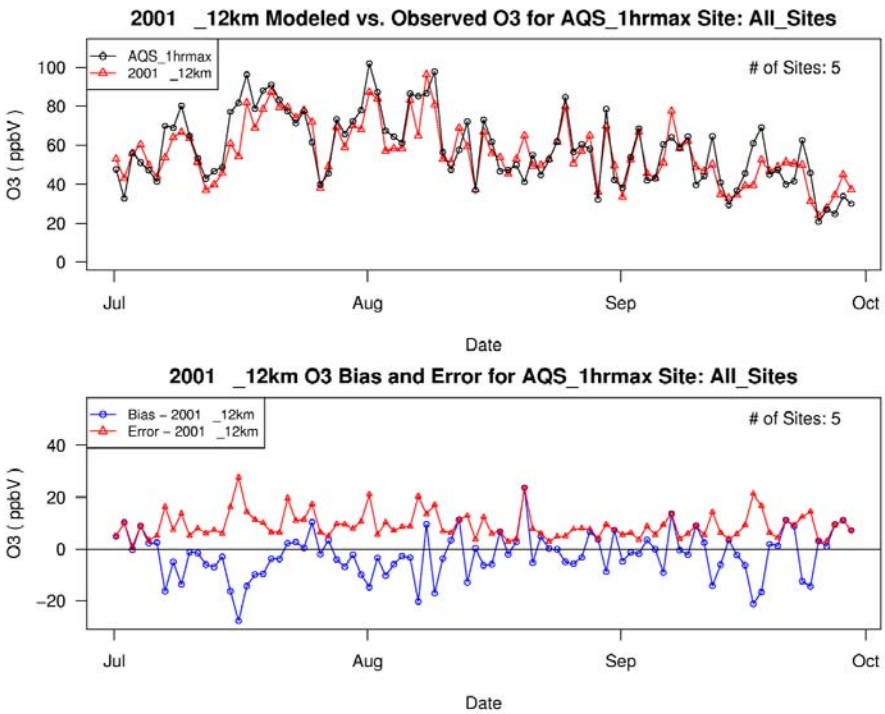


Figure 18.1. Sample time series analysis plot for a 12-km model vs. observed 2001 case. The upper plot shows a particular analysis for 1-hour maxima ozone, July through September performance averaged over five sites in Cleveland, Ohio. The bottom plot shows the bias and error (percentage) calculated based on the predicted 1-hour maxima ozone from the upper plot.

- Scatter plots of predicted and observed ozone at each site within the nonattainment area (and/or an appropriate subregion). These plots should be completed using: a) all hours within the modeling period for hourly ozone, and b) all 8-hour daily maxima within the modeling period. It may also be useful to develop separate plots for individual time periods or key subregions. These plots are useful for indicating if there is a particular part

of the distribution of observations that is poorly represented by the model[49]. See Figure 18.2 below for a scatter plot example, which was generated using the AMET.

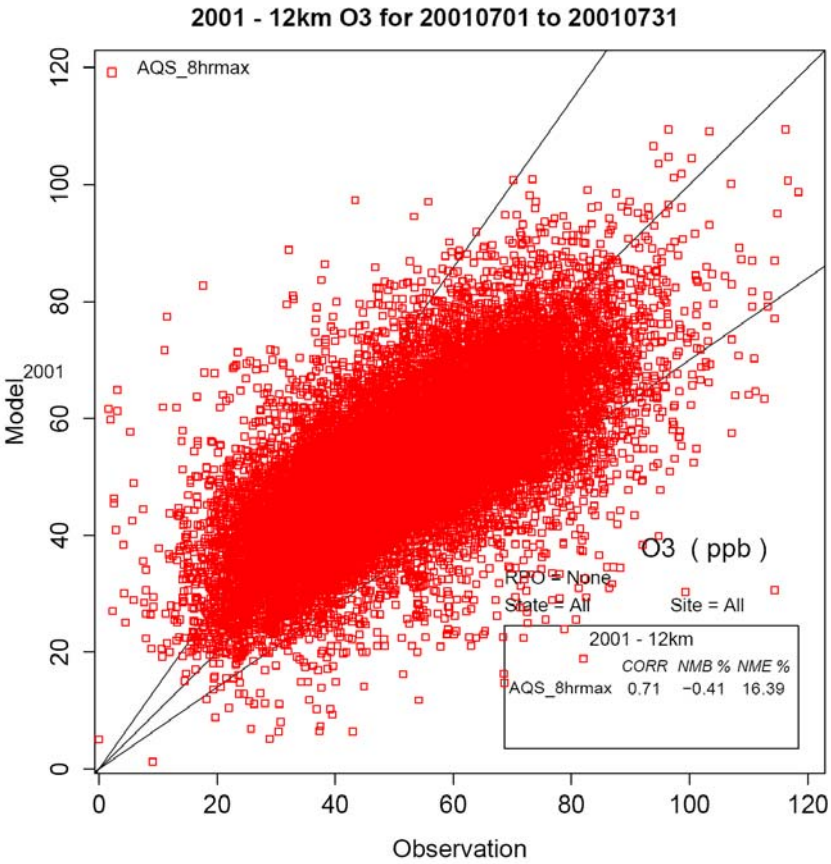


Figure 18.2. Scatter plot analysis for a 12-km model vs. observed 2001 case. The scatter plot shows a performance analysis for daily 8-hour maxima ozone pairs for the month of July.

- Daily tile plots of predicted ozone across the modeling domain with the actual observations as an overlay. Plots should be completed for both daily 1-hour maxima and daily 8-hour maxima. These plots can reveal locations where the model performs poorly. Superimposing observed hourly or daily maximum concentrations on the predicted isopleths reveals useful information on the spatial alignment of predicted and observed plumes.

---

[49] Quantile-quantile (Q-Q) plots may also provide additional information with regards to the distribution of the observations vs. predictions. But due to the fact that Q-Q plots are not paired in time, they may not always provide useful information. Care should be taken in interpreting the results.

- Animations of predicted hourly ozone concentrations for all episode days or for certain periods of interest. Animations are useful for examining the timing and location of ozone formation. Animations may also reveal transport patterns (especially when looking at ozone aloft). Animations can also be used to qualitatively compare model outputs with the conceptual model of particular ozone episodes.

18.4.2 PM/RH Operational Evaluation

An operational evaluation for $PM_{2.5}$ and regional haze is similar to that for ozone. Some important differences are that $PM_{2.5}$ consists of many components and is typically measured with a 24-hour averaging time. The individual components of $PM_{2.5}$ should be evaluated individually. In fact, it is more important to evaluate the components of $PM_{2.5}$ than to evaluate total $PM_{2.5}$ itself. Apparent "good performance" for total $PM_{2.5}$ does not indicate whether modeled $PM_{2.5}$ is predicted for "the right reasons" (the proper mix of components). If performance of the major components is good, then performance for total $PM_{2.5}$ should also be good.

This section contains some additional recommended statistics that have typically not been calculated for ozone performance, but have been found to be particularly useful for PM analyses (such as fractional bias and error). We also show examples of some new types of display plots such as "soccer plots" and "bugle plots". Soccer plots provide a convenient way to display a summary of model performance (including bias and error at the same time). Bugle plots have variable bias and error goals, based on ambient concentrations. This allows for a higher percentage error and bias at very low concentrations. This recognizes the fact that models often have difficulty in accurately predicting near background concentrations[50].

Similar to ozone evaluations, it may be useful to prioritize examination of model performance within and near the non-attainment area(s) and/or Class I areas of interest. Additionally, priority may be placed on examination of the days that are potentially used in the attainment test (20% best and worst days for visibility and the days > 65 ug/m3 for 24-hour $PM_{2.5}$).

Statistics: Statistics for PM and regional haze performance are similar to those calculated for ozone. We recommend calculating statistics for components of $PM_{2.5}$, and PM precursors. Useful metrics include mean fractional bias and mean fractional error, normalized mean bias, and normalized mean error. Formulas for estimating these metrics at an individual monitoring site, j, are shown in Equations (18.3) -(18.6). Other statistics such as mean bias, mean error, root mean square error, correlation coefficients, etc. should also be calculated to the extent that they provide meaningful information (see Appendix A for definitions).

---

[50]Predicting very low concentrations is not critical for modeling the $PM_{2.5}$ NAAQS. However, predicting low concentrations of PM species can be important for regional haze modeling, where in the cleanest Class I areas, PM concentrations may be less than a few ug/m3 (especially on the 20% best days).

- **Mean Fractional Bias (percent):** Normalized bias can become very large when a minimum threshold is not used. Fractional bias is used as a substitute. The fractional bias for cases with factors of 2 under- and over-prediction are -67 and + 67 percent, respectively (as opposed to -50 and +100 percent, when using normalized bias). Fractional bias is a useful indicator because it has the advantage of equally weighting positive and negative bias estimates. The single largest disadvantage is that the predicted concentration is found in both the numerator and denominator.

  Equation 18.3:

$$FBIAS = \frac{2}{N}\sum_{1}^{N}\left(\frac{(Model - Obs)}{(Model + Obs)}\right)\cdot 100\%$$

- **Mean Fractional Error (percent):** Normalized error can become very large when a minimum threshold is not used. Therefore fractional error is used as a substitute. It is similar to the fractional bias except the absolute value of the difference is used so that the error is always positive.

  Equation 18.4:

$$FERROR = \frac{2}{N}\sum_{1}^{N}\left(\frac{|Model - Obs|}{(Model + Obs)}\right)\cdot 100\%$$

- **Normalized Mean Bias (percent):** This performance statistic is used as a normalization to facilitate a range of concentration magnitudes. This statistic averages the difference (model - observed) over the sum of observed values. Normalized mean bias is a useful model performance indicator because it avoids over inflating the observed range of values.

  Equation 18.5:

$$NMB = \frac{\sum_{1}^{N}(Model - Obs)}{\sum_{1}^{N}(Obs)}\cdot 100\%$$

- **Normalized Mean Error (percent):** This performance statistic is used as a normalization of the mean error to facilitate a range of concentration magnitudes. This statistic averages the difference (model - observed) over the sum of observed values. Normalized mean error is a useful model performance indicator because it avoid over inflating the observed range of values.

  Equation 18.6:

$$NME = \frac{\sum_{1}^{N} |Model - Obs|}{\sum_{1}^{N} (Obs)} \cdot 100\%$$

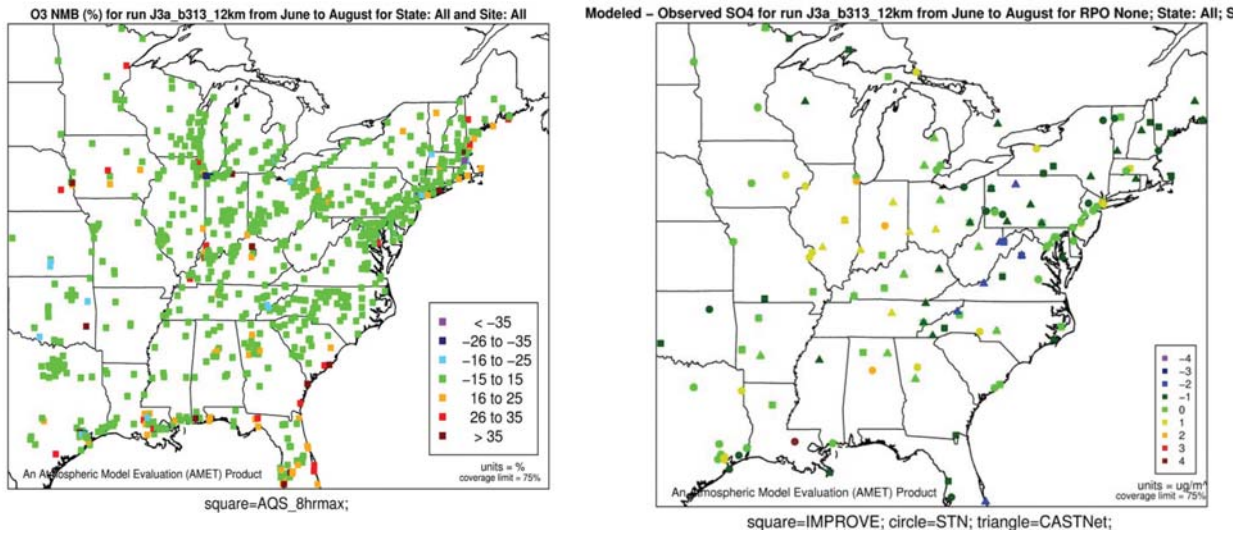
Averaging Times: Note that units of time associated with model and observed concentrations can be days (i.e., usually for particulate matter and its species), hours (i.e., usually for species with continuous measurements, like ozone) or sometimes weeks (CASTNet filter pack measurements). Also note that the preceding metrics may not be meaningful if the number of modeled days with monitored data is limited at a site.

Since modeling for the $PM_{2.5}$ NAAQS and regional haze will likely require modeling different times of year, season-specific statistics and graphic displays are helpful for evaluating and diagnosing model performance. Statistics and graphics can be averaged for various time scales. For example, statistical metrics and scatterplots can show daily averaged ambient-modeled pairs, monthly averaged pairs, quarterly (or seasonal averaged) pairs, or annual average pairs. Each of these averaging times can provide useful information. We recommend a range of different averaging times for annual or seasonal modeling. At a minimum, States should examine daily averaged pairs and seasonal (or quarterly) averaged pairs. It should be noted that statistics and plots tend to look “better” as the averaging time increases from daily to monthly to quarterly to annual. As such, daily pairs should always be examined to ensure a detailed look at model performance on the time scale of the FRM and STN measurements (24-hour average).

Plots/Graphics: As mentioned above in the ozone operational evaluation, graphics are a useful means for understanding *how* predictions and observations differ. For PM evaluations we recommend creating time series plots, tile plots, and scatter plots. Time series plots tell whether there is any particular time of day, day(s) of the week, or months (seasons) of the year when the model performs poorly. 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 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 component $PM_{2.5}$ observations and predictions which just happen to balance, producing low estimated aggregated bias. As mentioned above, since the NAAQS for $PM_{2.5}$ and the regional haze goals will likely require modeling different times of year, season-specific graphic displays are helpful for evaluating and diagnosing model performance.

These above mentioned graphical plots have been provided as examples in the ozone operational evaluation section. Other graphical analysis displays can also be developed to better inform a model performance evaluation. Additional types of graphical plots are shown below. These plots were generated using the AMET tool.

Figure 18.3 shows an example of spatial plots of normalized mean bias and mean bias by monitoring site. These plots allow an analysis of regional differences in model performance (e.g. inland vs. coastal areas).

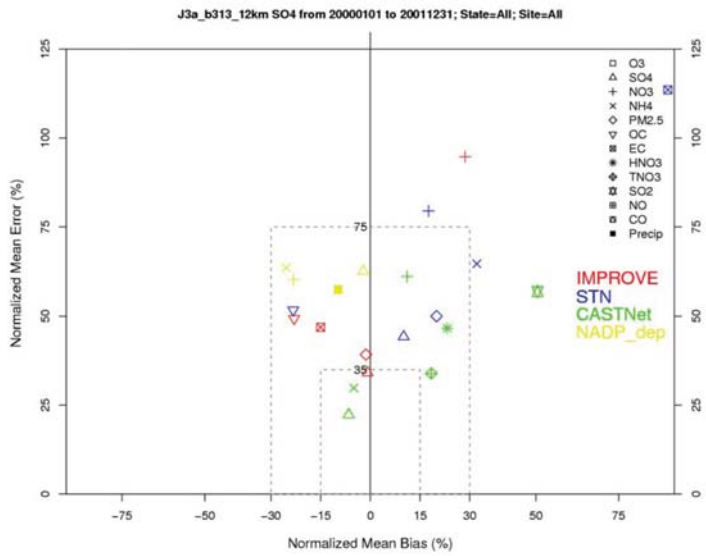


(a) (b)

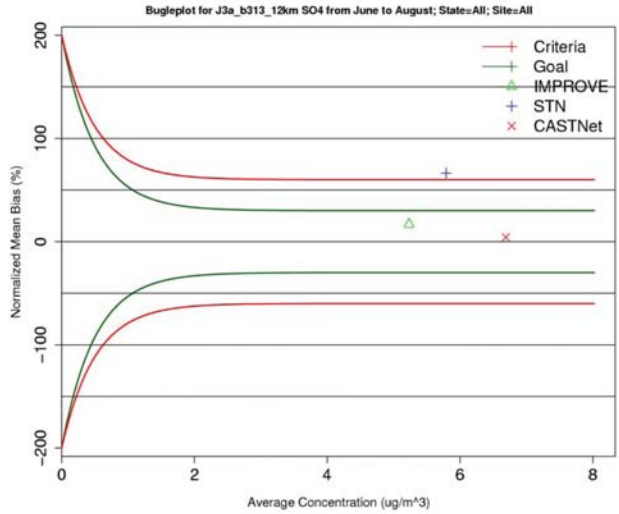
Figure 18.3. Example of (a) Spatial plot of NMB; (b) model-observed difference plot (mean bias).

Figure 18.4 shows an example of a “soccer plot” and “bugle plots” (Boylan, 2006) (Tesche, 2006). The soccer plot is so named because the dotted lines resemble a soccer goal. The plot is a convenient way to visualize model performance, as measures of both bias and error are shown on a single plot. As bias and error approach zero, the points are plotted closer to or within the “goal”, represented here by the dashed boxes. The “bugle plot”, named for the shape formed by the criteria and goal lines, is another plot available for model performance evaluation. The bugle plots are shaped as such because the goal and criteria lines are adjusted based on the average concentration of the observed species. As the average concentration becomes smaller, the criteria and goal lines become larger to adjust for the model’s ability to predict at low concentrations. We recommend allowing for larger bias and error when the ambient concentration falls below 2 ug/m3.

(a)



(b)



(c)

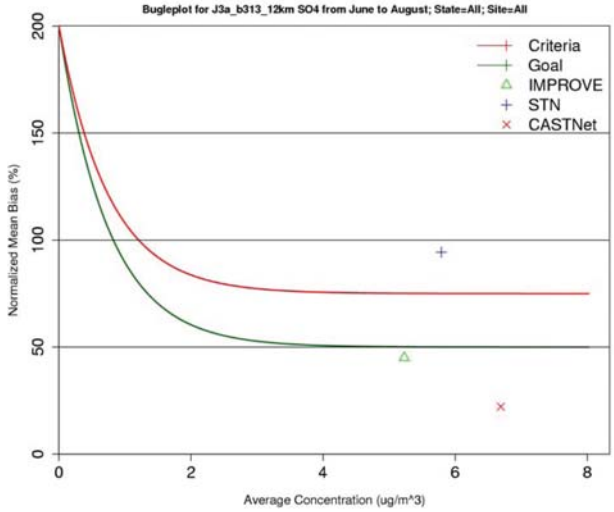


Figure 18.4. (a) The "soccer goal" plot measures both bias and error. The "goal" is represented by the dashed boxes. (b) A "bugle plot" of bias and (c) a "bugle plot" of error. As the average concentration becomes smaller, the criteria and goal lines become larger to adjust for the model's ability to predict at low concentrations.

Figure 18.5 shows an example of two types of box plots can be developed for model performance evaluation. One displays monthly (left) and one that displays hourly (right) information. The box plots show several quantities: the shading represents the 25% to 75% quartiles, while the lines represent the median values, for both model and observation concentrations. The monthly box plot can be used to quickly visualize model performance across the entire year, highlighting the seasonal change in model performance. The hourly or "diurnal" box plot is used with hourly data, and shows how the model predictions compare against observations throughout an entire day.

(a) (b)

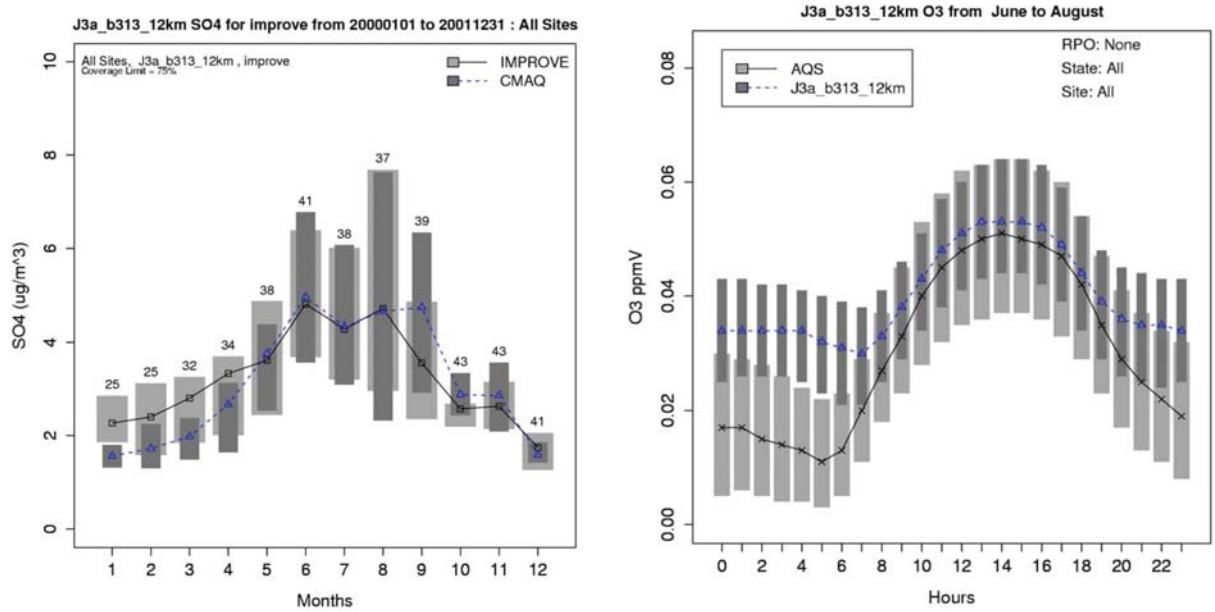


Figure 18.5. The box plots show several quantities: the shading represents the 25% to 75% quartiles, while the lines represent the median values, for both model and observation concentrations. (a) displays monthly data (left) and (b) displays hourly data (right).

18.5 How Do I Assess Model Performance and What Types of Analyses Can be Done to Evaluate the Accuracy of the Model Response: Diagnostic Evaluations?

This section lists possible analyses that could be performed to investigate the ability of the model to accurately forecast changes in predicted concentrations of ozone and PM resulting from changes in ozone and $PM_{2.5}$ precursor emissions. Diagnostic evaluations/tests are used to explain model performance and to provide clues about how to improve the reliability of predictions.

The diagnostic analyses discussed in this section include observational models, probing tools, receptor models, retrospective analyses, and sensitivity analyses. This represents a large variety of tools that can aid in diagnostic analyses. Some of these techniques are based on analysis of ambient data, some require running the photochemical many times, and some of the techniques involve coding of the photochemical model to allow additional outputs. States/Tribes are encouraged to complete as many of these types of diagnostic analyses as possible throughout the modeling process in order to help understand model performance and to help develop information which may help improve performance, as well as increase confidence in the modeled attainment projections.

18.5.1 Observational models

A performance evaluation which includes comparisons between modeled and observed ratios of indicator species carries with it a large potential advantage. For ozone, measurements of certain "indicator species ratios" are a potentially useful way to assess whether local ozone formation is VOC- or NOx-limited at any particular point in space and time as well as, help reveal whether the model is correctly predicting the sensitivity of ozone to VOC and/or NOx controls (e.g., comparisons between modeled and observed ratios: O3/NOy, O3/HNO3) (Sillman, 1995, 1997, 1998, and 2002; Lu and Chang, 1998; Pun and Seigneur, 1999). For PM, 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 (Ansari and Pandis, 1998; Blanchard et al., 2000; Pun and Seigneur, 2001). If a model accurately predicts observed ratios of indicator species, then one can conclude with additional confidence that the predicted change in future year ozone or PM may be accurate. One precaution with respect to the use of indicator species 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 response to controls, as predicted by the model is correct (especially for secondary particulate matter to changes in precursors). 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. A second precaution is that application of this method often requires more measurements than are commonly made. In some cases, it may be difficult to achieve the required precision with current routine monitoring. Finally, much of the work done to date with indicator species has focused on peak hourly concentrations of ozone. Despite these precautions, comparing predicted and observed ratios of indicator species provides a means of assessing a model's ability to accurately characterize the sensitivity of predicted ozone and predicted secondary components of $PM_{2.5}$ to changes in precursors.

Many studies and 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 continues to gain more experience with chemical mechanisms incorporating secondary particulate formation and more speciated particulate and gas phase ambient measurements become available.

Other observational methodologies exist and can be used in a similar manner. The Smog Production (SP) algorithm is another means by which ambient data can be used to assess areas that are NOx or VOC-limited (Blanchard, 1999). Additionally, it has been postulated that differences in weekend-weekday ozone and $PM_{2.5}$ patterns may also provide real-world information on which precursors are most responsible for ozone and $PM_{2.5}$ formation in any given area (Pun, 2001 and 2003; Heuss, 2003; Fujita, 2003; Blanchard and Tanenbaum, 2003 and 2005; Chinkin, 2003; Yarwood, 2003; Qin, 2004). For example, 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. In areas where there are large differences between average weekend and weekday ambient ozone and $PM_{2.5}$ concentrations over the span of several seasons, it would be useful to compare statistical model performance for weekends versus weekdays (e.g., does the model accurately reflect observed differences in component concentrations in summer vs. winter?). This would allow one to assess whether the model is capturing the effect of the emissions differences which are presumably driving the real-world concentration differences.

18.5.2 Probing Tools

Recently, techniques have been developed to embed procedures within the code of an air quality model which enable users to assess the contributions of specific source categories or of specific geographic regions to predicted model concentrations at specified sites (e.g. for ozone predictions, Zhang, 2003). Various techniques have been implemented into various air quality models, but three of the most commonly used probing tools are *photochemical source apportionment* (Dunker, 2002a; ENVIRON, 2006a; Tonnesen and Wang, 2004; SAI, 2005; Yarwood, 2005; Douglas, 2006), the *decoupled direct method* (DDM) (Dunker, 1980 1981, and 1984; Yang, 1997a and 1997b; Dunker, 2002b; ENVIRON, 2006a; Cohan, 2002, 2004, and 2005; Hakami, 2003 and 2004), *process analysis* (Jeffries, 1994 and 1997; Jeffries, 1996; Jang, 1995; Lo, 1997; Byun and Ching, 1999; Morris, 2001 and 2003; ENVIRON, 2006a; Henderson, 2006). In the context of model performance evaluation, these attribution procedures are useful in that they allow one to "track" the importance of various emissions categories or phenomena contributing to predicted ozone and particulate matter at a given location. This can provide

valuable insight into whether the model is adequately representing the conceptual description of ozone and PM patterns in the nonattainment area. In cases where model performance is subpar, these analyses can be useful for indicating where model input or model algorithm improvements are most needed.

Caution should be applied to the tools and techniques mentioned here as they are limited by the accuracy of the underlying air quality photochemical model, when the tools utilize the same algorithms/modules and assumptions. An overview of these aforementioned probing tools are provided below.

Photochemical source apportionment: A photochemical source apportionment tool tags and tracks the release, transport, chemical transformation, and deposition of precursor species from primary emission sources, source categories, source regions, initial conditions and/or boundary conditions from a photochemical grid model. Therefore, the contribution of tagged sources to simulated concentrations (including secondary pollutants) and deposition can be measured in a single model run. This source apportionment tool can be used to estimate how emissions from individual source areas and regions affect modeled ozone and PM concentrations over space and time. For example, this is achieved by using multiple tracer species to track the fate of ozone precursor emissions (VOC and NOx) and the ozone formation caused by these emissions within an air quality model simulation. There are challenges to consider during implementation of this methodology, i.e., tracking the spatial and temporal relationships between separate groups of emission sources and ozone and particulate matter formation (Douglass, 2006), (Environ, 2006a), (Environ, 2006b). Among them are insuring compatibility with the underlying air quality model formulation so that derived source-receptor relationships will be consistent with model response to emission changes.

The information gleaned from source apportionment can be useful in considering potential control strategies. However, a key consideration in weighing alternative strategies is which types of emissions are cost effective and practical to control, a consideration not addressed by source apportionment modeling.

Decoupled Direct Method (DDM): The DDM calculates the local, first-order sensitivity coefficients of gas-phase pollutant concentrations to perturbations of model inputs, e.g. initial conditions, boundary conditions, emissions. Theoretically, the DDM is applicable to the calculation of higher order sensitivity coefficients as well, but these are much more numerous than the first order coefficients and generally will be less useful (Environ, 2006a). DDM can also be used to explore model sensitivities to perturbations in meteorological parameters and chemical rate constants, although resource intensive and at present somewhat less useful. The decoupled direct method is advantageous both in terms of computational efficiency and stability of the numerical solution. DDM has been used as a sensitivity analysis tool in several studies. A single model run using DDM can potentially provide the information equivalent to tens, if not hundreds of brute force emissions reductions model runs. DDM has been incorporated into several models including CAMx (Environ, 2006a) and CMAQ (Cohan, 2005).

Process analysis: Process analysis techniques allows one to quantitatively track (mass balancing) the contributions of individual physical and chemical atmospheric processes to the overall change in pollutant concentrations at a particular time and location. Thus, process analysis will assess the relative importance of each physical and chemical process for a particular model simulation as well as explain how model predictions are obtained, which is valuable when modeling nonlinear systems such as atmospheric photochemistry. Since models used to simulate ozone and secondary particulate matter are similar, process analysis should also be useful for addressing both ozone and $PM_{2.5}$ (and visibility related) applications.

Three of the most common process analysis tools implemented in grid models are the integrated process rate (IPR) analysis, the integrated reaction rate (IRR) analysis, and the chemical process analysis (CPA). The IPR analysis is used to determine the relative contributions of individual physical (e.g., advection, diffusion, emissions) and chemical (e.g., chemical reaction, aqueous chemistry, aerosol production) processes, potentially revealing unexpected low or high process contributions. The IRR analysis is used to identify and explain the details of the chemical transformations (chemical pathways and key chemical characteristics) in a particular model simulation. Typically, IRR analyses have been used to determine and understand model prediction differences between chemical mechanisms (e.g., CB-IV, SAPRC). Likewise, the CPA is an improvement on the IRR analysis technique where part of the processing of IRR information is internalized within the air quality grid model to yield key chemical parameters directly (e.g., budget terms for ozone and NOx).

Process analysis requires a substantial amount of expertise to be interpreted to full advantage. However, useful insights are also possible with less detailed analyses. An analysis can focus on selected grid cells in a small area. Process analysis then takes advantage of the fact that a numerical grid model addresses physical and chemical factors affecting secondary pollutants 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 wish to discount that prediction in the attainment demonstration or uniform rate of progress assessment.

18.5.3 Receptor Models (Multi-variate Source Apportionment)

Receptor models (also referred to as multi-variate source apportionment models) are mathematical or statistical procedures for identifying and quantifying the sources of air pollutants at a receptor location (Henry, 1991; Hopke, 1991, 1999, 2001; Lewis, 2003; Coulter, 2000; Willis, 2000; Watson, 2002; Maykut, 2003; Mukerjee, 2004). Receptor models are generally used to examine primary pollutants such as the primary components of $PM_{2.5}$. They are not typically used to examine ozone or secondary $PM_{2.5}$ species. Unlike photochemical and dispersion air quality models, receptor models do not use pollutant emissions, meteorological data and chemical transformation mechanisms to estimate the contribution of sources to receptor concentrations. Instead, receptor models use the chemical and physical characteristics of gases and particles measured at source and receptor to both identify the presence of and to quantify source contributions to receptor concentrations. Receptor models can identify possible air quality management strategy solutions especially when used in conjunction with examination of the local emission inventory (Brook, 2004). These models are therefore a natural complement to other air quality models and can be used as part of State Implementation Plans (SIPs) for identifying sources contributing to air quality problems.

Receptor models provide scientific support for current ambient air quality standards and for implementation of those standards by identifying and quantifying contributions of various source types. As mentioned in sections 18.2 and 18.3, the richness of the ambient air quality data sets has been increasing, due to more species being measured, species being stratified by particle size, shorter durations of sampling, and measurements not only at the surface but also aloft. To take advantage of these richer data sets the receptor models have become more complex. To ensure that receptor modeling tools, both simple and complex, are available for use in the development and implementation of air quality standards, the United States Environmental Protection Agency's Office of Research and Development (ORD) has and continues to develop a suite of receptor models (multivariate statistical techniques) that are freely distributed to the air quality management community. The EPA has developed the Chemical Mass Balance (CMB) and UNMIX models as well as the Positive Matrix Factorization (PMF) method for use in air quality management. CMB fully apportions receptor concentrations to chemically distinct source-types depending upon the source profile database, while UNMIX and PMF internally generate source profiles from the ambient data. Details of CMB, UNMIX and PMF are described below.

The Chemical Mass Balance (CMB) Model (most recent version, EPA-CMBv8.2) is one of several receptor models that has been applied to air quality problems over the last two decades (U.S. EPA, 2004d and 2004e; Seigneur, 1997, Coulter, 2000). Based on an effective-variance least squares method (EVLS), EPA has supported CMB as a regulatory planning tool through its approval of numerous State Implementation Plans (SIPs) which have a source apportionment component. The chemical mass balance model 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. CMB requires speciated profiles of potentially contributing sources and the

corresponding ambient data from analyzed samples collected at a single receptor site. CMB is ideal for localized nonattainment problems and has proven to be a useful tool in applications where steady-state Gaussian plume models are inappropriate, as well as for confirming or adjusting emissions inventories.

UNMIX is named for its function, which is to "unmix" the concentrations of chemical species measured in the ambient air to identify the contributing sources. The particular mathematical approach used by UNMIX is based on a form of Factor Analysis, but its novelty is that physically-meaningful constraints are imposed which are intended to remove the undesirable ambiguity of the multiple solutions that are characteristic of ordinary Factor Analysis. For a given selection of species, UNMIX estimates the number of sources, the source compositions, and source contributions to each sample. Chemical profiles of the sources are not required, but instead are generated from the ambient data.

The PMF technique is a form of factor analysis where the underlying co-variability of many variables (e.g., sample to sample variation in PM species) is described by a smaller set of factors (e.g., PM sources) to which the original variables are related. The structure of PMF permits maximum use of available data and better treatment of missing and below-detection-limit values. Also available is a document which discusses the PMF methodology: A Guide to Positive Matrix Factorization (PDF).

18.5.4 Retrospective Analyses

A retrospective analysis is intended to examine the ability of the model to respond to emissions changes by comparing recent trends in observed ozone or $PM_{2.5}$ concentrations to the model-predicted trend over the same time period. The approach is a direct assessment of what we are most interested in---does the model accurately predict changes in air quality as a result of changes in emissions? As part of this analysis the model is run for current episodes or time periods and episodes in one or more historical time periods using the emissions and meteorological inputs appropriate for each time period modeled. While retrospective analyses may be useful, it may be difficult to obtain meteorological and emissions inputs for the historical time period(s) that are calculated using techniques and assumptions which are consistent with the calculation of these same inputs for the current time period. Using inconsistent inputs will confound the interpretation of the predicted trend.

Because differences in meteorology between years can confound the apparent change in pollutants, EPA recommends that, in most cases, retrospective analyses use constant meteorology and backcasted emissions. The model should respond in a predictable way if the emissions changes are large enough. Hence, backcasting an emissions change of only a few percent would not be viable. However, if NOx or SO2 emissions have been reduced by a large percentage (e.g. 30% or more), then that may be a good test of the response of the model. In Section 8, we noted that a retrospective analysis can be a useful tool for diagnosing why an area has not attained the NAAQS. Therefore, retrospective analyses are one of the few tools that can be used to determine

if the model is responding "adequately" to control measures. Retrospective analyses may be an important tool to show that the models are responding "correctly" to emissions changes.

18.5.5 Alternative Base Cases (Sensitivity tests)

In some cases it may be useful to evaluate how the response of the model to emissions reductions varies as a function of alternative model inputs or model algorithms. Outcomes of these types 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 (Arnold, 2003; Gilliland, 2006; Morris, 2006). Second, sensitivity tests may help prioritize additional data gathering efforts so that a subsequent review/diagnosis can be performed at the time of a mid-course review or required attainment date. Third, the tests can be used to assess the robustness of a control strategy. As an example, States/Tribes could consider the effects of assumed boundary conditions on predicted effectiveness of a control strategy. If the model response does not differ greatly over a variety of alternative plausible configurations, this increases confidence in the model results.

The parameters for sensitivity tests can include, but are not limited to: different chemical mechanisms, finer or coarser grid resolution, meteorological inputs from alternative, credible meteorological model(s), different initial/boundary conditions, and multiple sets of reasonable emission projections. Sensitivity tests can and should be applied throughout the modeling process, not just when model performance is being evaluated. In cases where the operational model performance is considered to be poor, these tests may help indicate where base case input/algorithm changes are warranted.

A new technique has been developed to more efficiently examine a large range of emissions control scenarios. Response Surface Modeling (RSM), has been developed by utilizing advanced statistical techniques to characterize the relationship between model outputs and input parameters in a highly economical manner (U.S. EPA, 2006c). The RSM is a metamodel of a model (i.e., air quality model); it is a reduced-form prediction model using statistical correlation structures to approximate model functions through the design of complex multi-dimension experiments.

The RSM is based on a new approach known as air quality metamodeling that aggregates numerous pre-specified air quality modeling simulations into a multi-dimensional air quality "response surface". Simply, this metamodeling technique is a "model of the model" and can be shown to reproduce the results from an individual modeling simulation with little bias or error (U.S. EPA, 2006c, Hubbell, 2005). The RSM is based on statistical relationships between model inputs and outputs to provide real-time estimate of these air quality changes. The RSM provides a wide breadth of model outputs, which can be used to develop emissions control scenarios. The RSM approach can inform the selection and evaluation of various control scenarios. This approach allows for the rapid assessment of air quality impacts and changes of different combinations of emissions reductions. While the RSM may not provide a complete picture of all changes necessary to reach various alternative standards nationwide, it is highly useful in the

context of providing illustrative control scenarios for selected areas, and understanding the contribution of different source categories, source regions and pollutant emissions to air quality across the U.S. The RSM can be used in a variety of ways: (1) strategy design and assessment (e.g. comparison of urban vs. regional controls; comparison across sectors; comparison across pollutants); (2) optimization (develop optimal combinations of controls to attain standards at minimum cost); (3) model sensitivity (systematically evaluate the relative sensitivity of modeled ozone and PM levels to changes in emissions inputs.

18.6 How Should the Results of the Model Evaluation be Assessed (Interpreted)?

In EPA guidance for the 1-hour ozone attainment demonstrations (U.S. EPA, 1991a), several statistical goals were identified for operational model performance. These goals were identified by assessing past modeling applications of ozone models and determining common ranges of bias, error, and accuracy (Tesche et al., 1990). The 1-hour guidance noted that because of differences in the quality of the applications considered, it was inappropriate to establish "rigid criterion for model acceptance or rejection" (i.e., no pass/fail test). It was recommended that these ranges should be used in conjunction with the additional qualitative procedures to assess overall model performance.[51]

With the additional experience of more than a decade of photochemical modeling, it is clear that there is no single definitive test for evaluating model performance. All of the tests identified in Sections 18.2 and 18.3 have strengths and weaknesses. Further, even within a single performance test, it is not appropriate to assign “bright line” criteria that distinguish between adequate and inadequate model performance. In this regard, EPA recommends that a “weight of evidence” approach (like that described in Section 7) be used to determine whether a particular modeling application is valid for assessing the future attainment status of an area. EPA recommends that States/Tribes undertake a variety of performance tests and weigh them qualitatively to assess model performance. Provided suitable data bases are available, greater weight should be given to those tests which assess the model capabilities most closely related to how the model is used in the modeled attainment test. Generally, additional confidence should be attributed to model base case applications in which a variety of the tests described above are applied and the results indicate that the model is performing well. From an operational standpoint, EPA recommends that States/Tribes compare their evaluation results against similar modeling exercises to ensure that the model performance approximates the quality of other applications. To aid in this comparison, we have summarized performance metrics (including information on models and notes on inputs) of recent ozone, $PM_{2.5}$, and regional haze applications in Appendix B.

---

[51] In practice, however, most 1-hour ozone modeling applications using the 1991 guidance tended to focus almost entirely on meeting the three statistical "goals" for bias, error, and accuracy at the expense of more diagnostic evaluation.

In summary, we continue to believe that for ozone, $PM_{2.5}$, and regional haze; 1) there should be no bright line performance criteria; 2) the evaluation of statistical performance measures should be compared to current and past modeling applications; and 3) statistical measures should be used in conjunction with diagnostic tests and other qualitative analyses to assess overall model performance.

REFERENCES

Abt Associates, (2003), "Environmental Benefits Mapping and Analysis Program (BenMAP), User's Manual: http://www.epa.gov/ttn/ecas/benmodels.html.

ABT Associates, (2007), "Model Attainment Test Software (MATS) User's Manual".

Air Quality Management Work Group, (2005), "Recommendations to the Clean Air Advisory Committee: Air Quality Management Work Group, Phase 1 and Next Steps", available at http://www.epa.gov/air/caaac/pdfs/report1-17-05.pdf.

Ansari, A.S., and S.N. Pandis, (1998), "Response of Inorganic PM to Precursor Concentrations", *Environ. Sci. Technol.*, **32**, 2706-2714.

Appel, W., S. Howard, R. Gilliam, and S. Phillips, (2005), "The Atmospheric Model Evaluation Tool: Air Quality Module", 4th Annual CMAS Models-3 Users' Conference, September 26-28, 2005, Chapel Hill, NC.

Arnold, J.R., R.L. Dennis, and G.S. Tonnesen, (2003), "Diagnostic Evaluation of Numerical Air Quality Models with Specialized Ambient Observations; Testing the Community Multiscale Air Quality Modeling System (CMAQ) at Selected SOS 95 Ground Sites", *Atmospheric Environment*, **37**, 1185-1198.

Arunachalam, S., Z. Adelman, R. Mathur, and J. M. Vukovich, (2002), "The Impacts of Biogenic Emissions Estimates from BEIS-3 on Ozone Modeling in the Southeastern U.S.", 11th Annual Emissions Inventory Conference of the U.S. EPA", http://www.epa.gov/ttn/chief/conference/ei11/index.html

Arunachalam, S., A. Holland, B. Do, and M. Abraczinskas, (2006), A Quantitative Assessment of the Influence of Grid Resolution on Predictions of Future-Year Air Quality In North Carolina, USA, *Atmospheric Environment*, **40** (26), 5010-5026.

Atmospheric and Environmental Research, Inc., (2002), "Evaluation of Probing Tools Implemented in CAMX", CRC Project A-37-1, Document Number CP099-1-2, August, http://www.crcao.com/reports/recentstudies00-02/A-37-1%20Final%20Report.pdf

Baker, K., (2004a), "Midwest Regional Planning Organization Modeling Protocol", Lake Michigan Air Directors Consortium / Midwest Regional Planning Organization, Des Plaines IL. http://www.ladco.org/tech/photo/docs/RPOModelingProtocolV2_dec2004.pdf

Baker K., (2004b), "Meteorological Modeling Protocol For Application to $PM_{2.5}$/Haze/OzoneModeling Projects", Lake Michigan Air Directors Consortium / Midwest Regional Planning Organization, Des Plaines IL.

Battelle, 2004, “Characterizing Ozone Episodes to Inform Regulatory Modeling”, Technical Report, U.S. EPA Contract No. 68-D-02-061, Work Assignment 2-07

Biswas, J., and S.T. Rao, (2001), “Uncertainties in Episodic Ozone Modeling Stemming from Uncertainties in the Meteorological Fields”, *J. Applied Met*., **40**, 117-136

Blanchard, C.L., and P.M. Roth, (1997), “User’s Guide Ozone M.A.P.P.E.R., Measurement-based Analysis of Preferences in Planned Emission Reductions, Version 1.1", Final Report prepared for the U.S. EPA pursuant to Contract 68-D6-0064 Work Assignment, W.M.Cox, EPA Work Assignment Manager.

Blanchard, C.L., F.W. Lurmann, P.M. Roth, H.E. Jeffries and M. Korc, (1999), “The Use of Ambient Data to Corroborate Analyses of Ozone Control Strategies”, *Atmospheric Environment*, **33**, 369-381.

Blanchard, C. L., (2000), “Ozone process insights from field experiments- Part III: Extent of reaction and ozone formation”, *Atmospheric Environment*, **34**, 2035-2043.

Blanchard, C. L., and T. Stoeckenius, (2001), “Ozone response to precursor controls: comparison of data analysis methods with the predictions of photochemical air quality simulation models”, *Atmospheric Environment*, **35**, 1203-1216.

Blanchard, C.L., and S.t. Tanenbaum, (2003), “Differences between Weekday and Weekend Air Pollutant Levels in South California”, *J. Air & Waste Management Association*, **53**, 816-828.

Blanchard, C.L., and S.T. Tanenbaum, (2005), “Weekday/Weekend Differences in Ambient Ozone and Particulate Matter Concentrations in Atlanta and the Southeastern United States”, Envair, Final Report, CRC Project No. A-47, pp. 78.

Boylan, J.W. and Russell, A.G, (2006), “PM and Light Extinction Model Performance Metrics, Goals, and Criteria for Three-Dimensional Air Quality Models”, *Atmos. Environ.*, **40**, 4946-4959.

Brook, J., E. Vega, and J. Watson, (2004), “Receptor Methods” Chapter 7, In: *Particulate Matter Science for Policy Makers, A NARSTO Assessment*, P.H. McMurry, M.F. Shepherd, J.S. Vickery, ed., Cambridge University Press, Cambridge, UK, 235-281.

Byun, D.W., and J.K.S. Ching, (1999), “Science Algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) Modeling System,” Office of Research and Development, USEPA, EPA/600/R-99/030.

CARB, (1996), “Performance Evaluation of SAQM in Central California and Attainment Demonstration for the August 3-6, 1990 Ozone Episode”, California Air Resources Board, Sacramento, CA.

Cabada, J.C., S. N. Pandis, R. Subramanian., A. L. Robinson, A. Polidori, and B. Turpin, (2004), “Estimating the Secondary Organic Aerosol Contribution to $PM_{2.5}$ Using the EC Tracer Method”, *Aerosol Science and Technology*, 38, 140-155.

Cardelino, C.A., and W.L. Chameides, (1995), “An Observation Based Model for Analyzing Ozone Precursor Relationships in the Urban Atmosphere”, *J. Air & Waste Mgmt. Assoc.*, **45**, 161-181.

Chang, J.S., S. Jin, Y. Li, M. Beauharnois, C.L. Lu, and H. Huang, (1997), “SARMAP Air Quality Model”, prepared by Atmospheric Sciences Research Center, State University of New York (Albany).

Chinkin, L.R., D.L. Coe, T.H. Funk, H.R. Hafner, P.T. Roberts, P.A. Ryan, and D.R. Lawson, (2003), “Weekday versus Weekend Activity Patterns for Ozone Precursor Emissions in California’s South Coast Air Basin”, *J. Air & Waste Management Association*, **53**, 829-843.

Chow, J. C., J. G. Watson, L.-W. A. Chen, W.P. Arnott, and H. Moosmüller, (2004), “Equivalence of Elemental Carbon by Thermal/Optical Reflectance and Transmittance with Different Temperature Protocols”, *Environmental Science and Technology*, 38(16), 4414-4422.

Chu, S.-H., (2005), “Stable estimate of primary OC/EC ratios in the EC tracer method”, *Atmospheric Environment*, 39, 1383–1392.

Clegg, S.L., P. Brimblecombe, and A. S. Wexler, (1998), “Aerosol Inorganics Model; A Thermodynamic Model of the System H+ - NH4+ - SO42- - NO3- - H2O at Tropospheric Temperatures.”, *J. Phys. Chem*., 102A, 2137-2154. http://www.aim.env.uea.ac.uk/aim/project.htm

Clinton, W.J., (July 16, 1997), Subject: “Implementation of Revised Air Quality Standards for Ozone and Particulate Matter”, Memorandum to the Administrator of the Environmental Protection Agency.

Croes, B.E., L.J. Dolislager, L.C. Larsen, and J.N. Pitts, (2003), “The O3 'Weekend Effect' and NOx Control Strategies—Scientific and Public Health Findings and Their Regulatory Implications”; EM, July 2003, 27-35.

Cohan, D.S., (2002), “Implementation of a Direct Sensitivity Method into CMAQ”, Paper 5.3, Presented at 1st CMAS Models-3 User’s Conference - Models-3 Performance and Applications, RTP, NC, (October 2002).

Cohan, D.S., (2004), “Applicability of CMAQ-DDM to Source Apportionment and Control Strategy Development”, Paper 5.3, Presented at 3rd Annual CMAS Models-3 Users’ Conference, RTP, NC, (October 2004).

Cohan, D. S., A. Hakami, Y.T. Hu, and A.G. Russell, (2005), “Nonlinear Response of Ozone to Emissions: Source Apportionment and Sensitivity Analysis.” *Environ. Sci. Technol.* **39**, 6379-6748.

Cohan, D.S., Y. Hu, A.G. Russell, (2006), “Dependence of ozone sensitivity analysis on grid resolution”, *Atmospheric Environment*, 40, 126-135.

Collett, Jeffrey, (2004), Technical note on IMPROVE study.

Coulter, T., R.A. Wagoner, and C.W. Lewis, (2000), “Chemical Mass Balance Software: EPA-CMB8.2", Submitted for inclusion in Proceedings of the AWMA/EPA Symposium, Measurement of Toxic and Related Air Pollutants, September 12-14, 2000, Research Triangle Park, NC.

Cox, W.M., and S. Chu, (1993), “Meteorologically Adjusted Ozone Trends in Urban Areas: A Probabilistic Approach”, *Atmospheric Environment*, **27B**, (4), 425-434.

Cox, W.M., and S. Chu, (1996), “Assessment of Interannual Ozone Variation in Urban Areas from a Climatological Perspective”, *Atmospheric Environment*, **30**, 2615-2625.

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, (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).

Dolwick, P.D., (2002), "USEPA/OAQPS Meteorological Modeling Analyses", presented at the 2002 Ad Hoc Meteorological Modeling Meeting, Des Plaines IL.

Doty. K.G., T.W. Tesche, D.E. McNally, B. Timin, and S. F. Mueller, (2001), “Meteorological Modeling for the Southern Appalachian Mountains Initiative (SAMI)”, Final Report, Southern Appalachian Mountain Initiative. Asheville, NC.

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).

Douglas, S., T. Myers, and Y. Wei, (2006). “Implementation of Sulfur and Nitrogen Tagging in the Community Multi-scale Air Quality (CMAQ) Model.” Prepared for the U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, North Carolina. Prepared by ICF International, San Rafael, California (06-076).

Dunker, A.M., (1980), “The response of an atmospheric reaction-transport model to changes in input functions”, *Atmospheric Environment*, **14**, 671-679.

Dunker, A.M., (1981), “Efficient calculations of sensitivity coefficients for complex atmospheric models”, *Atmospheric Environment*, **15**, 1155-1161.

Dunker, A.M., (1984), “The Decoupled Direct Method For Calculating Sensitivity Coefficients In Chemical-Kinetics”, *J. Chem. Phys*. **81**, 2385-2393.

Dunker, A.M., G. Yarwood, J.P. Ortmann, and G.M. Wilson, (2002a), "Comparison of source apportionment and source sensitivity of ozone in a three-dimensional air quality model", *Environ. Sci. Technol.*, **36**, 2953-2964.

Dunker, A.M., G. Yarwood, J.P. Ortmann, and G.M. Wilson, (2002b), "The decoupled direct method for sensitivity analysis in a three-dimensional air quality model–implementation, accuracy, and efficiency", *Environ. Sci. Technol.*, **36**, 2965-2976.

Eberly, S., (2005), "EPA PMF 1.1 User's Guide", June 30, 2005. USEPA, National Exposure Research Laboratory, http://www.epa.gov/heasd/products/pmf/pmf.htm .

Eder, B.K., J.M. Davis, and P. Bloomfield, (1993a), “A Characterization of the Spatiotemporal Variability of Non-urban ozone concentrations over the Eastern United States”, *Atmospheric Environment*, **27A**, No. 16, 2645-2668.

Eder, B.K., J.M. Davis, and P. Bloomfield, (1993b), “An automated Classification Scheme Designed to Better Elucidate the Dependence of Ozone on Meteorology”, *J. Applied Met*., **33**, 1182-1198.

Edgerton E.S., B.E. Hartsell, R.D. Saylor, J.J. Jansen, D.A. Hansen, and G.M. Hidy, (2005), “The Southeastern Aerosol Research and Characterization Study: Part II. Filter-based

measurements of fine and coarse particulate matter mass and composition", *J. Air Waste Mange. Assoc.*, **55,** 1527-1542.

Emery, C., and E. Tai, (2001), "Enhanced Meteorological Modeling and Performance Evaluation for Two Texas Ozone Episodes ", prepared for the Texas Near Non-Attainment Areas through the Alamo Area Council of Governments", by ENVIRON International Corp, Novato, CA., http://www.tnrcc.state.tx.us/air/aqp/airquality_contracts.html#met01

ENRIRON International Corporation, (2004), "Annual PM Modeling and Analysis of Episode Aggregation Uncertainties for Annual Average $PM_{2.5}$ and Visibility Modeling", CRC Project A-44a/b, http://www.crcao.com/ .

ENVIRON/Alpine/UCR, (2005), "White Paper- Visibility Projection Approaches- Preliminary Approaches for VISTAS 2018 Modeling".

ENVIRON International Corporation, (2006a), "User's Guide to the Comprehensive Air Quality Model with Extensions (CAMx) version 4.30", Novato, CA. Http://www.camx.com .

ENVIRON International Corporation, (2006b), "Guidance for the Application of the CAMx Hybrid Photochemical Grid Model to Assess Visibility Impacts of Texas BART Sources at Class I Areas", Novato, CA. Prepared for Texas Commission on Environmental Quality (TCEQ) (http://www.tceq.state.tx.us/assets/public/implementation/air/sip/bart/BART_TX_CAMx_PSAT_Guid.pdf).

Frank, N., 2006: "Retained Nitrate, Hydrated Sulfates, and Carbonaceous Mass in Federal Reference Method Fine Particulate Matter for Six Eastern U.S. Cities" *J. Air Waste Mange. Assoc.*, 56, 500-511.

Fiore, A.M., D.J. Jacob, H. Liu, R.M. Yantosca, T.D. Fairlie, and Q. Li,(2003), "Variability in surface ozone background over the United States: Implications for air quality policy", *J. Geophys. Res.*, 108, 4787, 2003.

Flaum, J. B., S.T. Rao, and I.G. Zurbenko, (1996), "Moderating the influence of meteorological conditions on ambient ozone concentrations", *Journal of the Air and Waste Management Association* **46**, 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**, 9107-9119.

Gego, E., H. Hogreffe, G. Kallos, A. Voudouri, J. Irwin, and S. T. Rao, (2005), "Examination of Model Predictions at Different Horizontal Grid Resolutions", *Environmental Fluid Mechanics*, 5, 63-85.

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**, 12,773-12,791.

Gilliam R., (2005), The Atmospheric Model Evaluation Tool: Meteorology Module, 4th Annual CMAS Models-3 Users' Conference, September 26-28, 2005, Chapel Hill, NC.

Gilliland, A., K.W. Appel, R.W. Pinder, and R. Dennis, (2006), "Seasonal NH3 Emissions for the Continental United States: Inverse Model Estimation and Evaluation", *Atmospheric Environment*, **40** (26), 4986-4998.

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**, 2205-2230.

Hakami, A., Odman M. T., and Russell, A. G., (2003), "High-order, direct sensitivity analysis of multidimensional air quality models" *Environ. Sci. Technol*, **37**, 2442-2452.

Hakami, A., Odman, M. T., and Russell, A. G., (2004), "Nonlinearity in atmospheric response: A direct sensitivity analysis approach",. *J. Geophys. Res.*, **109**, D15303.

Hand, J.L. and W.C. Malm, (2006), "Review of the IMPROVE Equation for Estimating Ambient Light Extinction Coefficients- Final Report", March 2006, http://vista.cira.colostate.edu/improve/Publications/GrayLit/016_IMPROVEeqReview/IMPROVEeqReview.htm

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.

Hansen, D.A., E.S. Edgerton, B.E. Hartsell, J.J.Jansen, N. Kandasamy, G.M. Hidy, and C.L. Blanchard, (2003), "The Southeastern Aerosol Research and Characterization Study: Part 1 – Overview", *J. Air Waste Mange. Assoc.*, **53,** 1460-1471.

Henderson, B., W. Vizuete, and H. Jeffries, (2006), "Development of a Lagrangian pyPA (Process Analysis in python) tool", Presented at the 5th Annual CMAS Conference, October 16-18, 2006, http://www.cmascenter.org/conference/2006/agenda.cfm

Henry, R.C., (1991), "Multivariate Receptor Models", In: *Receptor Modeling for Air Quality Management*, P.K. Hopke, ed., Elsevier Science Publishers, Amsterdam, 117-147.

Hering, S. and G. Cass, (1999), "The Magnitude of Bias in Measurement of $PM_{2.5}$ Arising from Volatilization of Particulate Nitrate from Teflon Filters." *JAWMA* 49:725-733.

Hogrefe, C., S.T. Rao, I.G. Zurbenko, and P.S. Porter, (2000), "Interpreting the information in time series of ozone observations and model predictions relevant to regulatory policies in the eastern United States", *Bull. Amer. Met. Soc.*, **81**, 2083-2106.

Holland, D.M., W.M. Cox, R. Scheffe, A.J. Cimorelli, D. Nychka, and P.K. Hopke, (2003), "Spatial Prediction of Air Quality Data", *EM, August 2003*, 31-35.

Hopke, P.K., ed. (1991), "Receptor Modeling for Air Quality Management", Elsevier Science, Amsterdam.

Hopke, P.K., ed. (1999), "An Introduction to Source Receptor Modeling", In: *Elemental Analysis of Airborne Particles*, Landsberger, S., Creatchman, M., Eds. Gordon and Breach Science, Amsterdam, 273-315.

Hopke, P.K., ed. (2001), "Advances in Receptor Modeling", *Journal of Aerosol Science*, **32**, 363-368.

Houyoux, M.R., J.M. Vukovich, C.J. Coats Jr., N.W. Wheeler, and P.S. Kasibhatla, (2000), "Emission Inventory Development and Processing for the Seasonal Model for Regional Air Quality (SMRAQ) project." *Journal of Geophysical Research, Atmospheres* 105:D7:9079-9090.

Hu, Y., D. Cohen, M.T. Odman, and T. Russell, (2004), "Fall Line Air Quality Study- Draft Final Report- Air Quality Modeling of the August 11-20, 2000 Air Pollution Episode."

Huang J-P., J. Fung, Y. Zhang, A. Lau, R. Kwok, and J. Lo, 2005: Improvement of Air Quality Modeling in Hong Kong by Using MM5 Coupled with LSM, 4th Annual CMAS Models-3 Users' Conference, September 26-28, 2005, Chapel Hill, NC.

Hubbell, B.J., P.D. Dolwick, D. Mooney, and M. Morara, (2005), "Evaluating the relative effectiveness of ozone precursor controls: Design of computer experiments applied to the comprehensive air quality model with extensions (CAMx)", *Air and Waste Management Association Conference Proceedings*

IMPROVE, (2000), "Spatial and Seasonal Patterns and Temporal Variability of Haze and its Constituents in the United States: Report III." Cooperative Institute for Research in the Atmosphere, ISSN: 0737-5352-47. IMPROVE Web Site: http://vista.cira.colostate.edu/improve.

IMPROVE, (2006), “Revised IMPROVE Algorithm for Estimating Light Extinction from Particle Speciation Data”, January 2006, http://vista.cira.colostate.edu/improve/Publications/GrayLit/gray_literature.htm

Isakov, V., J.S. Irwin, and J. Ching, (2007), Using CMAQ for exposure modeling and characterizing the sub-grid variability for exposure estimates”, *Journal of Applied Meteorology and Climatology,* (in press).

Jacob, D.J., J. A. Logan, and P.P. Murti, (1999), "Effect of Rising Asian Emissions on Surface Ozone in the United States", *Geophy. Res. Lett.*, 26, 2175-2178.

Jaffe D., McKendry I., Anderson T., and Price H.,(2003), “Six 'new' episodes of trans-Pacific transport of air pollutants”, *Atmos. Envir.* 37, 391-404.

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), 3101-3114.

Jeffries, H.E., (1994), “Process Analysis for UAM Episode 287", memo to Brock Nicholson, NC DEHNR, April 8, 1994, ftp://airsite.unc.edu/pdfs/ese_unc/jeffries/projprpt/uammod, panel287.pdf.

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, ftp://airsite.unc.edu/pdfs/ese_unc/jeffries/reports/gri, 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, ftp://airsite.unc.edu/pdfs/ese_unc/jeffries/ipradocs, sourcewksp.pdf, and http://www.epa.gov/ttn/faca/stissu.html, Source Att WS-Proc.analysis/mass track--Jeffries.

Johnson, M., (2003), "Iowa DNR 2002 Annual MM5 Modeling Project", presented at the 2003 Ad Hoc Meteorological Modeling Meeting, Des Plaines IL.

Jones, J. M., C. Hogrefe, R. Henry, J. Ku, and G. Sistla, (2005), “An Assessment of the Sensitivity and Reliability of the Relative response factor Approach in the Development of 8-hr Ozone Attainment Plans”, JAWMA, 55 (1), 13-19.

Kenski, D., (2004), “Analysis of Historic Ozone Episodes using CART”, http://www.ladco.org/tech/monitoring/docs_gifs/CART%20Analysis%20of%20Historic%20Ozone%20Episodes.pdf

Koerber, M. and D. Kenski, (2005), "The Case for Using Weight of Evidence Demonstrations in State Implementation Planning"; EM, April 2005, 24-28.

Ku, J., H. Mao, K. Zhang, K. Civerolo, S.T. Rao, C.R. Philbrick, B. Doddridge and R. Clark, (2001), "Numerical Investigation of the Effects of Boundary-Layer Evolution on the Predictions of Ozone and the Efficacy of Emission Control Options in the Northeastern United States/Tribes", *Environmental Fluid Mechanics*, **1**, 209-233

Kumar, N., M.T.Odman, and A.G. Russell, (1994), "Multiscale Air Quality Modeling: Application to Southern California", *J. Geophysical Research*, **99**, 5385-5397.

Kumar, N., and A.G. Russell, (1996), "Multiscale Air Quality Modeling of the Northeastern United States", *Atmospheric Environment*, **30**, 1099-1116.

LADCO, (1999), Personal communication.

Lefohn, A.S., D.S. Shadwick, and S.D. Ziman, (1998), "The Difficult Challenge of Attaining EPA's New Ozone Standard", *Environmental Science and Technology, Policy Analysis*, **32**, No.11, (June 1, 1998), 276A-282A.

Lehman, J., K. Swinton, S. Bortnick, C. Hamilton, E. Baldridge, B. Eder and B. Cox, (2004), "Spatial-Temporal characterization of Tropospheric Ozone Across the Eastern United States", *Atmospheric Environment*, **38:26**, 4357-4369, www.sciencedirect.com

Lewis, C.W., G.A. Norris, R.C. Henry, and T.L. Conner, (2003), "Source Apportionment of Phoenix PM-2.5 Aerosol with the Receptor Model", *Journal of the Air & Waste Management Association*, **53**, 325-338.

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), ftp://airsite.unc.edu/pdfs/ese_unc/jeffries/ipradocs, 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**, 3453-3462.

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.

Maykut, Lewtas, Kim, and Larson, (2003), "Source Apportionment of $PM_{2.5}$ at an Urban IMPROVE Site in Seattle, Washington", *Environmental Science and Technology*, **37**, 5135-5142. (This paper uses UNMIX and PMF)

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.

MCNC, (1999), MAQSIP air quality model, http://www.iceis.mcnc.org/products/maqsip/.

McNally, D.E., (2002), "A Comparison of MM5 Model Estimates for February and July 2001 Using Alternative Model Physics", report prepared for the United States Environmental Protection Agency, prepared by Alpine Geophysics, LLC, Arvada CO.

McQueen, J., P. Lee, M. Tsidulko, G. DiMego, R. Mathur, T. Otte, J. Pleim, G. Pouliot, D.Kang, K. Schere, J. Gorline, M. Schenk, and P. Davidson, (2004), "Update of The Eta-CMAQ Forecast Model run at NCEP operations and its performance for the Summer 2004", presented at the 2004 CMAS Workshop, Chapel Hill, NC.

Milanchus, M. L., S.T. Rao, and I.G. Zurbenko, (1998), "Evaluating the effectiveness of ozone management efforts in the presence of meteorological variability", *Journal of the Air and Waste Management Association* **48,** 201-215.

Morris R.E., G. Yarwood, C.A. Emery, and G.M. Wilson, (2001), "Recent Advances in CAMx Air Quality Modeling", Presented at the 2001Air and Waste Management Association Annual Conference, http://www.camx.com/publ/pdfs/camx934_AWMA_2001.pdf

Morris R.E., G. Yarwood, C.A. Emery, and G.M. Wilson, (2003), "Recent Advances in Photochemical Air Quality Modeling Using the CAMx Model: Current Update and Ozone Modeling of Point Source Impacts", Presented at the 2003Air and Waste Management Association Annual Conference, www.camx.com/publ/pdfs/camx_advances_AWMA_2002_rev3.pdf

Morris, R.E., B. Koo, A.Guenther, G. Yarwood, D. McNally, T.W. Teshe, G. Tonnesen, J. Boylan, and P. Brewer, (2006), "Model Sensitivity Evaluation for Organic Carbon using Two Multi-pollutant Air Quality Models that Simulate Regional Haze in the Southeastern United States", *Atmospheric Environment*, **40** (26), 4960-4972.

Mukerjee, S., G.A. Norris, L.A. Smith, C.A Noble, L.M. Neas, H.A. Özkaynak, and M. Gonzales, (2004), "Receptor Model Comparisons and Wind Direction Analyses of Volatile Organic Compounds and Submicrometer Particles in an Arid, Binational, Urban Air Shed", *Enviromental Science and Technology*, **38(8)**, 2317-2327.

Na, K. S., A.A. Sawant, C. Song, and D.R. Cocker, (2004), "Primary and secondary carbonaceous species in the atmosphere of Western Riverside County, California", *Atmospheric Environment*, 38, 1345–1355.

NADP, (2002), National Acid Deposition Program 2002 Annual Summary.

Nielsen-Gammon, John W., (2002), "Evaluation and Comparison of Preliminary Meteorological Modeling for the August 2000 Houston-Galveston Ozone Episode (Feb. 5, 2002 interim report)", http://www.met.tamu.edu/results/.

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.

Olerud, D., K. Alapaty, and N. Wheeler, (2000), "Meteorological Modeling of 1996 for the United States with MM5", MCNC-Environmental Programs, Research Triangle Park, NC.

Olerud, D., and A. Sims, (2003), "MM5 Sensitivity modeling in support of VISTAS", prepared for Mike Abraczinskas of the VISTAS Technical Analysis Workgroup by Baron Advanced Meteorological Systems, LLC, Raleigh NC.

Otte, T., (2004), "What's New in MCIP2", presented at the 2004 CMAS Workshop, Chapel Hill, NC.

Pacific Environmental Services, Inc., (1997), "Draft Technical Memorandum Analysis of 8-hour Ozone Values: 1980-1995", prepared under EPA Contract No.68D30032, Work Assignment III-88, Edwin L. Meyer, Jr., Work assignment manager.

Pielke, R.A., W.R. Cotton, R.L. Walko, C.J. Tremback, W.A. Lyons, L.D. Grasso, M.E. Nicholls, M.D. Moran, D.A. Wesley, T.J. Lee, and J.H. Copeland, (1992), "A Comprehensive Meteorological Modeling System - RAMS", *Meteor. Atmos. Phys.*, **49**, 69-91.

Pierce, T., C. Geron, G. Pouliot, J. Vukovich, and E. Kinnee, (2004), "Integration of the Biogenic Emissions Inventory System (BEIS3) into the Community Multiscale Air Quality (CMAQ) Modeling System", 12th Joint Conference on the Application of Air Pollutant Meteorology with the Air and Waste Management Association, http://ams.confex.com/ams/AFMAPUE/12AirPoll/abstracts/37962.htm.

Pitchford, M.L., and W.C. Malm, (1994), "Development and Applications of a Standard Visual Index", *Atmos. Environ*., **28**(5),1049-1054.

Poirot, Wishinski, Hopke, and Polissar, (2001), "Comparative Application of Multiple Receptor Methods to Identify Aerosol Sources in Northern Vermont", *Environmental Science and Technology*, **35**, 4622-4636 (This paper uses Unmix, PMF, and Ensemble back trajectories (CAPITA's Monte Carlo, PSCF, and RTA)).

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**, 4865-4875.

Pun, B.K., and C. Seigneur, (2001), "Sensitivity of Particulate Matter Nitrate Formation to Precursor Emissions in the California San Joaquin Valley", *Environ. Sci. Technol*., **35** (14), 2979-2987.

Pun, B.K, C. Seigneur, and W. White, (2001), "Data Analysis for a Better Understanding of the Weekday/Weekend Ozone and PM Differences", Final report prepared by Atmospheric and Environmental Research, Inc., San Ramon, CA, under CRC contract Number A-36B for the Coordinating Research Council, Alpharetta, GA, June 2001.

Pun, B.K., C. Seigneur, and W. White, (2003), "Day-of-Week Behavior of Atmospheric Ozone in Three U.S. Cities", *J. Air & Waste Management Association*, **53**, 789-801.

Qin, Y., G.S. Tonnesen, and Z. Wang, (2004), "Weekend/Weekday Differences of Ozone, Nox, CO, VOCs, PM_{10} and the Light Scatter during Ozone Season in Southern California", *Atmospheric Environment*, **38** (19), 3069-3087.

Rao, S. T., E. Zalewsky, E., and I.G. Zurbenko, (1995), "Determining temporal and spatial variations in ozone air quality", *Journal of the Air and Waste Management Association* **45**, 57-61.

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**, 1977-1988 .

Roth, P.M., S.D. Reynolds, and T.W. Tesche, (2005), "Air Quality Modeling and Decisions for Ozone Reduction Strategies", *Journal of the Air and Waste Management Association*, **55**, 1558-1573.

Russell, A., and R. Dennis, (2000), "NARSTO critical review of photochemical models and modeling", *Atmospheric Environment*, **34**, 2283-2324.

Saylor, R. D., E. S. Edgerton, and B. E. Hartsell, (2006), "Linear regression techniques for use in the EC tracer method of secondary organic aerosol estimation", *Atmospheric Environment*, 40, 7546-7556.

Saxena, P. and Hildemann, L, (1996), "Water Soluble Organics in Atmospheric Particles: A Critical Review of the Literature and Application of thermodynamics to Identify Candidate
Compounds", *J. Atmos. Chem*., 24:57-109.

Scire, J.S., R.J. Yamartino, G.R. Carmichael, and Y.S. Chang, (1989), “CALGRID: A Mesoscale Photochemical Grid Model”, Volume II: User’s Guide, Sigma Research Corp., Concord, MA.

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.

Scire, J.S., D.G. Strimaitis, R.J. Yamartino, (2000), "A User's Guide for the CALPUFF Dispersion Model", Jan 2000, http://www.src.com/calpuff/calpuff1.htm.

Seaman, N.L., (1997), “Use of Model-Generated Wind fields 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, http://www.epa.gov/ttn/faca/stissu.html, Source Attribution Workshop Materials.

Seaman, N.L., (2000), “Meteorological modeling for air quality assessments”, *Atmospheric Environment*, **34**, 2231-2259.

Seigneur, C., P. Pai, J. Louis, P. Hopke, and D. Grosjean, (1997), “Review of Air Quality Models for Particulate Matter”, Document Number CP015-97-1a , prepared for American Petroleum Institute.

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.

Sillman, S., (1995), “The Use of NOy, H_2O_2, and HNO_3 as Indicators for O3-NOx-ROG Sensitivity in Urban Locations”, *J. Geophys. Res.* **100**, 14,175-14,188.

Sillman, S., D. He, C. Cardelino, and R.E. Imhoff, (1997), “The Use of Photochemical Indicators to Evaluate Ozone-NOx-Hydrocarbon Sensitivity: Case Studies from Atlanta, New York and Los Angeles”, *J. Air and Waste Mgmt. Assoc.* , **47** (10), 1030-1040. (Oct. 1997)

Sillman, S., (1998), “Evaluating the Relation Between Ozone, NO_x and Hydrocarbons: The Method of Photochemical Indicators”, EPA/600R-98/022, http://www-personal.engin.umich.edu/~sillman/publications.htm.

Sillman, S., and D. He, (2002), “Some theoretical results concerning O3-NO_x-VOC chemistry and NOx-VOC indicators”, *J. Geophys. Res.*, **107**, D22, 4659, doi:10.1029/2001JD001123, 2002, http://www-personal.engin.umich.edu/~sillman/publications.htm.

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, Colorado State University, Fort Collins, CO.

Sistla, G., C. Hogrefe, W. Hao, J. Ku, R. Henry, E. Zalewsky, and K. Civerolo, (2004), "An Operational Assessment of the Application of the Relative response factors in the Demonstration of Attainment of the 8-Hour Ozone National Ambient Air Quality Standard", JAWMA, 54 (8), 950-959.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang and J. G. Powers, (2005), A Description of the Advanced Research WRF Version 2, http://www.wrf-model.org/wrfadmin/docs/arw_v2.pdf.

Smith, James H.(1998), "Hierarchical Emissions Inventory Development for Nested Photochemical Modeling, Part I: Conceptual Overview", Proceedings of the Air & Waste Management Association Specialty Conference, Emissions Inventory: Living in a Global Environment, New Orleans, La., December 1998.

Stehr, J., (2004), "How's the Air Up There? A Guide to Mid-Atlantic Air Quality." Presentaion at MANE-VU/MARAMA Science Meeting, January 27, 2004.

Strader, R., F. Lurmann, and S.N. Pandis, (1999), "Evaluation of secondary organic aerosol formation in winter", *Atmospheric Environment*, 33, 4849–4863.

Strum, M., G. Gipson, W. Benjey, R., M. Houyoux, C. Seppanen, and G. Stella, (2003), "The Use of SMOKE to Process Multipollutant Inventories - Integration of Hazardous Air Pollutant Emissions with Volatile Organic Compound Emissions," *12th International Emissions Inventory Conference, Session 10*, http://www.epa.gov/ttn/chief/conference/ei12/index.html.

Systems Applications International, (1996), "Users Guide to the Variable Grid Urban Airshed Model (UAMV)", SYSAPP 96-95/27r, http://uamv.saintl.com/.

Systems Applications Internation, Inc. (SAI), (2005), "Overview of the REMSAD Modeling System", November 15, 2005, see website: http://remsad.saintl.com/overview.htm

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., P. Georgopoulos, F.L. Lurmann and P.M. Roth,1990, "Improvement of Procedures for Evaluating Photochemical Models", Draft Final Report, California Air Resources Board,
Sacramento, CA.

Tesche, T. W., and D. E. McNally, (2001), "Evaluation of the MM5, RAMS, and SAIMM Meteorological Model for the 6-11 September 1993 COAST and 25 August-1 September 2000 TexAQS2000 Ozone SIP Modeling Episodes", report prepared for the Business

Coalition for Clean Air-Appeals Group, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Tesche T. W., D.E. McNally, and C. Tremback, (2002), "Operational evaluation of the MM5 meteorological model over the continental United States: Protocol for annual and episodic evaluation." Submitted to USEPA as part of Task Order 4TCG-68027015. (July 2002)

Tesche T.W., R. Morris, G. Tonnesen, D. McNally, J. Boylan, and P. Brewer, (2006), CMAQ/CAMx annual 2002 performance evaluation over the eastern US",Atmospheric Environment, **40**, 4906-4919.

Timin, B., (2005a), "Draft Final Ozone Guidance Comments and Proposed Changes", presented at the 3rd PM/RH/O3 Modeling Workshop, New Orleans, LA, pages 25-27. http://www.cleanairinfo.com/modelingworkshop/presentations/O3_Guidance_Timin.pdf .

Timin, B., (2005b), "Draft Final Ozone Guidance Comments and Proposed Changes", presented at the 3rd PM/RH/O3 Modeling Workshop, New Orleans, LA, pages 8-22. http://www.cleanairinfo.com/modelingworkshop/presentations/O3_Guidance_Timin.pdf .

Tonnesen, G., and B. Wang, (2004), "CMAQ Tagged Species Source Apportionment", July 22, 2004; http://www.wrapair.org/forums/aoh/meetings/040722/UCR_tssa_tracer_v2.ppt

Turpin, B.J., and J.J Huntzicker, (1995),." Identification of secondary organic aerosol episodes and quantitation of primary and secondary organic aerosol concentrations during SCAQS", *Atmospheric Environment*, 29, 3527–3544.

Turpin, B. J. and H-J Lim, (2001), "Species Contributions to $PM_{2.5}$ Mass Concentrations: Revisiting Common Assumptions for Estimating Organic Mass." *Aerosol Science and Technology* 35: 602-610.

U.S. EPA, (1991a), "Guideline for Regulatory Application of the Urban Airshed Model", EPA-450/4-91-013, http://www.epa.gov/ttn/scram/guidance_sip.htm.

U.S. EPA, (1991b), "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, www.epa.gov/ttn/chief/software.html#speciate

U.S. EPA, (1994a), Office of Transportation and Air Quality (OTAQ), http://www.epa.gov/otaq/models.htm, MOBILE Model: MOBILE5, MOBILE6

U.S. EPA, (1994b), "Guidance on Urban Airshed Model (UAM) Reporting Requirements for Attainment Demonstration", EPA-454/R-93-056.

U.S. EPA, (1995a), "User's Guide for the Industrial Source Complex (ISC3) Dispersion Models: Volume I - User Instructions", Office of Air Quality Planning and Standards, Emissions, Monitoring, and Analysis Division; September 1995; EPA-454/B-95-003a, RTP, NC.

U.S. EPA, (1995b), "User's Guide for the Industrial Source Complex (ISC3) Dispersion Models: Volume II - Description of Model Algorithms", Office of Air Quality Planning and Standards, Emissions, Monitoring, and Analysis Division; September 1995, EPA-454/B-95-003b, RTP, NC.

U.S. EPA, (1996a), "Guidance on Use of Modeled Results to Demonstrate Attainment of the Ozone NAAQS", http://www.epa.gov/ttn/scram/guidance_sip.htm, EPA-454/B-95-007.

U.S. EPA, (1997a): 40 CFR Part 50 National Ambient Air Quality Standards for Particulate Matter; Final Rule, Federal Register / Vol. 62, No. 138 / Friday, July 18, 1997. http://www.epa.gov/ttn/amtic/files/cfr/recent/pmnaaqs.pdf

U.S. EPA, (1997b), "Guidance For Network Design and Optimum Site Exposure For $PM_{2.5}$ and PM_{10}", OAQPS, Research Triangle Park, NC, EPA report no. EPA-454/R-99-022.

U.S. EPA, (1997c), "EIIP Volume I, Introduction and Use of EIIP Guidance for Emission Inventory Development", July 1997, EPA-454/R-97-004a.

U.S. EPA, (1998a), "EPA Third-Generation Air Quality Modeling System, Models-3 Volume 9b Users Manual", EPA-600/R-98/069(b), http://www.cmascenter.org/, and http://www.epa.gov/asmdnerl/models3/cmaq.html.

U.S EPA, (1998b)"Finding of Significant Contribution and Rulemaking for Certain States in the Ozone Transport Assessment Group Region for Purposes of Reducing Regional Transport of Ozone; Final Rule," 63 FR 57,356 (October 27, 1998).

U.S. EPA, (1999a), "Draft Guidance on the Use of Models and Other Analyses in Attainment Demonstrations for the 8-Hour Ozone NAAQS", EPA-454/R-99-004, http://www.epa.gov/ttn/scram, Modeling Guidance, DRAFT8HR. (May 1999)

U.S. EPA, (1999b), "Implementation Guidance for the Revised Ozone and Particulate Matter (PM) National Ambient Air Quality Standards (NAAQS) and Regional Haze Program".

U.S. EPA, (1999c), "Guidance on the Reasonably Available Control Measures (RACM) Requirement and Attainment Demonstration Submissions for Ozone Nonattainment Areas.", http://www.epa.gov/ttn/oarpg/t1pgm.html .

U.S. EPA, (2001), "(Draft) Guidance for Demonstrating Attainment of Air Quality Goals for $PM_{2.5}$ and Regional Haze", http://www.epa.gov/ttn/scram/guidance_sip.htm,

U.S. EPA, (2002a), "3 Year Quality Assurance Report- Calendar Years 1999, 2000, and 2001-$PM_{2.5}$ Performance Evaluation Program."

U.S. EPA, (2002b): Clean Air Status and Trends Network (CASTNet), 2001 annual report. http://www.epa.gov/castnet/

U.S. EPA, (2003a), "Federal Reference Method Quality Assurance Report".

U.S. EPA, (2003b), "Guidance for Tracking Progress Under the Regional Haze Rule", September 2003, EPA-454/B-03-004, http://www.epa.gov/ttn/oarpg/t1/memoranda/rh_tpurhr_gd.pdf .

U.S. EPA, (2004a), "The Ozone Report: Measuring Progress Through 2003." EPA 454/k-04-001. http://www.epa.gov/air/airtrends/aqtrnd04/ozone.html

U.S. EPA, (2004b), "The Particle Pollution Report: Current Understanding of Air Quality and Emissions Through 2003." EPA 454/R-04-002. http://www.epa.gov/air/airtrends/aqtrnd04/pm.html

U.S. EPA, (2004c), "Developing Spatially Interpolated Surfaces and Estimating Uncertainty", EPA-454-R-04-0004, http://nsdi.epa.gov/oar/oaqps/pm25/docs/dsisurfaces.pdf.

U.S. EPA, (2004d), "EPA-CMB8.2 Users Manual", December 2004, EPA-452/R-04-011, Office of Air Quality Planning and Standards RTP, NC, .http://www.epa.gov/scram001/models/receptor/EPA-CMB82Manual.pdf

U.S. EPA, (2004e), "Protocol for Applying and Validating the CMB Model for PM2.5 and VOC", December 2004, EPA-451/R-04-001, Office of Air Quality Planning and Standards, RTP, NC, http://www.epa.gov/scram001/models/receptor/CMB_Protocol.pdf

U.S. EPA, (2004f), "User's Guide for the AMS/EPA Regulatory Model - AERMOD", EPA-454/B-03-001, September 2004, Office of Air Quality Planning and Standards, RTP, NC.

U.S. EPA, (2004g), "AERMOD: Description of Model Formulation", EPA-454/R-03-004, September 2004, Office of Air Quality Planning and Standards, RTP, NC.

U.S. EPA, (2004h), Emission Inventory Improvement Program (EIIP), series of technical documents, http://www.epa.gov/ttn/chief/eiip.

U.S. EPA, (2005a), "Regulatory Impact Analysis for the Final Clean Air Interstate Rule", http://www.epa.gov/interstateairquality/pdfs/finaltech08.pdf

U.S. EPA, (2005b), "Technical Support Document for the Final Clean Air Interstate Rule- Air Quality Modeling", March 2005. Docket number OAR-2003-0053-0162. http://www.epa.gov/CAIR/pdfs/finaltech02.pdf

U.S. EPA, (2005c), "Evaluating Ozone Control Programs in the Eastern United States: Focus on the NOx Budget Trading Program, 2004", http://www.epa.gov/air/airtrends/2005/ozonenbp/

U.S. EPA, (2005d), 40CFR, Part 51, Appendix W, Revision to the Guideline on Air Quality Models, 68 FR 68235-68236, November 9, 2005. http://www.epa.gov/scram001/guidance/guide/appw_05.pdf

U.S. EPA, (2005e), "Emission Inventory Guidance For Implementation Of Ozone And Particulate Matter National Ambient Air Quality Standards (NAAQS) and Regional Haze Regulations, http://www.epa.gov/ttn/chief/eidocs/eiguid/eiguidfinal_nov2005.pdf .

U.S. EPA, (2005f), "EPA's Final Draft National Ambient Air Monitoring Strategy", EPA-SAB-CASAC-05-006, May 2005; posted at http://www.epa.gov/sab/pdf/casac-05-006.pdf

U.S. EPA, (2006a), "SPECIATE 4.0 Speciation Database Development Documentation- Final Report", EPA/600/r-06-161, November 2006, Office of Research and Development, RTP, NC, http://www.epa.gov/ttn/chief/software/speciate/speciate4/documentation/speciatedoc_1206.pdf

U.S. EPA, (2006b), "AQS Data Dictionary", Version 2.12, http://www.epa.gov/ttn/airs/airsaqs/manuals/manuals.htm and http://www.epa.gov/ttn/airs/airsaqs/manuals/AQS%20Data%20Dictionary%20v2.12.pdf

U.S. EPA, (2006c), "Technical Support Document for the Proposed PM NAAQS Rule- Response Surface Modeling", February 2006, Office of Air Quality Planning and Standards, RTP, NC, http://www.epa.gov/scram001/reports/pmnaaqs_tsd_rsm_all_021606.pdf

U.S. EPA. (2007) "Guidance for Setting Reasonable Progress Goals Under the Regional Haze Program".

Vukovich, J.M., and T. Pierce, (2002), "The Implementation of BEIS3 within the SMOKE modeling framework", *11th Annual Emissions Inventory Conference of the U.S. EPA*, http://www.epa.gov/ttn/chief/conference/ei11/, Session 10, or http://www.epa.gov/ttn/chief/conference/ei11/modeling/vukovich.pdf

Watson, J.G., (1997), "Observational Models: The Use of Chemical Mass Balance Methods", Prepared for the U.S. EPA Source Attribution Workshop, Research Triangle Park, NC,

July 16-18, 1997, http://www.epa.gov/ttn/faca/stissu.html, Source Attribution Workshop Materials.

Watson, J.G., J.C. Chow, and E.M. Fujita, (2001), "Review of volatile organic compound source apportionment by chemical mass balance", *Atmospheric Environment*, **35**, 1567-1584.

Watson, J.G., T. Zhu, J.C. Chow, J.P. Engelbrecht, E.M. Fujita, and W.E. Wilson, (2002), "Receptor modeling application framework for particle source apportionment", *Chemosphere*, **49**(9), 1093-1136.

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, and J.B. Milford, (1995), "Uncertainties in Incremental Reactivities of Volatile Organic Compounds", *Environmental Science and Technology*, **29**, 1336-1345.

Yang, Y.J., J.G. Wilkinson and A.G. Russell, (1997a), "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, http://www.epa.gov/ttn/faca/stissu.html, SAW-Direct Sensi Analys.--PowerP.--T.Russell.

Yang, Y.J., J.G. Wilkinson, and A.G. Russell, (1997b), "Fast, Direct Sensitivity Analysis of Multidimensional Photochemical Models", *Environmental Science and Technology*, **31**, 2859-2868.

Yarwood, G., and R. Morris, (1997a), "Description of the CAMx Source Attribution Algorithm", prepared for U.S. EPA Source Attribution Workshop, Research Triangle Park, NC, July 16-18, 1997, http://www.epa.gov/ttn/faca/stissu.html.

Yarwood, G., G. Wilson, R.E. Morris, and M.A. Yocke, (1997b), "User's Guide to the Ozone Tool: Ozone Source Apportionment Technology for UAM-IV", prepared for Thomas Chico, South Coast Air Quality Management District, Diamond Bar, CA, March 28, 1997.

Yarwood, G., T.E. Stoeckenius,, J.G. Heiken, and A.M. Dunker, (2003), "Modeling Weekday/Weekend Ozone Differences in the Los Angeles Region for 1997", *J.Air & Waste Management Association*, **53**, 864-875.

Yarwood, G., Morris, R. E., and Wilson, G.M., (2005), "Particulate Matter Source Apportionment Technology (PSAT) in the CAMx Photochemical Grid Model", www.camx.com/publ/pdfs/___Yarwood_ITM_paper.pdf

Yua, J. Z., H. Yanga, H. Zhanga, and A.K.H. Lau, (2004), " Size distributions of water-soluble organic carbon in ambient aerosols and its size-resolved thermal characteristics", *Atmospheric Environment* 38,1061–1071.

Yuan., Z.B., J. Z. Yu, A. K. H. Lau, P. K. K. Louie, and J. C. H. Fung, (2006), " Application of positive matrix factorization in estimating aerosol secondary organic carbon in Hong Kong and its relationship with secondary sulfate", *Atmospheric Chemistry and Physics*, 6, 25-34.

Zhang, X. and P. H. McMurry, (1992), "Evaporative losses of fine particulate nitrates during sampling", *Atmospheric Environment*, 26, 3305-3312.

Zheng, M., G.R. Cass, J.J. Schauer, and E.S. Edgerton, (2002), "Source apportionment of PM2.5 in the southeastern United States using solvent-extractable organic compounds as tracers", *Environ. Sci. Technol.*, **36,** 2361-2371.

Glossary

Modeled 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 an "unmonitored area analysis" and a review of a diverse set of model outputs and emissions, air quality and meteorological data for consideration in a weight of evidence determination to assess whether attainment of the NAAQS is likely with the proposed control strategy.

Modeled attainment test - This test takes the ratio of mean predicted future and current ozone or $PM_{2.5}$ species concentrations averaged over multiple days and multiplies this ratio times the site-specific monitored design value at each monitoring location. If the product is less than the NAAQS near all monitoring sites, the test is passed.

Modeling system - This is a group of models used to predict ambient ozone and/or $PM_{2.5}$ concentrations. The group includes an emissions model which converts countywide emission information into gridded speciated emissions which vary diurnally and reflect environmental conditions. It also includes a meteorological model which provides 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 hourly pollutant concentrations.

Relative response factor (RRF) - The ratio of predicted ozone and/or $PM_{2.5}$ concentrations averaged over multiple days near a monitoring site with future emissions to corresponding predictions obtained with current emissions.

Uniform Rate Of Progress (a.k.a Glidepath)- The amount of visibility improvement needed (in each review period) to reach natural background visibilty conditions by 2064 at each Class I area by achieving a linear rate of improvement in visibility between the 2000-2004 base period and 2064.

Unmonitored Area Analysis - An analysis used to ensure that a proposed control strategy will be effective in reducing ozone and/or $PM_{2.5}$ at locations without air quality monitors so that attainment is shown throughout a nonattainment area. The purpose of the analysis is to use a combination of model output and ambient data to identify areas that might exceed the NAAQS if monitors were located there.

Weight of evidence determination (WOE) - This is a set of diverse analyses used to judge whether attainment of the NAAQS is likely. The credibility of each analysis is assessed and an outcome consistent with an hypothesis that the NAAQS will be met is identified beforehand. If the set of outcomes, on balance, is consistent with attainment, then the WOE can be used to show attainment. A weight of evidence determination includes results from the modeled attainment test, the unmonitored area analysis, other model outputs and several recommended analyses of air quality, emissions and meteorological data.

APPENDIX A

Below are the definitions of model performance statistics suggested as part of this ozone modeling guidance.

Mean Observation: The time-average mean observed value (in ppb)

$$OBS = \frac{1}{N}\sum_{1}^{N} Obs$$

Mean Prediction: The time-average mean predicted value (in ppb) paired in time and space with the observations.

$$Model = \frac{1}{N}\sum_{1}^{N} Model$$

Ratio of the Means: Ratio of the predicted over the observed values. A ratio of greater than 1 indicates on overprediction and a ratio of less than 1 indicates an underprediction.

$$Ratio = \frac{1}{N}\sum_{1}^{N} \frac{Model}{Obs}$$

Mean Bias (ppb): This performance statistic averages the difference (model - observed) over all pairs in which the observed values were greater than zero. A mean bias of zero indicates that the model over predictions and model under predictions exactly cancel each other out. Note that the model bias is defined such that positive values indicate that the model prediction exceeds the observation, whereas negative values indicate an underestimate of observations by the model. This model performance estimate is used to make statements about the absolute or unnormalized bias in the model simulation.

$$BIAS = \frac{1}{N}\sum_{1}^{N} (Model - Obs)$$

Normalized Mean Bias (percent): This statistic averages the difference (model - observed) over the sum of observed values. Normalized mean bias is a useful model performance indicator because it avoids over inflating the observed range of values.

$$NMB = \frac{\sum_{1}^{N} (Model - Obs)}{\sum_{1}^{N} (Obs)} \cdot 100\%$$

Mean Fractional Bias (percent): Normalized bias can become very large when a minimum threshold is not used. Fractional bias is used as a substitute. The fractional bias for cases with factors of 2 under- and over-prediction are -67 and + 67 percent, respectively (as opposed to -50 and +100 percent, when using normalized bias). Fractional bias is a useful indicator because it has the advantage of equally weighting positive and negative bias estimates. The single largest disadvantage is that the predicted concentration is found in both the numerator and denominator.

$$FBIAS = \frac{2}{N} \sum_{1}^{N} \left(\frac{(Model - Obs)}{(Model + Obs)} \right) \cdot 100\%$$

Mean Error (ppb): This performance statistic averages the absolute value of the difference (model - observed) over all pairs in which the observed values are greater than zero. It is similar to mean bias except that the absolute value of the difference is used so that the error is always positive.

$$ERR = \frac{1}{N} \sum_{1}^{N} |Model - Obs|$$

Normalized Mean Error (percent): This performance statistic is used to normalize the mean error relative to the observations. This statistic averages the difference (model - observed) over the sum of observed values. Normalized mean error is a useful model performance indicator because it avoids over inflating the observed range of values.

$$NME = \frac{\sum_{1}^{N} |Model - Obs|}{\sum_{1}^{N} (Obs)} \cdot 100\%$$

Mean Fractional Error (percent): Normalized error can become very large when a minimum threshold is not used. Therefore fractional error is used as a substitute. It is similar to the fractional bias except the absolute value of the difference is used so that the error is always positive.

$$FERROR = \frac{2}{N}\sum_{1}^{N}\left(\frac{|Model - Obs|}{(Model + Obs)}\right)\cdot 100\%$$

Correlation Coefficient (R^2): This performance statistic measures the degree to which two variables are linearly related. A correlation coefficient of 1 indicates a perfect linear relationship; whereas a correlation coefficient of 0 means that there is no linear relationship between the variables.

$$CORRCOEFF = \frac{\sum_{1}^{N}\left(Model - \overline{Model}\right)\left(Obs - \overline{Obs}\right)}{\sqrt{\sum_{1}^{N}\left(Model - \overline{Model}\right)^2\sum_{1}^{N}\left(Obs - \overline{Obs}\right)^2}}$$

Appendix B: Summary of recent model performance evaluations conducted by other modeling groups

1. Boylan, J., VISTAS, "PM Model Performance Goal and Criteria", National RPO Modeling Meeting, Denver, CO, 2005a.

- Based on benchmarking with a combination of data from a number of PM modeling studies (SAMI, VISTAS, WRAP, EPA, MANE-NU, EPRI, and Midwest RPO). Proposed performance goals (close to best achievable) and performance criteria (acceptable) for these metrics and showed where the modeling studies fall for the various components of PM.
- Proposed to use mean fractional bias (MFB) and mean fractional error (MFE) as the standard performance metrics.

  Goal: MFE <= 50%, MFB <= +-30%
  Criteria: MFE <= 75%, MFB <= +-60%
  Less abundant species should have
  less stringent goal & criteria
- Proposed to use asymptotically approaching goals & criteria when data are greater than 2.5 um, approaching +200% MFE and +-200% MFB for extreme small model & observed data (formula of logarithmic MFB & MFE are proposed)
- Based on combined modeling studies described above, for more abundant conditions, MFE and MFB are typical in the range of

  Sulfate: MFE = 30% ~ 77%, MFB = -45% ~ +51% (> 2 ug/m3)
  Nitrate: MFE = 55% ~ 125%, MFB = +3% ~ +82% (> 1 ug/m3)
  Organic: MFE = 35% ~ 95%, MFB = -70% ~ +35% (> 1.5 ug/m3)
  EC: MFE = 50% ~ 95%, MFB = -45% ~ +50% (> 0.5 ug/m3)
  $PM_{2.5}$: MFE = 50% ~ 85%, MFB = -55% ~ +60% (> 5 ug/m3)
- Suggested to conduct performance evaluation on episode-by-episode basis or month-by-month for annual modeling
- Different performance goals & criteria may be needed for gaseous precursors and wet depositions.
- Benchmarking should be done for the entire modeling system (meteorology, emissions inventory, and model).

2. Morris, R., et al., "Application of Multiple Models to Simulation Fine Particulate in the Southeastern US", National RPO Modeling Meeting, Denver, CO, 2005a.

- Based on model multiple model applications over VISTAS modeling:
  1July 1999 & July 2001: CMAQ and CMAQ 36-km & 12-km
  2January 2002 & July 2001: CMAQ and CMAQ MADRID 36-km only

3Used same horizontal & vertical grids, CMAQ-to-CAMx emissions, ICs/BCs, but different MM5 interface (MCIP vs. MM5CAMx).

- Both models performed reasonably well, with CMAQ performing better for SO4 and CAMx performing better for OM. Both models did not perform well for NO3, soil, and coarse PM, but CAMx seemed to have much higher positive bias in winter nitrate than CMAQ
- Used fractional bias and fractional error (instead of normalized ones) to illustrate model performance because these statistics do not exhibit such extreme fluctuation.
- Demonstrated the usefulness of "soccer goal" plots for MFB & MFE. Suggested 15%/35% (MFB/MFE) for O3, and illustrated 50%/75% for $PM_{2.5}$ species and bounded at 100%/200%
- Based on modeling studies described above (July 1999 & 2001), MFE and MFB are typical in the range of

  Sulfate: MFE = 25% ~ 70%, MFB = -25% ~ +51%
  Organic: MFE = 30% ~ 75%, MFB = -50% ~ +45%
  EC: MFE = 35% ~ 65%, MFB = -30% ~ +40%
- CMAQ performance improved for OM when they adjusted the Kz min value to 1.0 m^2/sec (from 0.1m^2/sec).
- Data (July 2001) of SO4 and total carbon mass (TCM). Both models seemed to captured the diurnal and daily trends well for both SO4 & TCM, but the magnitude can be significantly off for some of the days.

3. Boylan, J. and Baker, K., "Photochemical Model Performance and Consistency", National RPO Modeling Meeting, Denver, CO, 2005.

- Illustrated model performance comparisons of 36-km daily modeling results by MRPO (CAMx4), VISTAS (CMAQ-CB4), and MANE-VU (CMAQ-SARPC99) for two episodes (July 1999 and January 2002) at three IMPROVE sites; and by MRPO (CAMx4) and VISTAS (CMAQ-CB4 for one episode (July 1999) for the Pittsburgh supersite.
- Indicated that in some cases there was good agreement among the models and selected IMPROVE sites, but in others cases there were noticeable variations (even between the two RPOs running CMAQ).
- Sulfate were more consistent among the three models; nitrate and OM have higher discrepancies; All three models overpredicted nitrate and underpredicted OM (based on model vs. observed scatter plots comparisons)
- Suggested that where models diverge, they may show a different response to control strategies. Reasons for the model variations may include differences between CB4 and SAPRC99, and potential differences in emissions inventories, differences in Kz_min values, differences in met. & land use/soil methodologies, etc.
- Also described a comparison of hourly modeling results by MRPO and VISTAS for a July 1999 episode at the Pittsburgh super site. In some cases these hourly results were

better than daily results and generally captured the diurnal patterns for $PM_{2.5}$ species and gaseous species (O3, NOx, HNO3, SO2, etc.), although the magnitude was considerably off in some cases.
4. Seigneur, C., "Review of CMAQ and REMSAD Performance for Regional PM Modeling", AER Report for UARG, Doc# CP163-03-01b, March 2003. • Describe a review of summary performance of CMAQ, CMAQ-MADRID 1, CMAQ-MADRID 2, and REMSAD. Based on the comparison studies by others given below: a. BRAVO 1999 study: CMAQ-MADRID 1 & REMSAD b. WRAP 1996 studies: CMAQ & REMSAD c. WRAP Aug. 1999 & Jan. 2000 studies: CMAQ & REMSAD d. EPA/ORD July 1999 studies: CMAQ & REMSAD e. Southeast PM - July 1999 study (Nashville/Atlanta):CMAQ, CMAQ-MADRID1,2 & REMSAD • AER Suggested normalized errors of 50% as the benchmark for sulfate and $PM_{2.5}$ (note: this suggestion was not agreed by other modeling groups, see references #2, #5, #7: the MNE is most biased metrics and has high fluctuations). The review showed only SE PM study meets this criteria for sulfate and $PM_{2.5}$, and all the rest of species failed. No model showed consistently better performance than the others. Suggested that the model inputs has more effect on performance than model formulations. • Indicated that "the current performance of air quality models for PM is poor"..."There is a dire need for improving model inputs and model formulations in order to obtain acceptable model performance"... "3-D air quality models are the best tools available to address the PM source-receptors relationships because they take into account the non-linearities that affect the formation of secondary PM".
5. Boylan, J., VISTAS, "Calculating Statistics: Concentration Related Performance Goals", PM Model Performance Workshop, RTP, NC, 2004b. • Illustrated a set of standard bias and errors calculations commonly used for model performance statistics and proposed model performance goal. • Indicated Mean Normalized Bias and Errors (MNBE) are most biased and least useful among "MNBE", "NMBE" and "MFBE". The Mean Fractional Bias and Errors (MFBE) is least biased and most useful among the three metrics. • Recommended MFB & MBE and proposed performance goal: MFE<=50% and MFB <=+-30% for more abundant species (eg., sulfate & $PM_{2.5}$) and less stringent for less abundant species (eg., nitrate, OC, EC, soil, etc.). Performance goal is not criteria and should be prohibit the modeling from being used if it fails to meet the goal. • Proposed to use "2.5ug/m3" as the "grayline" for 50% MFE & +-30% MFB and asymmetrically approaching 200% MFE & +-200% MFB to extremely small concentrations.

- Recommended to use monthly avg. for annual modeling.

6. Morris, R., et al., "Model and Chemistry Inter-comparison: CMAQ with CB4, CB4-2002, SAPRC99", National RPO Modeling Meeting, Denver, CO, 2005b.

- Based on US (36-km) and VISTAS (12-km) January 2002 modeling, conducted chemistry mechanisms inter-comparisons for CMAQ with CB4, CB4-2002,and SAPRC99.
- The performance of CB4 and CB4-2002 was similar for PM, and superior to SAPRC99 overall (for the Jan02 case).
- The model performance for CMAQ/CB4, US 36-km domain is in the range of:

  Sulfate: MFE = 42% ~ 73%, MFB = -21% ~ +14%
  Nitrate: MFE = 62% ~ 105%, MFB = -21% ~ +46%
  Organic: MFE = 50% ~ 77%, MFB = +3% ~ +59%
  EC: MFE = 59% ~ 88%, MFB = +2% ~ +70%
  Soil: MFE = 165% ~ 180%, MFB =+164% ~ +180%
  $PM_{2.5}$: MFE = 48% ~ 88%, MFB = +25% ~ +81%
- Given that the computational cost of SAPRC99 is twice that of CB4, suggested to use 36 and 12 km grids with CB4 chemistry for PM modeling for the time being.
- Noted that both CB4 and SAPRC underpredicted winter O3 significantly.

7. Tonnesen, G., et al., "Regional Haze Modeling: Recent Modeling Results for VISTAS and WRAP", CMAS Annual workshop, RTP, NC, 2003.

- Illustrated the WRAP 1996 CMAQ 36-km modeling and performance evaluation in the Western U.S. and VISTAS CMAQ 12-km modeling for 3 episodes: January 2002, July 1999, July 2001. Recommended to use the performance metrics of Mean Fractional Bias (MFB)and Mean Fractional errors (MFE) over mean normalized bias & errors (used in earlier WRAP model evaluation).

  Sulfate: MFB = -47% ~ +48% (1996 WRAP domain)
  Nitrate: MFB = -95% ~ +30% (1996 WRAP domain)
  Organic: MFB = -20% ~ +70% (1996 WRAP domain)
  OC: MFB = -45% ~ +3% (1996 WRAP domain)
- VISTAS modeling key findings (1) sulfate performance reasonably well (2) nitrate overpredictions in the winter, underpredictions in summer, may need better NH3 emissions (3) Kv min =1 improved performance, mixing height is important (4) minor differences in 19 vs. 34 layers

8. Zhang, Y., et al., "Performance Evaluation of CMAQ and PM-CAMx for July 1999 SOS Episode", AER Report for CRC, Doc# CP131-03-01, April 2003.

- Conducted CMAQ (2002 version) and PM-CAMx performance evaluation based on July 1999 SOS episode (6/29-7/11) modeling study (32-km nested w/ 8-km in the SE U.S., including Atlanta & Nashville)
- Ozone performance: use MNB & MNE w/ 60 ppb threshold for O3 (CMAQ performed better):

  CMAQ: MNB < 1%, MNE = 18%

  CAMx: MNB = 27%, MNE = 33%
- PM performance: CMAQ & PM-CAMx are generally consistent in the rural areas (vs. IMPROVE); differs significantly over urban/suburban; in general, CMAQ performs much better than PM-CAMx

  $PM_{2.5}$: CMAQ: MNB = 38%, MNE = -7%

  CAMx: MNB = 55%, MNE = 35%

  Sulfate: CMAQ: MNB = 9%, MNE = 45%

  CAMx: MNB = 44%, MNE = 63%

  Nitrate: CMAQ: MNB = -50%, MNE = 98%

  CAMx: MNB =137%, MNE = 158%
- The performance of CMAQ for PM and O3 is consistent with the performance expected for air quality models (however, nitrate issue existed); the performance of CAMx does not generally meet current expectations for AQM.

9. Morris. R., et al.,"Evaluation of CAMx: Issues Related to Section Models ", PM Model Performance Workshop,RTP, NC, 2004c.

- Illustrated a WRAP comparison of CMAQ (v4.3), REMSAD (v7), CAMx (bimodal PM), and CAMx (4-section PM) based on January and July 1996 and annual 1996 over the 36-km WRAP domain.
- Indicated that all models exhibited variations in performance, but no clear winner across all species and periods. Sulfate predictions were reasonable, but nitrate was significantly overpredicted. For all three models,

  Sulfate : MFE = 40% ~ 60%, MFB = -40% ~ +14% (July 96)

  Nitrate : MFE = 105% ~ 200%, MFB = +45% ~> +100% (Jan. 96)

  Organic: MFE = 50% ~ 75%, MFB = -65% ~ +5% (July 96)

  EC : MFE = 47% ~ 105%, MFB = -48% ~ +25% (Jan. 96)
- The 1996 model performance is less than stellar, indicating potential issues in 1996 MM5 and emissions.
- Showed the effects of sectional PM distribution on PM Modeling in the Western US. Based on a study in the South Coast Air Basin using CAMx4+, which allows side-by-side comparisons of aqueous-phase chemistry modules (bulk vs. variable size resolution) and of PM size distribution treatments (bimodal vs. sectional).

10. Tonnesen, G., et al..,"Model Performance Metrics, Ambient Data Sets and Evaluation Tools", PM Model Performance Workshop, RTP, NC, 2004a.

- Illustrated model performance metrics, available $PM_{2.5}$ and O3 data, and evaluation software tool. Suggested that air quality modeling should include model evaluation as part of the system.
- Comparing performance metrics may not be enough since performance metrics often show mixed response and it is possible for a better model to have poorer metrics (e.g., bias in met & emissions inputs). Diagnostic evaluation is needed to judge finer grid performance since coarser grid may have compensating errors. But should we assume that finer grid modeling always gives better simulations/physics?
- An example of VISTAS modeling (July 1999) showed differences of hourly sulfate and its wet deposition between CMAQ results using 12-km and 36-km grids, possible due to regional transport (wind speed & direction), precipitations/clouds, and numerical diffusion.
- Recommended bias factor as the best metric for evaluating haze.

11. Wang, Z., et al.," Comparison and Diagnostic Evaluation of Air Quality Models for Particulate Matter: CAMx, CMAQ, CMAQ-MADRID", National RPO Modeling Meeting, 2004.

- Conducted model evaluation based on 1999 SOS episode (6/29-7/10) EPRI modeling study,
- CAMx has higher positive bias than CMAQ and CMAQ-Madrid in predicting sulfate; CMAQ underestimated nitrate and CAMx and CMAQ-MADRID overestimated nitrate.
- All three models underestimated OM. However, there was no clear winner in model performance.
- The three models responded differently to a 50% increase in ammonia emissions, indicating a need to further look at the models' responsiveness to changes in emissions.

12. Ku, C., CENRAP, "CMAQ and CAMx Simulations for January and July 2002", National RPO Modeling Meeting, Denver, CO, 2005.

- Compared CMAQ and CAMx 36-km simulations Based on January and July 2002 (basB) over a continental US domain.
- Indicated that the results were mixed for the models over CENRAP generally. Both models performed acceptably for $PM_{2.5}$ and sulfate in the summer (when sulfate is abundant) but they were overpredicted in the winter compared against IMPROVE network.
- CAMx significantly overpredicted nitrate in the winter (higher prediction than CMAQ), but has better performance in OM (lower prediction than CMAQ), based on performance measures of normalized bias and errors.
- The study also showed mixed results for the models in three climatically different regions in CENRAP that contain Class I areas, with varying performance depending on region and season. This finding points up the difficulty in improving model performance over the whole CENRAP domain.

13. Eder B. and Yu S., "An Evaluation of the 2003 Release of Models-3/CMAQ", CMAS Annual workshop, RTP, NC, 2003. • Illustrated CMAQ 2003 evaluation for two episodes (winter 2002 and Summer 1999) for O3 and $PM_{2.5}$ species against AIRS, CASTNet, IMPROVE, and STN. Suggested the use of the performance metrics of Normalized Mean Bias (NMB)and Normalized Mean errors (NMB) (and correlation coefficient, R). • Ozone is fairly unbiased and accurate (NMB<10%) and NME=~20%); sulfate performance is quite well, even for STN. NO3 was the worst in the winter (NME ~67% for STN and ~96% for IMPROVE)
14. Frank, N.,"Use of National $PM_{2.5}$ and Speciation Network Measurements for Model Evaluation ", PM Model Performance Workshop, RTP, NC, 2004. • Compared correlated speciated monitoring sites (e.g., IMPROVE, CASTNet, STN) for ambient PM species measurements. • The sulfate agrees very well, but nitrate has both positive & negative bias site-by-site. • OM is more uncertain than other species, but is still somewhat robust for model evaluation (ie., uncertainty is relatively small compared to current range of uncertainty in modeling). OM is about 50%-80% of urban excess of $PM_{2.5}$.
15. Hu, Y., et.al., "Evaluation of CMAQ with FAQS Episode of August 11th-20th", 2000, CMAS Annual workshop, RTP, NC, 2003. • Conducted CMAQ (36/12/4 km nesting) performance evaluation for O3 based on Fall Line Air Quality Study (FAQS) 8/11-20. • Indicated that CMAQ had a good O3 performance, but has a nighttime problem, which could be due to min Kz used. Analysis suggested that an optimal of min Kz may lie between 0.1~1 m^2/s. • The Isoprene emissions may be overestimated in the rural area and CO emissions may be underestimated.
16. Tonnesen, G., et al., "Prelim Preliminary Results CMAQ and CMAQ-AIM with SAPRC99", National RPO Modeling Meeting, 2004b. • Compared CMAQ and CMAQ-AIM with SAPRC99 based on 2001 VISTAS modeling study. CMAQ-AIM with SAPRC99 has larger negative bias and lower predictions of Sulfate PM than standard released CMAQ. • Conducted model & chemistry inter-comparisons: CMAQ with CB4, CB4-2002,

SAPRC99. Some chemistry differences were observed in the models based on the 36-km U.S. and the 12-km VISTAS modeling; the performance of CB4 and CB4-2002 was similar, and superior to SAPRC99 overall.

- Given that the computational cost of SAPRC99 is twice that of CB4, suggested to use 36 and 12 km grids with CB4 chemistry for current VISTAS modeling study.

17. Yu, S., et al., "Simulation of Primary and Secondary Organic Aerosols over the US by CMAQ: Evaluation and Analysis", CMAS Annual workshop, RTP, NC, 2003.

- Evaluated CMAQ w/ SAPRC99 performance of Primary Organic Aerosols (POA), EC, and secondary Organic Aerosols (SOA) based on IMPROVE, SEARCH, and SOS data.

- Suggested that model captured general trends & patterns of most EC & POA within a factor of two (based on IMPROVE & SEARCH).
- Slight underprediction in the Eastern U.S., likely due to underprediction in biogenic SOA.

18. Chien, C. Y., et al., "CMAQ Model Performance Evaluation with the Updated CB4-2002", CMAS Annual workshop, RTP, NC, 2003.

- Evaluated CMAQ w/ CB4 vs. CB4-2002 (and CB4-2002 with removed N2O5 gaseous reaction) against IMPROVE, CASTNet and AQS based on January & July 1996 cases. Described the key updates of CB4-2002 (HNO3, N2O5 & PAN rxns).
- CB4-2002 has lower O3, lower N2O5, lower PAN, higher HNO3, higher nitrate, slightly higher sulfate, and slightly lower SOA prediction than CB4.
- The performance of CMAQ CB4 and CB4-2002 was similar for PM species. However, CB4-2002 has higher positive bias for winter nitrate (since it predicted higher nitrate).

19. Dennis, R.,"Time-Resolved and In-Depth Evaluation of PM and PM Precursors Using CMAQ ", PM Model Performance Workshop, RTP, NC, 2004.

- Evaluated the CMAQ (2003 version) performance of inorganic PM, EC, and gaseous precursors against Atlanta 1999 supersite summer data and 2002 Pittsburgh supersite winter data.
- Suggested that collapse of evening PBL is too fast and rise of morning PBL too slow (based on higher modeled EC, NOy & CO against Atlanta data at sunrise & sunset)
- Performance for sulfate & ammonium was fairly good;
- Overprediction for nitrate, but updated chemistry in CMAQ (included in 2004 version) has improved the nitrate performance (updated reaction probability for HNO3 heterogeneous rxn and remove gaseous N2O5->HNO3 rxn)
- Use ratio to define nitrate PM formation being HNO3- or NH3-limited: (NHx-

2*SO4)/(HNO3+NO3)

- Nitrate PM predictions are very sensitive to NHx, and thus the NH3 emissions need serious attention.

20. Kumar, N.,"$PM_{2.5}$ Model Performance: Lessons Learned and Recommendations ", PM Model Performance Workshop, RTP, NC, 2004.

- Illustrated the use of model performance statistics based on two CMAQ-MADRID applications, SOS 1999 episode in the SE U.S. and 12-km BRAVO summer 1999 episode (nested within REMSAD BRAVO 36-km modeling results)
- Indicated model evaluation issues regarding local vs. regional, daily/weekly vs. month/seasonal

Sulfate : MNB= +20% / +51%, MNE = 51% / 89% (SEARCH/IMPROVE)
Nitrate : MNB= +72% / -25%, MNE = 72% / 46% (SEARCH/IMPROVE)
EC : MNB= +14% / -8%, MNE = 52% / 54% (SEARCH/IMPROVE)
$PM_{2.5}$: MNB= -19% / -8%, MNE = 32% / 49% (SEARCH/IMPROVE)

- Described the available performance metrics and illustrated the use of logarithmic and fractional bias and errors as potential benchmarks

21. Baker K., Midwest RPO, "Fine Grid Model Performance", National RPO Modeling Meeting, Denver, CO, 2005.

- Based on the Midwest RPO 36-km and 12-km Modeling study using CAMx4
- Indicated that the model performance for $PM_{2.5}$ species generally was similar for 36 and 12 km grids; there were some differences in running control scenarios for sulfate/nitrate, but we could not determine which one is better

22. Morris, R., et al., "VISTAS Grid Resolution Sensitivity", National RPO Modeling Meeting, Denver, CO, 2005d.

- Based on VISTAS Southeast U.S. modeling study for Phase I episode (Jan. 2002, July 1999, and July 2001), compared results for CMAQ w/ CB4 and SAPRC99 using the 36 km national RPO domain and the 12 km grid VISTAS domain
- Effects of grid resolution on model performance was mixed (performance was not necessarily improving using 12-km)
- CMAQ w/ SAPRC99 (mixed for sulfate, worse for nitrate) was not performing better than w/ CB4
- Compared July 1999 episode for ozone using 36-km & 12-km CMAQ and CAMx. CMAQ O3 performance degraded to underprediction with finer grid; CAMx is similar at 36-km and 12-km.
- Examined the performance of the MM5 model configurations using various cloud schemes (including Kain-Fritsch and Reisner schemes) for the 12 km (WRAP) and 36 km domains. The results showed that cool and moist bias found in the West,

overprediction of convective precipitation using KF.

23. Zhang, Y., et al., "Development and Application of MADRID: A New Aerosol Module in CMAQ", CMAS Annual workshop, RTP, NC, 2003.

- Evaluated CMAQ-MADRID against SCAQS 1987 episode. The performance of CMAQ-MADRID seemed to perform well in O3, $PM_{2.5}$, and sulfates, and comparable in other $PM_{2.5}$ species to other models (UAM-IV/CALGRID/UAM-AERO, GATOR, SMOG, CIT, SAQM-AERO)
- Described the development and sciences of MADRID aerosol module included in CMAQ. Key features: sectional (size bins), hybrid equilibrium, more sophisticated SOA treatment, CMU aqueous chemistry

24. Morris, R., et al., "WRAP Multi-Model Evaluation Using the 1996 36 km Section 309 Database", National RPO Modeling Meeting, Denver, CO, 2005e.

- Conducted WRAP 1996 multi-model evaluation. Compared CMAQ (v4.3), REMSAD (v7), CAMx (bimodal PM), and CAMx (4-section PM) using the 1996 36-km section 309 database.
- Indicated that sulfate was the best performing species on an annual basis, with a winter overprediction compensating for a summer underprediction; NO3 was predicted poorly by all models; OC, EC, and CM were underpredicted by all models. concluded that modeling is more challenging in the West than in the Midwest and Southeast.

25. Seigneur, C., "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, 2000b.

- Based on a limited set of comparisons with limited temporal and spatial model studies.
- Statistics calculated for aggregated gross error (AGE) and aggregated normalized bias (ANB). Note that the statistics for AGE and AGB were not obtained with the same models nor the same data base. Thus, performance for the two metrics in the table are not associated with each other.

$PM_{2.5}$: AGE ~ +30% - +50%, AGB ~ +10% (based on one study)
Sulfate : AGE ~ +30% - +50%, AGB ~ -20% - -30%
Nitrate : AGE ~ +20% - +70%, AGB ~ -15 - +50%
EC : AGE ~ +15% - +60%, AGB = none available
OC : AGE ~ -40% - +50%, AGB ~ 38% (based on one study)

United States
Environmental Protection
Agency

Office of Air Quality Planning and Standards
Air Quality Analysis Division
Research Triangle Park, NC

Publication No. EPA-454/B-07-002
April 2007