



**Guidance on the Use of Models and
Other Analyses in Attainment
Demonstrations for the 8-hour Ozone
NAAQS**

EPA-454/R-05-002
October 2005

Guidance on the Use of Models and Other Analyses in Attainment Demonstrations for the 8-hour Ozone NAAQS

U.S. Environmental Protection Agency
Office of Air Quality Planning and Standards
Emissions, Monitoring, and 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, the STAPPA/ALAPCO emissions/modeling committee 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 (NYDEC) and Computer Sciences Corporation (CSC) for testing our ideas for a modeled attainment test and sharing the results with us.

TABLE OF CONTENTS

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 Affect My Attainment Demonstration?	2
1.4 What Topics Are Covered In This Guidance?	5
Part I. How Do I Use Results Of Models And Other Analyses To Help Demonstrate Attainment?	7
2.0 What Is A Modeled Attainment Demonstration?--An Overview	8
2.1 What Is The Recommended Modeled Attainment Test?--An Overview	8
2.2 What Does A Recommended Supplemental Analysis/Weight Of Evidence Determination Consist Of? --An Overview	9
2.3 Why Should A Model Be Used In A “Relative” Sense And Why May Corroboratory Analyses Be Used In A Weight Of Evidence Determination?10	
3.0 What Is The Recommended Modeled Attainment Test?	12
3.1 Calculating site-specific baseline concentrations.	13
3.2 Identifying surface grid cells near a monitoring site.	15
3.3 Choosing model predictions to calculate a relative reduction factor (RRF)_I near a monitor.	16
3.4 Estimating design values at unmonitored locations: what is an unmonitored area analysis and why is it needed?	18
3.4.1 Why does the unmonitored area analysis need to use both ambient data and model output?	18
3.4.2 Implementation of Model Adjusted Spatial Fields	19
3.4.3 Using the Results of the Unmonitored Area Analysis	20
3.5 Limiting modeled 8-hour daily maxima chosen to calculate RRF.	21
3.6 Which base year emissions inventory should be projected to the future for the purpose of calculating RRFs?	24
3.7 Choosing a year to project future emissions.	25
3.8 How Do I Apply The Recommended Modeled Attainment Test?	26
4.0 How Can Additional Analyses Can Be Used to Support the Attainment Demonstration?	28
4.1 What Types of Additional Analyses Should Be Completed as Part of the	

	Attainment Demonstration?	28
4.2	If I Use A Weight Of Evidence Determination, What Does This Entail? ...	31
5.0	What Additional Analyses Can Be Completed to Assess Progress towards Attainment	34
6.0	What Documentation Do I Need To Support My Attainment Demonstration?	36
Part II.	How Should I Apply Air Quality Models To Produce Results Needed To Help Demonstrate Attainment?	41
7.0	How Do I Apply Air Quality Models?-- An Overview	42
8.0	How Do I Get Started?- A “Conceptual Description”	46
8.1	What Is A “Conceptual Description”?	46
8.2	What Types Of Analyses Might Be Useful For Developing And Refining A Conceptual Description?	49
8.2.1.	Is regional transport an important factor affecting the nonattainment area?	49
8.2.2.	What types of meteorological episodes lead to high ozone?	49
8.2.3.	Is ozone limited by availability of VOC, NOx or combinations of the two? Which source categories may be most important?	50
9.0	What Does A Modeling/Analysis Protocol Do, And What Does Developing One Entail?	51
9.1	What Is The Protocol’s Function?	51
9.2	What Subjects Should Be Addressed In The Protocol?	51
10.0	What Should I Consider In Choosing An Air Quality Model?	53
10.1	What Prerequisites Should An Air Quality Model Meet To Qualify For Use In An Attainment Demonstration?	53
10.2	What Factors Affect My Choice of A Model For A Specific Application? .	54
10.3	What Are Some Examples Of Air Quality Models Which May Be Considered?	55
11.0	How are the Meteorological Time Periods (Episodes) Selected?	57
11.1	What Are The Most Important Criteria For Choosing Episodes?	58
11.1.1	Choose a mix of episodes which represents a variety of meteorological conditions which frequently correspond with observed 8-hour daily maxima exceeding 84 ppb.	58
11.1.2	Choose episodes having days with monitored 8-hour daily maxima close to observed average 4th high daily maximum ozone	

	concentrations.	59
	11.1.3 Choose days with intensive data bases.	60
	11.1.4 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.	60
11.2	What Additional, Secondary Criteria May Be Useful For Selecting Episodes?	64
12.0	What Should Be Considered When Selecting The Size And Horizontal/Vertical Resolution Of The Modeling Domain?	66
12.1	How is the Size of the Modeling Domain Chosen?	66
12.2	How are the Initial and Boundary Conditions Specified?	67
12.3	What Horizontal Grid Cell Size Is Necessary?	68
12.4	How Should the Vertical Layers Be Selected?	70
13.0	How are the Meteorological Inputs Prepared for Air Quality Modeling?	72
13.1	What Issues are Involved in the Generation and Acquisition of Meteorological Modeling Data?	72
13.2	How Should the Meteorological Modeling Analysis be Configured?	73
13.3	How Should the Performance of the Meteorological Modeling Be Evaluated?	74
14.0	How Are the Emission Inputs Developed?	77
14.1	Can The National Emission Inventory Be Used As a Starting Point?	77
14.2	What Emission Inventory Data are Needed to Support Air Quality Models?	78
14.3	What Other Data are Needed to Support Emissions Modeling?	80
14.4	How Are Inventory Data Converted Into Air Quality Model Input?	82
14.5	Are there Other Emissions Modeling Issues?	84
14.6	How Are Emissions Estimated for Future Years?	86
15.0	What are the Procedures for Evaluating Model Performance and What is the Role of Diagnostic Analyses?	95
15.1	What are the Procedures for Evaluating An Air Quality Model?	95
15.2	How Should the Operational Evaluation of Performance Be Completed?	96
15.3	What Types of Analyses Can be Done to Evaluate the Accuracy of the Model Response: Diagnostic Evaluations?	99
15.4	How Should the Results of the Model Evaluation be Assessed?	100
	REFERENCES	102
	Glossary	115

LIST OF TABLES

Table 2.1 Guidelines For Weight of Evidence Determinations 9

Table 3.1 Example Illustrating Calculation Of Baseline Design Values 15

Table 3.2. Default Recommendations For Nearby Grid Cells Used To Calculate RRF's 16

Table 3.3 Mean RRFs and standard deviations as a function of various minimum thresholds. 23

Table 3.4 Example Calculation of a Site-Specific Future Design Value (DVF)_I 27

Table 6.1 Recommended Documentation for Demonstrating Attainment of the 8-hour NAAQS for Ozone 37

Table 10.1 Current Air Quality Models Used To Model Ozone 56

Table 10.2 Other Air Quality Models Used to Model Ozone 56

Table 11.1 Number of Days Needed to Replicate the 25/50/100 Day Dataset Mean RRF to Within ±1% and ± 2% , With a 95% Confidence Interval 62

Table 11.2 Examples of the Recommended Hierarchy in Choosing the Number of Days in the Mean RRF Calculation vs. the Minimum Threshold 63

LIST OF FIGURES

Figure 3.1 Daily Relative Reduction Factors as a Function of Daily Maximum Base Modeled Concentrations for Monitors in the Baltimore Nonattainment Area 22

Figure 11.1 Mean RRF as a Function of the Subset of Days (3-25) for a Harford County, MD Ozone Monitoring Site 61

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 attainment demonstrations 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 recommends procedures for estimating whether a control strategy to reduce emissions of ozone precursors will lead to attainment of the 8-hour national ambient air quality standard (NAAQS) for ozone. The document also describes how to apply air quality models to generate the predictions later used to see whether attainment is shown. Guidance in this document applies to nonattainment areas¹ for which modeling is needed, or desired.

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?” It explains what we mean by a *modeled attainment demonstration*, a *modeled attainment test*, and a *weight of evidence determination*. It also identifies 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. 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 episodes, or 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. Evaluate the performance of the air quality model and 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.

¹While 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.

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 national ambient air quality standard (NAAQS) for 8-hour daily maximum ozone concentrations will occur by the appropriate attainment 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.

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 8-hour ozone NAAQS. Areas required to submit an attainment demonstration are encouraged to follow the procedures described in this document. Details on when a State is required to submit a modeled attainment demonstration can be found in the 8-hour implementation rule and preamble².

Implementation plan revisions are due three years after an area is designated “nonattainment” (e.g., June 15, 2007 for areas whose effective designation dates are June 15, 2004). Attainment demonstrations supporting these revisions should be completed in time to allow sufficient time to complete the rulemaking process by June 15, 2007, at the latest.

1.3 How Does The Perceived Nature Of Ozone Affect My Attainment Demonstration?

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

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.

²Ozone Implementation Rule- Phase 2

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)³. 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 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 concentrations approach levels specified in the NAAQS throughout much of the eastern U.S. and in large parts of California. A number of analyses (EPA, 1998 and EPA, 2004) show that regional ozone transport can impact areas several hundred miles or more downwind. The regional extent of moderate to high ozone and transport patterns and distances in some areas will likely necessitate nested regional model applications.

This guidance identifies several *modeling systems*⁴ 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. 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

³ Ambient design values are based on observations made at monitor locations.

⁴A 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.

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 concentrations approaching 0.08 ppm in large portions of the U.S. While we believe that existing and future regional reductions in NO_x emissions will reduce ozone over much of the eastern U.S., regional ozone concentrations approaching the level specified in the NAAQS will continue to affect local strategies needed to attain the NAAQS in the remaining nonattainment areas.

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)⁵, LADCO, and the Ozone Transport Commission (OTC)).

Premise 4. High concentrations of ozone, PM_{2.5}⁶ 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. 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.

Integration of strategies to reduce ozone, PM_{2.5} and regional haze may be complicated by the different dates by which SIP revisions may be due (e.g., 2007 for ozone, circa 2007-2008 for PM_{2.5} and regional haze, etc.). This guidance identifies activities which could yield useful information for a subsequent review of the impact of ozone control strategies on PM_{2.5} and regional haze.

⁵EPA provides funding to five regional planning organizations to address regional haze and related issues. <http://www.epa.gov/air/visibility/regional.html>

⁶PM_{2.5} are particles having aerodynamic diameters less than or equal to 2.5 micrometers.

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 5 sections (i.e., Sections 2-6). Part II consists of 9 sections (Sections 7-15).

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.

Section 3 describes the recommended modeled attainment test in detail. The section also includes examples illustrating the use of the recommended tests.

Section 4 describes how supporting analyses should be performed to complement the attainment test, as well as how it should be used in a weight of evidence determination.

Section 5 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 (e.g. mid-course reviews).

Section 6 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 7 with an overview of the topics to be covered.

Section 8 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 9 describes the purpose, function, and contents of a modeling protocol.

Section 10 addresses what criteria should be considered in choosing a model to support the attainment demonstration of the ozone NAAQS. Several guidelines are identified for accepting the use of a model for this purpose.

Section 11 provides guidance for selecting suitable episodes to model for an ozone attainment demonstration. Topics include a discussion of the form of the NAAQS and its resulting implications for episode selection.

Section 12 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 13 addresses how to develop and evaluate meteorological inputs for use in a modeling exercise supporting an attainment demonstration.

Section 14 discusses how to develop appropriate emissions estimates for use in the selected air quality model.

Section 15 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.

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 ozone 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 by using the *modeled attainment test*. 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.

2.1 What Is The Recommended Modeled Attainment Test?--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 concentrations are ≤ 84 ppb, then this element of the attainment test is satisfied⁷. 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 ozone monitors. We call these ratios, *relative reduction factors*. Future ozone concentrations are estimated at existing monitoring sites by multiplying a modeled relative reduction factor at locations “near” each monitor by the observation-based, monitor-specific, “baseline” ozone design value. The resulting predicted “future concentrations” are compared to 84 ppb.

The recommended modeled attainment test predicts whether or not all estimated future design values will be less than or equal to the concentration level specified in the ozone NAAQS under meteorological conditions similar to those which have been simulated. The monitor-based test does not consider future ozone in areas that are not near a monitor. Therefore, we recommend a supplemental *unmonitored area analysis* to identify other locations where passing the attainment test may be problematic if monitoring data were available. Details of the unmonitored area analysis are in Section 3.4.

⁷As detailed in Section 4, additional corroboratory analyses are still needed to supplement the modeled attainment test, even when predicted ozone concentrations are ≤ 84 ppb.

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

As we describe in more detail in Section 4, 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 greater than or equal to 88 ppb 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 attained⁸. 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
Future Design Value < 82 ppb, 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	A weight of evidence demonstration should be conducted to determine if aggregate supplemental analyses support the modeled attainment test
Future Design Value \geq 88 ppb, 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.

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,

⁸ Regional ozone modeling completed by EPA indicates that, on average, considerable amounts of precursor control (e.g., 20-25 percent) will be needed to lower projected ozone design values by 3 ppb or more.

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 ozone 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.

2.3 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, 1996c) for ozone 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. The form of the 8-hour NAAQS necessitates such an attainment test. The 8-hour NAAQS for ozone requires the fourth highest 8-hour daily maximum ozone concentration, averaged over three consecutive years, to be ≤ 0.08 ppm at each monitoring site⁹. The feature of the NAAQS requiring averaging over three years presents difficulties using the resource-intensive Eulerian models we believe are 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 many days and, perhaps, many strategies. This problem is reduced by using the *monitored* design value, as an inherent part of the modeled attainment test.

2. 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

⁹See 40CFR Part 50.10, Appendix I, paragraph 2.3. Because of the stipulations for rounding significant figures, this equates to a modeling target of ≤ 84 ppb. Because non-significant figures are truncated, a modeling estimate < 85 ppb is equivalent to ≤ 84 ppb.

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 an individual day are reduced by the new procedure.

3. 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 (i.e., the attainment test indicates that the strategy will not reduce ozone to 84 ppb or less), 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 NAAQS. These results address relative changes in the frequency and intensity of high modeled ozone 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.

4. 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. 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. However, there are several decisions which must be made before the recommended test can be applied. In this section, we identify a series of issues regarding selection of inputs to the test, and recommend solutions. We also describe, in more detail, the unmonitored area analysis procedure. Finally, we describe how to apply the test and illustrate this with examples.

Equation (3.1) describes the recommended modeled attainment test, applied near monitoring site I.

$$(\text{DVF})_I = (\text{RRF})_I (\text{DVB})_I \quad (3.1)$$

where

$(\text{DVB})_I$ = the baseline concentration monitored at site I, units in ppb;

$(\text{RRF})_I$ = the relative reduction factor, calculated near site I, unitless
The relative reduction factor is the ratio of the future 8-hour daily maximum concentration predicted near a monitor (averaged over multiple days) to the baseline 8-hour daily maximum concentration predicted near the monitor (averaged over the same days), and

$(\text{DVF})_I$ = the estimated future design value for the time attainment is required, ppb.

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

- (1) How is a “site-specific” baseline design value $((\text{DVB})_I)$ calculated?
- (2) In calculating the $(\text{RRF})_I$, 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 $(\text{RRF})_I$?
- (4) How do you calculate future design values in unmonitored areas?
- (5) Should any days be excluded when computing a relative reduction factor?
- (6) Which base year emissions inventory should be projected to the future for the purpose of calculating RRFs?
- (7) 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 8-hour NAAQS for ozone through use of monitored design values¹⁰. 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 8-hour ozone standard.
- 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 test we recommend using the average of the three design value periods (choice number 4 from above) which include the baseline inventory year. Based on the attributes listed above, the average of the three design value periods best represents the baseline ozone 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 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

¹⁰Design values at each monitoring site are calculated in accordance with 40 CFR Part 50.10, Appendix I. The design value is calculated as the 3 year average of the fourth highest monitored daily 8-hour maximum value at each monitoring site.

inventory 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). 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 design values was 4.0 ppb. The median change in the 5 year weighted average design values was only 0.8 ppb. These analyses show that the average design values are clearly more stable and will therefore provide a “best estimate” baseline year design value (DVB₁) for use in future year model projections.

The recommended averaging technique assumes that at least five complete years of ozone data is available at each monitor. In some cases there will be 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 values¹¹ to use in the attainment test

Example 3.1

Given:

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

¹¹The “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 averages of observed design values.

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	88.3¹²
2	86	84	91	87.0
3	88	86	85	86.3

3.2 Identifying surface grid cells near a monitoring site.

There are three 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, we believe that uncertainty in the formulation of the model and the model inputs is consistent with recognizing some leeway in the precision of the predicted location of 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 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.

¹² The average design value should carry one significant figure to the right of the decimal point. There are several calculations in the modeled attainment test which carry the tenths of a ppb digit. We have found that rounding and/or truncating ozone 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. 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 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 NO_x decreases may lead to ozone increases). 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, 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.

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 - 5¹³	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 reduction factor (RRF)_t near a monitor.

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.

¹³The appropriate size of the array for horizontal grid cells < 4km should be discussed with the appropriate U.S EPA Regional Office.

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

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

(3) The relative reduction factor for site I is

$$\begin{aligned}(\text{RRF})_I &= (\text{mean 8-hr daily max.})_{\text{future}} / (\text{mean 8-hr daily max.})_{\text{baseline}} \\ &= 82.1/95.2 = 0.862\end{aligned}$$

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 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 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 (outside of the monitor arrays) 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 ozone data combined with gridded modeled ozone outputs.

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

Ambient ozone data can be interpolated to provide a set of spatial fields. The spatial fields will provide an indication of ozone concentrations in monitored and unmonitored areas. But a simple interpolation of the ambient data cannot identify unmonitored areas with higher ozone 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 is 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 ozone values. But the absolute predictions from the model may not be entirely accurate. The model output is only as good 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 ozone 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. 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, as compared to assuming that the absolute predictions from the model are correct. For example, in areas where the model predicts relatively high ozone concentrations, the spatial fields can be adjusted upward. In areas where the model predicts relatively low ozone concentrations, the spatial fields can be adjusted downward. In this way, it may be possible to predict downwind unmonitored areas that may have high ozone concentrations. At the same time, ozone concentrations in rural areas, (which may be overly influenced by high monitored ozone near urban areas), may be adjusted downward. The combination of interpolated spatial fields and modeled output will be referred to as “model adjusted spatial fields”

3.4.2 Implementation of Model Adjusted Spatial Fields

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

- 1) Interpolate ambient ozone design value 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 model adjusted spatial fields.
- 4) Determine if any unmonitored areas are predicted to exceed the NAAQS in the future.

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. There is not a single recommended interpolation technique. EPA has provided example analyses in the past using the Kriging interpolation technique (U.S.EPA, 2004b). EPA's BenMAP software, which was used to create interpolated fields for the CAIR, uses the Voronoi Neighbor Averaging (VNA) technique (Abt, 2003).

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. The eVNA technique has been used in health benefits assessments (U.S. EPA, 2005a).

The next step is to create future year fields by multiplying the base year model 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 model adjusted spatial fields are then analyzed to determine if any grid cells are predicted to remain above the NAAQS.

EPA intends to provide software (similar to BenMAP) which will be able to spatially interpolate data (VNA), adjust the spatial fields based on model output (eVNA) and multiply the fields by model calculated RRFs. 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.

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 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. 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. Due to the uncertainty of the analysis, at a minimum, it is appropriate to commit to additional deployment of ozone monitors in areas where the unmonitored area analysis predicts future violations¹⁴. 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

¹⁴In most cases, States/Tribes can commit to additional emissions controls in lieu of additional monitoring in unmonitored areas.

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.

3.5 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 layers¹⁵. 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 reduction factor (RRF) is dependent on the magnitude of modeled current 8-hour daily maxima.

Figure 3.1 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 bias in the model results such that the model predicts less benefit from emissions reductions at lower concentrations. 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.

The analysis examined daily RRFs, but in practice, the RRFs are not calculated on a daily basis. 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

¹⁵See <http://www.epa.gov/cair/pdfs/finaltech02.pdf> for documentation of the base case and future year modeling.

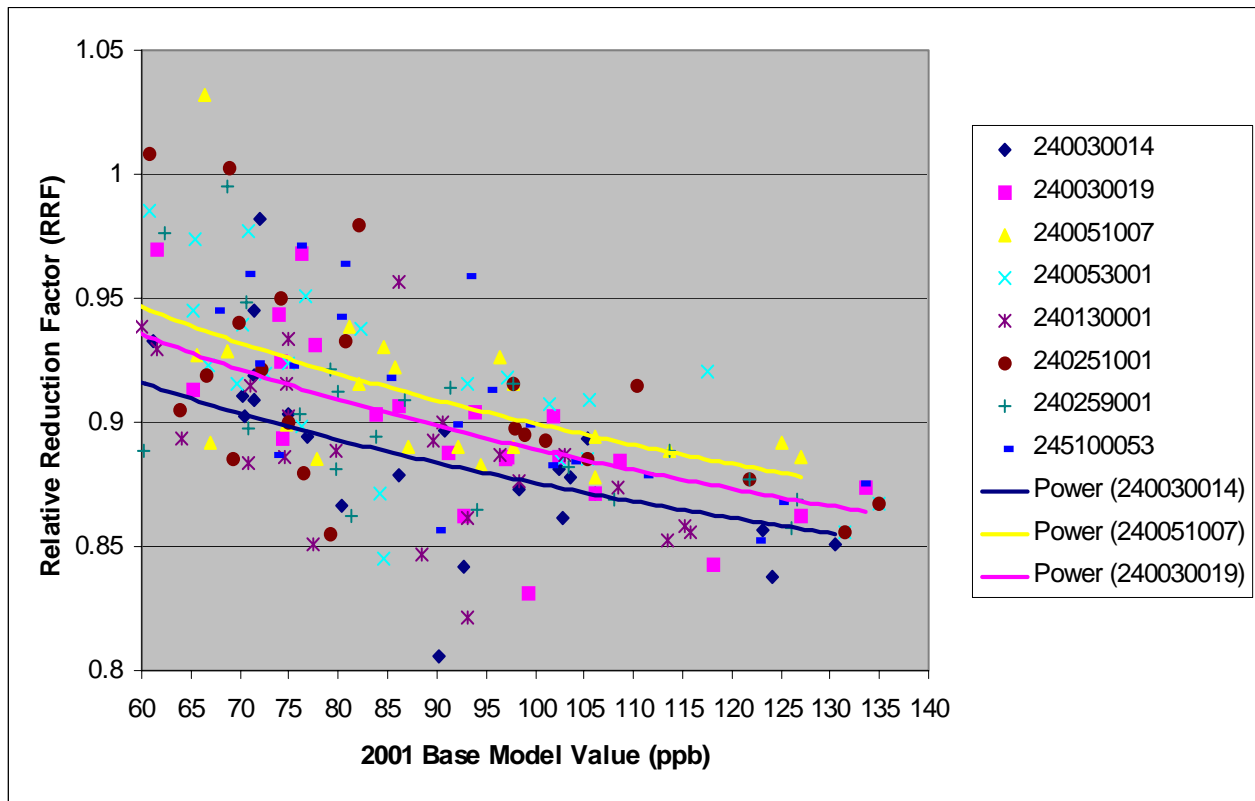


Figure 3.1 - Daily relative reduction factors as a function of daily maximum base modeled concentrations for monitors in the Baltimore nonattainment area.

concentration thresholds examined ranged from 70-85 ppb in 5 ppb increments¹⁶. 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 analysis¹⁷.

Table 3.3 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 larger. 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

¹⁶It 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 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 3.6).

¹⁷The same set of sites was used for each of the threshold concentrations (70-85 ppb).

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

As a result of the apparent bias 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 ppb¹⁸. This will result in less bias and provide for more robust RRFs and future design values.

Example 3.3

Example 3.3 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 ppb shown in Example 3.2).

Find: The relative reduction 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,

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

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

¹⁸The 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. It should be sufficient to judge the results based on modeled days that exceed the standard.

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

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

3.6 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 reduction factors”. It is important that emissions used in the attainment test correspond with the period reflected by the chosen design value period (e.g., 2000-2004). Deviations from this constraint will diminish the credibility of the relative reduction 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 1998 episode, “base case” emissions and meteorology would be for 1998. As described in Section 15, 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.

In section 14 we recommend using 2002 as the baseline inventory year for the current round of ozone SIPs. If States/Tribes use only episodes from 2002 (or the full 2002 ozone season) then the base case and baseline inventory years will be the same¹⁹. But if States/Tribes model 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

¹⁹ 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).

discouraged.

It is desirable to model meteorological episodes occurring during the period reflected by the baseline design value (e.g., 2000-2004). However, episodes 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 standard occur. The idea is to use selected representative episodes to capture sensitivity of predicted ozone to changes in emissions during commonly occurring conditions. There are at least three reasons why using episodes 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 ozone conditions compared to the baseline design value period and (3) availability of a past modeling analysis in which the model performed well.

3.7 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 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. Attainment dates are expressed as “no later than” three, five, six, or nine years after designation and nonattainment areas are required to attain as expeditiously as practicable. For example, moderate nonattainment areas that were designated on June 15, 2004, 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 advance their attainment date by at least a year. Requirements for the RACM analysis can be found in (U.S. EPA, 1999c).

For areas with an attainment date of no later than June 15th 2010, the emission reductions need to be implemented no later than the beginning of the 2009 ozone season. A determination of attainment will likely be based on air quality monitoring data collected in 2007, 2008, and 2009. Therefore, the year to project future emissions should be no later than the last year of the three year monitoring period; in this case 2009. 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 future projection year should be no later than 2008. 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.

3.8 How Do I Apply The Recommended Modeled Attainment Test?

States/Tribes should apply the modeled attainment test at all monitors within the nonattainment area plus other nearby counties within the State²⁰. 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 ≤ 0.08 ppm (i.e., ≤ 84 ppb using rounding conventions)²¹. Thus, if all resulting predicted future design values (DVF) are ≤ 84 ppb, the test is passed. The modeled attainment test is applied using 3 steps.

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 reduction factors. Use air quality modeling results to estimate a relative reduction 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. This has been illustrated in Examples 3.2 and 3.3. The relative reduction factor for site I is given by Equation 3.2.

$$(\text{RRF})_I = (\text{mean 8-hr daily max})_{\text{future}} / (\text{mean 8-hr daily max})_{\text{baseline}} \quad (3.2)$$

Using Equation (3.2), the relative reduction factor is calculated as shown in the column (5) in the last row of Table 3.3. 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 3.3, we have assumed that the same four days described previously in Example 3.3 have been simulated. Note that on day 3, model baseline 8-hour daily maximum ozone concentration was < 85 ppb. As discussed in Section 3.5, predictions for this day are not included in calculating the mean values shown in the last row of

²⁰States 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.

²¹40CFR Part 50.10, Appendix I, paragraph 2.3

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 reduction factors obtained in Step 2.

In Table 3.4, 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,

$$(DVF)_I = (102.0 \text{ ppb}) (0.867) = 88.4 \text{ ppb} = 88 \text{ ppb}$$

Note that the final future design value is *truncated*²² and in this example, the modeled attainment test is not passed at monitor site 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 reduction 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., 83.3/96.0)	88.4= 88 ppb

Table 3.4 Example Calculation of a Site-Specific Future Design Value (DVF)_I

²²This effectively defines attainment in the modeled test as ≤ 84.9 ppb and nonattainment as ≥ 85.0 ppb.

4.0 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 demonstrate that various emission reduction measures will bring an individual area into attainment of the 8-hour ozone standard 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 replace that, EPA believes that all attainment demonstrations will be strengthened by additional analyses that can help confirm that the planned emissions reductions will result in attainment.

Corroboratory 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., less than 82 ppb in the attainment year) will need more limited supporting material. For other attainment cases, in which the projected future design value is closer to 85 ppb, more supporting analyses should be completed. As noted earlier (see section 2.2), there may be some cases in which it is possible to either demonstrate attainment via a "weight of evidence" demonstration despite failing the model attainment test, or conversely, demonstrate that reaching attainment is not likely despite passing the model attainment test. This section of the guidance will discuss some specific additional analyses that can be used to corroborate the model projections, or refute them in the case of a weight of evidence determination. Additional examples of possible weight of evidence determinations are provided in existing EPA guidance (USEPA, 1996c). Again, it should be noted that no single corroborating analysis can serve as an adequate substitute for the air quality model, however, in aggregate, many such analyses can help inform the process.

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

Modeling: As discussed in Section 2, EPA has determined that the best approach to using models to demonstrate attainment of the 8-hour ozone standard is to use the 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 inventory²³. There are several metrics that can be considered as part of this type of additional analysis:

- percent change in total amount of ozone ≥ 85 within the nonattainment area
- percent change in grid cells ≥ 85 ppb within the nonattainment area
- percent change in grid cell-hours ≥ 85 ppb within the nonattainment area
- percent change in maximum modeled 8-hour ozone within the nonattainment area

While these metrics can be used to estimate the magnitude, frequency, and relative amount of eight-hour ozone 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) 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 show 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 in 8-hour ozone. 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 may help provide confidence that attainment will be reached.

In cases where attainment is demonstrated at all monitors, the above metrics are useful in showing the amount of expected reduction in other portions of the area where the unmonitored area analysis was applied. The results of the unmonitored area analysis can also be examined to indicate the magnitude and spatial extent of remaining nonattainment. If application of the modeled test in all grid cells indicates that most or all of the unmonitored areas will be attainment, then that information can support a modeled attainment demonstration and add positive evidence to a weight of evidence determination.

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. If model predicted future design values are close to or above the level of the NAAQS, placing a monitor in the area, may be the only way to address this issue in the future (absent of requiring additional emissions controls).

There are various other ways to use modeling results as supplemental evidence that supports

²³ Care should be taken in interpreting absolute metrics if the model evaluation shows a large underprediction or overprediction of 8-hour ozone concentrations. An underprediction of observed ozone concentrations will make it artificially easy to show progress towards absolute attainment levels and an overprediction of ozone will make it artificially difficult to show progress towards attainment.

(or questions) the modeled attainment test. These include, but are not limited to:

- use of available regional modeling applications that are suitable²⁴ 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,
- 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 reduction factors (RRF) and resulting projected design values to the variations in inputs or model formulations.
- use of the same modeled attainment demonstration but with future design values that are calculated in an alternative manner than that recommended in Section 3 of this guidance. Any alternate approaches for the calculation of the future design value 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.

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 8-hour ozone based on measured historical trends of air quality and emissions. There are several elements to this analysis that are difficult to quantify. First, the ambient data trends must be normalized to account for year-to-year meteorological variations. Second, one must have an accurate accounting of the year-to-year changes in actual emissions (NO_x and VOC) for the given area and any surrounding areas whose emissions may impact local ozone. Third, one must have a solid conceptual model of how ozone is formed in the local area (e.g., NO_x-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 “reduction factor” can be applied to the expected emissions reductions from a particular control strategy.

A simpler (and more uncertain) way to qualitatively assess progress toward attainment is to

²⁴ 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.

examine recently observed air quality and emissions trends. A downward trend in observed air quality and a downward trend in emissions (past and projected) is 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. EPA recently prepared a report that analyzed statistically significant trends in ozone (U.S. EPA, 2004c and U.S. EPA, 2005b) and ozone precursor emissions as well as a report which examined the . This report is a good template for States/Tribes considering similar analyses.

Observational Modeling: 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). There are numerous tools that can be used to determine whether ozone is sensitive to certain types of precursors (i.e., VOC or NO_x) or source sectors. 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. Additionally, receptor models, like chemical mass balance (CMB), positive matrix factorization (PMF), and Unmix may be useful for confirming whether a strategy is reducing the right sorts of sources (Maykut, 2003; Poirot et. al, 2001).

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.

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

As discussed in Section 2, it may be possible through the use of supplemental analyses to draw a conclusion differing from that implied by the modeled attainment test results. Past modeling analyses have shown that future design value uncertainties of 2-4 ppb 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 additional analyses, including the outcome of the 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 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 needs to be done 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 a WOE analysis to determine which additional analyses will be most appropriate for their particular area.

In a WOE determination, each type of analysis has an identified outcome that is consistent with the hypothesis that a proposed control strategy is sufficient to meet the NAAQS within the required time frame. 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 various 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 very close to the NAAQS (e.g., 82 to 87 ppb).
- 2) Multiple supplemental analyses in each of the three various categories discussed above (modeling, ozone/emissions trends, observational models).
- 3) A 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) A written description as to why the aggregate analyses leads to a conclusive determination regarding the future attainment status of the area that differs from the modeled attainment test.

5.0 What Additional Analyses 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). 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. 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 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.

Measurement of “indicator species” is a potentially useful means for assessing which precursor category (VOC or NO_x) limits further production of ozone at a monitor’s location at various times of day and under various sets of meteorological conditions. 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 NO_x. Comparisons are done by looking at ratios of these species. The following appear to be the most feasible for use by a regulatory agency: O₃/NO_y, O₃/(NO_y - NO_x) and O₃/HNO₃. Generally, high values for the ratios suggest ozone is limited by availability of NO_x emissions. Low values suggest availability of organic radicals (e.g., attributable to VOC emissions) may be the limiting factor. For these ratios to be most useful, instruments should be capable of measuring NO_y, NO_x, NO₂ and/or HNO₃ with high precision (i.e., greater than that often possible with frequently used “routine” NO_x measurements). Thus, realizing the potential of the “indicator species method” as a tool for model performance evaluation and for diagnosing why observed ozone concentrations do or do not meet previous expectations may depend on deploying additional monitors and/or measurements. 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.

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. 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, NO_y, NO, NO₂, as well as several relatively stable species like CO and selected VOC species. Measurements of SO₂ 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). This provides greater assurance that the air quality model correctly reflects the configuration of sources contributing to ozone formation.

Collecting locally applicable speciated emissions data. While the U.S. EPA maintains a library of default VOC emissions species profiles (U.S. EPA, 1993), 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 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.

Projecting emission estimates and comparing these to subsequent emission estimates. States/Tribes addressing traditional nonattainment areas with lengthy attainment dates may find it worthwhile to project emissions to multiple future years and retain the resulting data files for use in subsequent reviews. Intermediate projections could be useful during a mid-course review to help diagnose reasons for subsequent observed ozone trends which are inconsistent with earlier expectations obtained with the air quality model. Retention of projected emission data bases would enable States/Tribes to compare the projected inventory estimates with an inventory which is subsequently updated. These checks would be possible after the inventory updates for 2005 or 2008 become available.

Future diagnostic analyses using air quality models. To facilitate a subsequent, mid-course review, States/Tribes should retain all meteorological input data as well as current (e.g., 2002) and projected (e.g., 2009) emission input files developed to support the needed SIP revisions. When a model is applied with updated emissions estimates (e.g., 2009 projections from the later versions of national inventories) and/or with updated meteorological inputs indicative of more recent episodes, several useful comparisons are possible if the old files are retained. A State/Tribe would be better able to determine whether differences in observed and past-predicted air quality are explained by revised emission estimates, differences in meteorological episodes, or by changes which have occurred in the model formulation during the intervening years. Insights from such comparisons should help a State/Tribe explain why changes in the strategy reflected in its MCR SIP revision may or may not be necessary.

6.0 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 6.1 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 6.1.

Table 6.1 Recommended Documentation for Demonstrating Attainment of the 8-hour NAAQS for Ozone

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; Processes, conditions, and influences for ozone formation.
Modeling/Analysis Protocol	Communicate scope of the analysis and document stakeholder involvement.	Names of stakeholders participating in preparing and implementing the protocol; Types of analyses performed; Steps followed in each type of analyses; Rational 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, CO emissions by State/County for major source categories. Quality assurance/quality control procedures

Table 6.1 Recommended Documentation for Demonstrating Attainment of the 8-hour NAAQS for Ozone (continued)

Subject Area	Purpose of Documentation	Issues Included
Air Quality/Meteorology Preparations and Results	Assurance that representative air quality and meteorological inputs are used in analyses	<p>Description of data base and procedures used to derive and quality assure inputs for modeling;</p> <p>Departures from guidance and their underlying rationale.</p> <p>Performance of meteorological model used to generate meteorological inputs to the air quality model.</p>
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.	<p>Summary of observational data base available for comparison;</p> <p>Identification of performance tests used and their results;</p> <p>Ability to reproduce observed temporal and spatial patterns;</p> <p>Overall assessment of what the performance evaluation implies.</p>
Diagnostic Tests	Ensure rationale used to adjust model inputs or to discount certain results is physically justified and the remaining results make sense.	<p>Results from application prior to adjustments;</p> <p>Consistency with scientific understanding and expectations;</p> <p>Tests performed, changes made and accompanying justification;</p> <p>Short summary of final predictions.</p>

Table 6.1 Recommended Documentation for Demonstrating Attainment of the 8-hour NAAQS for Ozone (continued)

Subject Area	Purpose of Documentation	Issues Included
<p>Description of the Strategy Demonstrating Attainment</p>	<p>Provide the EPA and the public an overview of the plan selected in the attainment demonstration.</p>	<p>Qualitative description of the attainment strategy;</p> <p>Reductions in VOC, NOx, and/or CO emissions from each major source category for each State/county/Tribal land from current (identify) emission levels;</p> <p>Clean Air Act mandated reductions and other reductions;</p> <p>Show predicted 8-hour future design values for the selected control scenario and identify any location(s) which fails the unmonitored area test described in Section 3;</p> <p>Identification of authority for implementing emission reductions in the attainment strategy.</p> <p>Evidence that emissions remain at or below projected levels throughout the 3-year period used to determine future attainment.</p>
<p>Data Access</p>	<p>Enable the EPA or other interested parties to replicate model performance and attainment simulation results, as well as results obtained with other analyses.</p>	<p>Assurance that data files are archived and that provision has been made to maintain them;</p> <p>Technical procedures for accessing input and output files;</p> <p>Identify computer on which files were generated and can be read, as well as software necessary to process model outputs;</p> <p>Identification of contact person, means for downloading files and administrative procedures which need to be satisfied to access the files.</p>

Table 6.1 Recommended Documentation for Demonstrating Attainment of the 8-hour NAAQS for Ozone (concluded)

Subject Area	Purpose of Documentation	Issues Included
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 by the required time.	<p>Description of the modeled attainment test and observational data base used;</p> <p>Identification of air quality model used;</p> <p>Identification of other analyses performed;</p> <p>Outcome of each analysis, including the modeled attainment test;</p> <p>Assessment of the credibility associated with each type of analysis in this application;</p> <p>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.</p>
Review Procedures Used	Provide assurance to the EPA and the public that analyses performed in the attainment demonstration reflect sound practice	<p>Scope of technical review performed by those implementing the protocol;</p> <p>Assurance that methods used for analysis were peer reviewed by outside experts;</p> <p>Conclusions reached in the reviews and the response.</p>

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

7.0 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 ozone NAAQS 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 8 - 15.

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 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? 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 4 and 5 (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. 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 six steps needed to generate the modeling results and then apply the modeled attainment and screening tests 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 formation, transport, and removal during high ozone episodes, (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.

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 national ambient air quality standard and of the modeled attainment test described in Section 3. Finally, it requires a review of meteorological conditions which have been observed to accompany monitored exceedances of the 8-hour ozone NAAQS. The object of these reviews is to select time periods which: a) include days with observed concentrations close to site-specific design values and b) reflect a variety of meteorological conditions which have been commonly observed to accompany monitored exceedances. This latter objective is desirable, because it adds confidence that a proposed strategy will work under a variety of conditions.

Due to increased computer speeds, it is now prudent to recommend modeling relatively long time periods. 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 ozone season may simplify the episode selection process and provide a rich database with which to apply the modeled attainment test.

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 (i.e., ozone) data corresponding to the time periods that will be modeled need to be reviewed prior to choosing size of the area 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 of ozone precursors 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 episodes, (2) a baseline corresponding to that represented by the current monitored design value, (3) a future base case when attainment of the NAAQS needs to be demonstrated, and (4) control scenarios in which 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 air quality in a historical case. 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. In the past, the performance evaluation 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 concentrations with different combinations of these inputs. There is no guarantee that ozone 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 ozone 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 test recommended in Section 3. 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 ozone 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 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 ozone days. 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 calculating relative reduction 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 base year baseline. If attainment cannot be shown, then additional model runs which contain control measures are needed. Multiple future year control strategy runs may need to be completed until the attainment test is passed.

8.0 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 ozone monitoring site, as described in Section 3. 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 problem in the area which is the focus of a modeled attainment 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 attainment 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.

8.1 What Is A “Conceptual Description”?

A “conceptual description” is a qualitative way of characterizing the nature of an area’s nonattainment problem. It is best described by identifying key components of a description. Examples are listed below. The examples are not necessarily 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.

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 NO_x 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 is expected to benefit from the regional NO_x reductions from the NO_x 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 NO_x emissions throughout the regional domain. Reduced NO_x 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 ≥ 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 NO_x 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 NO_x emissions separately and at the same time.

It should be clear from the preceding example that the initial conceptual description of an area's nonattainment problem may draw 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 9.

8.2 What Types Of Analyses Might Be Useful For Developing And Refining A Conceptual Description?

Questions like those posed in Section 8.1 can be addressed using a variety of analyses ranging in complexity from an inspection of air quality data to sophisticated mathematical analyses. We anticipate the simpler analyses will often be used to develop the initial conceptual

description. These will be followed by more complex approaches or by approaches requiring more extensive data bases as the need later becomes apparent. In the following paragraphs, we revisit key parts of the conceptual description identified in Section 8.1. We note analyses which may help to develop a description of each part. The list serves as an illustration. It is not necessarily exhaustive.

8.2.1. Is regional transport an important factor affecting the nonattainment area?

- Are there other nonattainment areas within a day's transport of the nonattainment area?
- Do "upwind" 8-hour daily maximum ozone concentrations approach or exceed 84 ppb on some or all of the days with observed 8-hour daily maxima > 84 ppb in the nonattainment area?
- Are there major sources of emissions upwind?
- What is the size of the downwind/upwind gradient in 8-hour daily maximum ozone concentrations compared to the upwind values?
- Do ozone concentrations aloft but within the planetary boundary layer approach or exceed 84 ppb at night or in the morning hours prior to breakup of the nocturnal surface inversion?
- Is there a significant positive correlation between observed 8-hour daily maximum ozone concentrations at most monitoring sites within or near the nonattainment area?
- Is the timing of high observed ozone consistent with impacts estimated from upwind areas using trajectory models?
- Do available regional modeling simulations suggest that 8-hour daily maximum ozone concentrations within the nonattainment area respond to regional control measures?
- Does source apportionment modeling indicate significant contributions to local ozone from upwind emissions?

8.2.2. What types of meteorological episodes lead to high ozone?

- Examine the spatial patterns of 8-hour daily maxima occurring on days where the ozone is > 84 ppb and try to identify a limited number of distinctive patterns.
- Review synoptic weather charts for days having observed concentrations > 84 ppb to identify classes of synoptic scale features corresponding to high observed ozone.
- Perform statistical analyses between 8-hour daily maximum ozone and meteorological measurements at the surface and aloft to identify distinctive classes of days corresponding with

observed daily maxima > 84 ppb.

8.2.3. Is ozone limited by availability of VOC, NO_x or combinations of the two? Which source categories may be most important?

- What are the major source categories of VOC and NO_x and what is their relative importance in the most recent inventory?

- Review results from past modeling analyses to assess the likelihood that ozone in the nonattainment area will be more responsive to VOC or NO_x controls. Do conclusions vary for different locations?

- Apply modeling probing tools (e.g., source apportionment modeling) to determine which source sectors appear to contribute most to local ozone formation.

- Apply indicator species methods such as those described by Sillman (1998, 2002) and Blanchard (1999, 2000, 2001) at sites with appropriate measurements on days with 8-hour daily maximum ozone exceedances. Identify classes of days where further ozone formation appears limited by available NO_x versus classes of days where further ozone formation appears limited by available VOC. Do the conclusions differ for different days? Do the results differ on weekdays versus weekends?

- Apply receptor modeling approaches such as those described by Watson (1997, 2001), Henry (1994) and Henry (1997a, 1997b, 1997c) to identify source categories contributing to ambient VOC on days with high observed ozone. Do the conclusions differ on days when measured ozone is not high?

Additional analyses may be identified as issues arise in implementing a modeling/analysis protocol. 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 4, corroboratory analyses should be used to help assess whether a simulated control strategy is sufficient to meet the NAAQS.

9.0 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 ozone NAAQS, 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.

9.1 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.

9.2 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 analyses
 - e. Description of the conceptual model for the nonattainment area
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. Geographic area identified for application of the attainment test(s)
 - g. 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

10.0 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 for the 8-hour ozone NAAQS. 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. Meteorological and emissions models are discussed in Sections 13 and 14, respectively.

10.1 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. These general criteria are consistent with requirements in 40 CFR 51.112 and 40 CFR part 51, Appendix W (U.S. EPA, 2003). Note that, unlike in previous guidance (U.S. EPA, 1991a), we are not recommending a specific model for use in the attainment demonstration for the 8-hour NAAQS for ozone. At present, there is no single model which has been extensively tested and shown to be clearly superior than its alternatives. Thus, 40CFR Part 51 Appendix W does not identify a “preferred model” for use in attainment demonstrations of the 8-hour NAAQS for ozone. Based on the language in 40CFR Part 51 Appendix W, States/Tribes should consider nested regional air quality models or urban scale air quality models as “applicable models” for ozone.

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 code²⁵ 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 ozone attainment demonstration.

²⁵Air 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 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.

10.2 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 for the 8-hour ozone NAAQS. 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. 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 ozone 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 10.1, a State/Tribe may use the factors described in this section (Section 10.2) to show that it is appropriate for use in a specific application. Choosing 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 15.

10.3 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 10.1 lists several current generation air quality models which have been used to simulate ambient ozone concentrations. Table 10.2 lists several air quality models which have been used for various ozone applications over the past decade, but are not widely used at this time. 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 for the ozone NAAQS. 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 10.1 and 10.2 in selecting an air quality model for a specific application.

Table 10.1 Current Air Quality Models Used To Model Ozone

Air Quality Model	References
CAMx	Environ (2004)
CMAQ	U.S. EPA (1998a)
UAM-V	Systems Applications International (1996)

Table 10.2 Other Air Quality Models Used to Model Ozone

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)

11.0 How are the Meteorological Time Periods (Episodes) Selected?

Historically, ozone 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, several groups have recently modeled entire summers or even full years (Baker, 2004).

Additionally, recent research has shown that model performance evaluations and the response to emissions controls need to consider modeling results from 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 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. The following sections contain further recommendations for choosing appropriate time periods to model for attainment demonstrations.

At a minimum, four criteria should be used to select episodes²⁶ which are appropriate to model:

- 1) Choose a mix of episodes reflecting a variety of meteorological conditions which frequently correspond with observed 8-hour daily maxima > 84 ppb at multiple monitoring sites.
- 2) Model periods in which observed 8-hour daily maximum concentrations are close to the average 4th high 8-hour daily maximum ozone concentrations.
- 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 11.1.4).

These four criteria may often 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 ozone concentrations close to site-specific design values during the base period. Thus, tradeoffs among the four primary criteria may be necessary in specific applications.

²⁶If modeling a full ozone season or year, the ambient ozone data and meteorology should be evaluated to determine if the suggested criteria are met.

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, may result in its being chosen over an alternative. Another consideration should be to choose episodes occurring during the 5-year period which serves as the basis for the baseline average design value (DVB). If observed 8-hour daily maxima > 84 ppb occur on weekends, weekend days should be included within some of the selected episodes. 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 episodes 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.

11.1 What Are The Most Important Criteria For Choosing Episodes?

11.1.1 Choose a mix of episodes which represents a variety of meteorological conditions which frequently correspond with observed 8-hour daily maxima exceeding 84 ppb.

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. 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 8-hour daily maxima for ozone, 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 ozone 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 ozone 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 ozone.

11.1.2 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). Note that we said, “at the time of the selected episode” (i.e., the “base case period”) rather than “current or baseline period” in the preceding sentence. 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.

Air quality measurements recorded during the baseline/current period can also 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 other causes, 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 “ ± 10 ppb” as a default recommendation for purposes of prioritizing

choice of episodes²⁷. If the base and baseline/current periods do not coincide, “close to” is within ± 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 ≥ 85 ppb should receive greatest priority.

11.1.3 Choose days with intensive data bases.

Preference should be given to days with measurements aloft, available measurements of indicator species (see Section 15) 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.

11.1.4 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 3.2 indicates that the relative reduction 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 3.5) 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 number 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 days²⁸. The

²⁷The analysis in section 3.5 showed that low ozone predictions can introduce a bias in the relative results from the model. 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.

²⁸The 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.

standard deviation of the daily RRFs was used to create the datasets and measure the variability of the RRFs.

Figure 11.1 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%)²⁹. 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

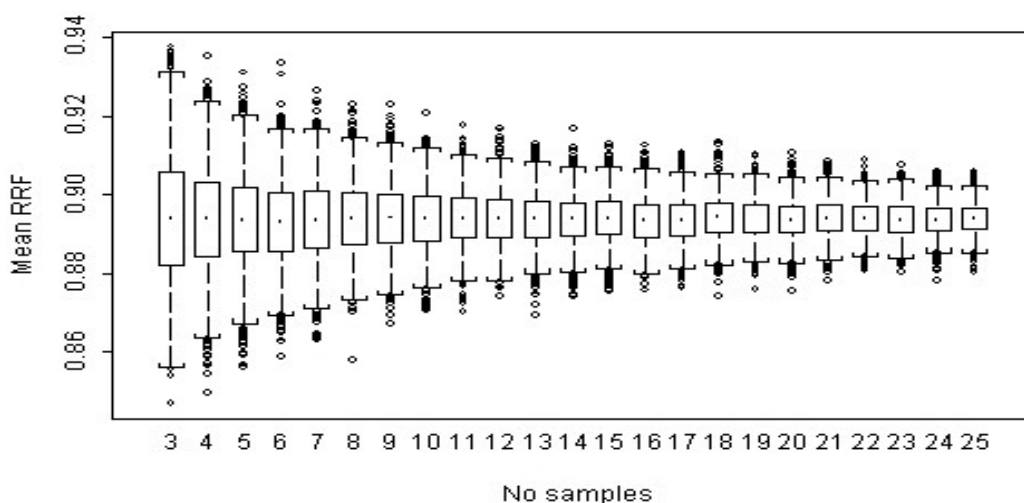


Figure 11.1- Mean RRF as a function of the subset of days (3-25 days) for a Harford County, MD ozone monitoring site. The full dataset was 50 days.

that the RRF is more variable when based on a small number of days.

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.

²⁹The standard deviation is in “RRF units”. For example, an RRF of 0.90 is equal to a 10% ozone reduction. A standard deviation of 3.4% is a measure of the variability such that $\pm 3.4\%$ is equal to a range in mean RRF of 0.866-0.934.

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 $\pm 1\%$ and $\pm 2\%$ 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 11.1 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 $\pm 1\%$ and $\pm 2\%$ 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 $\pm 1\%$ (95% of the time) and a minimum number of 5-6 days is needed to replicate the mean RRF to within $\pm 2\%$ (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	$\pm 1\%$ (25/50/100 days)	$\pm 2\%$ (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 11.1- Number of days needed to replicate the 25/50/100 day dataset mean RRF to within $\pm 1\%$ and $\pm 2\%$, 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 $\pm 1\%$ (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 3.5 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, 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 ppb³⁰.

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	Theshold < 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 11.2- 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.

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

³⁰Any 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).

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.

11.2 What Additional, Secondary Criteria May Be Useful For Selecting Episodes?

In Section 11.1, 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 episodes which have already been modeled. That is, of course, provided that past model performance evaluation for such an episode was successful in showing that the model worked well in replicating observations. Given that the four primary criteria are met approximately as well by such episodes as they are by other candidate episodes, a State/Tribe could likely save a substantial amount of work in evaluating model performance. However, it should be noted that large changes in ozone precursor levels or ratios may make the use of older episodes undesirable.

Choose episodes 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 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 possible in the modeled attainment test.

It is desirable to include weekend days among those chosen, especially if concentrations greater than 84 ppb are observed on weekends. Weekend days often reflect a different mix of emissions than occurs on weekdays³¹. This could also lead to different spatial patterns of 8-hour daily maxima in excess of 84 ppb. 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 do differ on weekends versus 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

³¹ <http://www.arb.ca.gov/aqd/weekendeffect/weekendeffect.htm>

simultaneously, choose episodes which meet the primary and secondary criteria in as many of these nonattainment areas as possible. As discussed in Section 10, 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 in several nonattainment areas at once. Time and resources needed for this effort could be reduced by choosing episodes which meet the primary and secondary criteria in several nonattainment areas which are modeled.

12.0 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/ozone 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.

12.1 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. We expect that most 8-hour ozone attainment demonstrations will utilize a regional nested grid modeling approach.

The principal determinants of model domain size are the nature of the ozone problem and the scale of the emissions which impact the nonattainment area. Isolated nonattainment areas that are not impacted by regional ozone and ozone precursors may be able to use a relatively small domain. 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 ozone precursors from hundreds of miles or more upwind of the receptor area (U.S EPA, 1998b). 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. In most cases, the modeling domain should be large enough to allow the use of clean or relatively clean boundary conditions.

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 during ozone episodes.
 - 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) due to the fact that air that leaves the fine grid never returns. 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.

12.2 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 at least 2-3 days prior to a period of interest to diminish the importance of initial conditions³². In this way, relatively clean static initial

³²Sensitivity 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.

conditions can be used to initialize the model. For nested model applications, initial conditions can be specified using model predictions from the outer grid if the nested grids are started a couple of days 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 12.1, 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 model³³. Another option is to use default initial and boundary concentration profiles representing relatively clean conditions which have been formulated from available measurements and 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.

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 not change in the future. The use of lowered initial or boundary conditions in the future year should be documented and justified.

12.3 What Horizontal Grid Cell Size Is Necessary?

As we discuss in Section 13, most applications will use a prognostic meteorological model to provide meteorological inputs needed to make air quality estimates. Typically, these models are

³³One 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, et al., 1999; Jaffe, et al., 2003; Fiore, et al., 2003).

set up to produce meteorological fields for nested grids with a 3:1 ratio. In past ozone 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.

In past guidance, we have recommended using horizontal grid cell sizes of 2-5 km in urban scale 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 NO_x 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 several studies, the model performance at 12 km resolution was equal to or better than performance at 4 km resolution (Irwin, 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 reduction factors in predicted 8-hour daily maxima 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, 2004). 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 NO_x “disbenefits” (i.e., ozone increases in oxidant limited areas when NO_x emissions are reduced). 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, 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 NO_x should be

considered as an alternative with coarser than desired grid cells.

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. Thus, 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.

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 ordinarily apply plume-in-grid algorithms to major point sources of NO_x. The use of plume-in-grid should be discussed with the appropriate EPA Regional Office.

12.4 How Should the Vertical Layers Be Selected?

As described in Section 13, 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) with a “one atmosphere” model which accounts 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 the 100 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 concentrations will be affected by how well the model is able to characterize dilution of precursors and ozone. This, in turn, depends in part on how precisely the model can estimate maximum afternoon 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). Because maximum mixing heights may vary on different days and it is necessary to simulate numerous days and locations, model predictions can be influenced by the number of 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 ozone precursors on high ozone days. Therefore, vertical resolution above the boundary layer is typically much coarser.

There is no correct minimum number of vertical layers needed in an ozone 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 et. al, 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.

13.0 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, in part, determine 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 episodes) 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, meteorological data sets sufficient for air quality modeling purposes. Additional suggestions are provided to assist in the configuration of the meteorological modeling analysis. The last section outlines procedures for evaluating whether the meteorological input is of sufficient quality for input into the air quality model. It is recommended that States/Tribes spend considerable effort in accurately characterizing the meteorological fields in view of several sensitivity runs which show that relatively small differences in meteorological inputs can have large impacts on resultant air quality modeling results (Dolwick, 2002).

13.1 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 steer (i.e., "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 include, but are not limited to, MM5 (Grell et al., 1994) and RAMS (Pielke et al., 1992). Recent advances in relatively low-cost computational power have resulted in widespread use of MM5 and RAMS for air pollution applications over the past decade (Olerud et al., 2000; Doty et al., 2001; Johnson, 2003, Baker, 2004) and EPA expects that all future attainment demonstration analyses will be based on data from these types of prognostic meteorological models. 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 the Eta, Rapid Update Cycle (RUC), and/or the Weather Research Forecast (WRF) model 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 et al., 2004).

13.2 How Should the Meteorological Modeling Analysis be Configured?

As with other parts of the air quality modeling system, determining how to configure the meteorological modeling can affect the quality and suitability of the air quality model predictions. Decisions regarding the configuration of complex prognostic meteorological models can be particularly challenging because of the amount of flexibility available to the user. The following are recommendations on how to configure a meteorological model for air quality analyses.

Selecting a Model Domain: As noted in Section 12, 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 ozone locally. In terms of selecting a meteorological modeling domain, one should extend the grid 3-6 cells beyond the bounds 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 meteorological models use the same grid resolution as desired for the air quality model applications. In some cases, however, this may not 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 situations it is recommended that the 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) to interpolate more finely resolved meteorological fields for air quality modeling.

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 a series of up front sensitivity tests which consider various combinations of physics options over specific time periods and regions. While these tests may not ultimately conclude that one configuration is clearly superior at all times and in all areas, it is recommended that sensitivity tests be completed as they should lead to a modeling analysis that is suited for the domain and period being simulated. Examples of sensitivity analyses can be found in (Olerud, 2003) and (McNally, 2002). 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.

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 ozone 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.

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, 2005), and RAMSCAMx (Environ, 2005). 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 missing yet needed variables.

13.3 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 parameters³⁴ as compared to observed data. Typical statistical comparisons of the key meteorological parameters 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 agreement. For modeling exercises over large domains and entire ozone seasons, it is recommended that the operational evaluation be broken down into individual segments such as geographic subregions 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 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.

It may be helpful when calculating domainwide and/or regional summary statistics to compare the results against previously accomplished 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. 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.

³⁴ It is difficult to say which meteorological parameters will most affect any particular modeling exercise. 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.

Phenomenological Evaluation: As discussed in Chapter 8, it is recommended that a conceptual description of the area's ozone problem be developed prior to the initiation of any air quality modeling study. Within the conceptual description of a particular modeling exercise, it is recommended that the specific meteorological parameters 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 phenomena³⁵ 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 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.

³⁵ Possible examples include: lake/sea breezes, low-level jets, amount of convection on a given day.

14.0 How Are the Emission Inputs Developed?

Air quality modeling for 8-hour ozone requires emission inputs for base case, baseline, and future modeling years. As explained in the EPA Emission Inventory Guidance (U.S. EPA, 2005c), 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 for the 8-hour ozone standard³⁶. The future-year depends on the nonattainment classification of the individual State or Tribe, as described in Section 3. Note that emissions data should be consistent with the data used in the modeled attainment test, described in Section 3.

Preparation of emissions data for air quality models for the base and future years requires several steps. First, States/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 must include a complete accounting of anthropogenic and biogenic VOC, NO_x, and CO. 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 14.1 and 14.2 summarize the issues for preparing emission inventories. Section 14.3 describes the needs for ancillary data. Section 14.4 summarizes the emissions modeling steps. Section 14.5 and Section 14.6 summarize the issues associated with modeling of future year emissions data.

14.1 Can The National Emission Inventory Be Used As a Starting Point?

It is recommended that States/Tribes start with available inventories suitable for air quality modeling of the modeling episode(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

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.

inventory development³⁷. However, the detail on emissions in the NEI may not be sufficient for use in attainment demonstration modeling. 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.

14.2 What Emission Inventory Data are Needed to Support Air Quality Models?

Emission inventory data from five 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, 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.

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/on-road/latest_version.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 14.3, link-based emissions can be based on Travel Demand Models (TDMs).

³⁷The final 2002 NEI is expected to be available at the end of 2005. Until the final NEI is available, the final 1999 (version 3) NEI or the draft 2002 NEI can be used as a source of national inventory data.

Nonroad Mobile- For nonroad mobile sources, the emissions should be compiled as country, state/tribe, county and SCC totals using the most current version of EPA’s NONROAD model (<http://www.epa.gov/oms/nonrdmdl.htm>) or alternative model.

Biogenic Emissions - Biogenic emissions from plants and soil contribute VOC and NO_x emissions and are best calculated as part of the emissions modeling step, as described in Section 14.4. These emissions require land use data, which is described as an ancillary dataset in Section 14.3. 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 an average-summer-day inventory. In any case, the temporal allocation step during emissions modeling (see Sections 14.3 and 14.4) 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 NO_x emissions and exit gas flow rates³⁸.

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, 2005c). 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, seven documents have been issued by the Emission Inventory Improvement Program (EIIP) for use in inventory development:

- Volume I: Introduction and Use of EIIP Guidance for Emissions Inventory Development (U.S. EPA, 1997a)

³⁸Day 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). Therefore, certain day specific inventory information should be removed and replaced with average data in the baseline inventory, before projecting the baseline to the future. This issue is not a concern for day-specific mobile source or biogenic emissions data which may be dependent on day specific (or even hourly) meteorological data for the time periods modeled.

- Volume II: Point Sources Preferred and Alternative Methods (U.S. EPA, 1997b)
- Volume III: Area Sources Preferred and Alternative Methods (U.S. EPA, 1997c)
- Volume IV: Mobile Sources preferred and Alternative Methods (U.S. EPA, 1997d)
- Volume V: Biogenics Sources Preferred and Alternative Methods (U.S. EPA, 1997e)
- Volume VI: Quality Assurance Procedures (U.S. EPA, 1997f)
- Volume VII: Data Management Procedures (U.S. EPA, 1997g)

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.

14.3 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/>.

The emissions models also need information about the chemical species of the VOC emissions for stationary point, stationary area, and all mobile sources. These data are used to disaggregate the total VOC 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 speciation data, which are available at: <http://www.epa.gov/ttn/chief/emch/speciation/>. For large or critical VOC sources in the modeling domain, modelers should consider determining the individual chemical compounds contributing to the total VOC. 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) and the MIMS Spatial Allocator (<http://www.cep.unc.edu/empd/projects/mims/spatial/>), which calculate the fraction of countywide emissions to allocate to each grid cell based on the surrogate type. 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 use patterns, population growth, and demographics, however, the impact and utility of such approaches is not well characterized³⁹.

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. 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 characterized⁴.

³⁹At 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.

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 14.2. 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.

14.4 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 14.3, 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 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). 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 et al., 2000; Strum et al., 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 NO_x SIP call rule to reduce regional NO_x emissions (U.S. EPA 1997h), 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).

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, NO_x, 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.

Other models for biogenic emissions include the BIOME model that is a part of EMS-2001 and the Global Biosphere Emissions and Interactions System (GloBEIS). The latter estimates emissions from natural (biogenic) sources and is designed for use in combination with photochemical modeling systems, such as CAMx (<http://www.globeis.com/>).

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.

14.5 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 layer⁴⁰ (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.

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 14.3.) 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

⁴⁰Point 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.

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.

14.6 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 over the next five to twenty years. For details on which future year(s) should be modeled for attainment demonstrations, refer to Section 3.

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 14.6.1).
- Collect the available data and models that can be used to project emissions for each of the sectors (Section 14.6.2).
- 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 14.6.3)
- Create inputs needed for emissions models, create future-year inventories, quality assure them, and create air quality model inputs from them (Section 14.6.4)

The remainder of this section provides additional details about these steps.

14.6.1 Identify and prioritize key sectors for projections

Emissions modelers should evaluate their inventories for those sectors 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 modeling 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.

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 this cases, it may be advisable to either leave such emissions out of the modeling or create an “average year” inventory and temporal distribution of the emissions.

14.6.2 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.

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. This issue was also noted in footnote 3 in Section 14.2. 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](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 superseded 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>.

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 14.6.3 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 show 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. 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>.

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>.

14.6.3 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. NO_x 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.

14.6.4 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 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.

15.0 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. 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 ozone and/or precursors (surface and aloft), and (2) how accurate is the model in characterizing sensitivity of ozone 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 test recommended in Section 3 uses models to predict the response of ozone to controls and then applies the resulting relative reduction factors to *observed* (rather than modeled) ozone. 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 ozone 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 identify and 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 concludes by identifying several 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 levels.

15.1 What are the Procedures for Evaluating An Air Quality Model?

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 reduction 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. For the 8-hour ozone modeling,

States/Tribes should compare all 1-hour observations and predictions (above a certain threshold), as well as all observed and predicted 8-hour daily maxima. Generally, if model performance is acceptable for the hourly pairs, one would expect the 8-hour performance to be acceptable as well.

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”. If ratios of observed indicator species are very high or very low, they provide a sense of whether further ozone production at the monitored location is likely to be limited by availability of NO_x or VOC (Sillman, 1995). Agreement between paired observed and predicted high (low) ratios suggests a model may correctly predict sensitivity of ozone 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 sensitivity of ozone to changes in emissions is to compare model projections after the fact 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 15.3. We recommend that diagnostic analyses be performed during the initial phase of the model application and during any mid-course review.

15.2 How Should the Operational Evaluation of Performance Be Completed?

This section describes the recommended statistical measures and other analytical techniques which should be considered as part of an operational evaluation of ozone model performance. Note that model predictions from the ramp-up days should be excluded from the analysis of model performance. It is recommended that, at a minimum, the following three statistical measures be calculated for hourly ozone and 8-hourly maxima 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 15.1.
- 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 15.2.
- 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 15.1 and 15.2), 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 reduction 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.

Equation 15.1

$$\text{MNB} = \frac{1}{N} \sum_1^N \left(\frac{(\text{Model} - \text{Obs})}{\text{Obs}} \right) \cdot 100\%$$

Equation 15.2

$$\text{MNGE} = \frac{1}{N} \sum_1^N \left(\frac{|\text{Model} - \text{Obs}|}{\text{Obs}} \right) \cdot 100\%$$

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 ppb⁴¹, and 2) for all pairs (no threshold)⁴². 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 locations⁴³. States/Tribes should

⁴¹ 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.

⁴² 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.

⁴³ 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

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.

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.

Other statistics such as normalized mean bias, normalized mean gross error, fractional bias, fractional error, root mean square error, and correlation coefficients should also be calculated to the extent that they provide meaningful information (see Appendix A for definitions). Wherever possible, these types of performance measures should also be calculated for ozone precursors and related gas-phase oxidants (NO_x, NO_y, CO, HNO₃, H₂O₂, VOCs and VOC species, etc.) and ozone (and precursors) aloft.

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.
- 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 model⁴⁴.
- 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

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.

⁴⁴Quantile-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.

observed hourly or daily maximum concentrations on the predicted isopleths reveals useful information on the spatial alignment of predicted and observed plumes.

- 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).

15.3 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 ozone resulting from changes in ozone precursor emissions. States/Tribes are encouraged to complete as many of these types of analyses as possible, in order to increase confidence in the modeled attainment projections.

Observational models: In Section 5 it was noted that measurements of certain “indicator species ratios” are a potentially useful way to assess whether local ozone formation is VOC- or NO_x-limited at any particular point in space and time. A performance evaluation which includes comparisons between modeled and observed ratios of indicator species (e.g., O₃/NO_y, O₃/HNO₃) can help reveal whether the model is correctly predicting the sensitivity of ozone to VOC and/or NO_x controls (Sillman, 1995 and 1998) and (Sillman, 1997 and 2002). If a model accurately predicts observed ratios of indicator species, then one can conclude with additional confidence that the predicted change in ozone 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. A second precaution is that application of this method often requires more measurements than are commonly made. 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 to changes in precursors.

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 NO_x or VOC-limited (Blanchard et al., 1999). Additionally, it has been postulated that differences in weekend-weekday ozone patterns may also provide real-world information on which precursors are most responsible for ozone formation in any given area (Heuss et al., 2003). In areas where there are large differences between average weekend and weekday ambient ozone concentrations over the span of several seasons, it would be useful to compare statistical model

performance for weekends versus weekdays. 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. This technique is not recommended if: 1) the number of days modeled is too few to result in an appropriate sample size of days, 2) there is no clear difference between ozone observations on the weekend versus weekdays, and/or 3) it is not possible to attribute differences in weekend/weekday differences to emissions differences. Despite these reservations associated with all of the various observational modeling approaches, these techniques allow one to evaluate the ability of the model to accurately predict changes in ozone concentrations. States/Tribes should include these comparisons in their efforts to evaluate model performance, whenever feasible.

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 ozone at specified sites (Zhang et al., 2003). Various techniques have been implemented into various air quality models, but three of the most commonly used probing tools are photochemical source apportionment (Environ, 2004), the direct decoupled method (DDM) (Dunker, 1980 and 1981), (Environ, 2004) and process analysis (Jeffries, 1994 and 1997); (Jeffries, 1996); (Jang, 1995); (Lo, 1997). 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 at a given location. This can provide valuable insight into whether the model is adequately representing the conceptual description of ozone patterns in the nonattainment area. In the cases where model performance is subpar, these analyses can be useful for indicating where model input or model algorithm improvements are most needed.

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 concentrations to the model-predicted trend over the same time period. As part of this analysis the model is run for current episodes 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. In Section 5, we noted that a retrospective analysis can be a useful tool for diagnosing why an area has not attained the NAAQS. To that end, it is recommended that States/Tribes archive all modeling files and document assumptions and procedures used for calculating model inputs in order to facilitate replications of the modeled analyses at future dates.

Alternative Base Cases: 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. These types of 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.

15.4 How Should the Results of the Model Evaluation be Assessed?

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.⁴⁵

With the additional experience of another 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 15.2 and 15.3 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 4) 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.

⁴⁵ 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.

REFERENCES

- Abt Associates, (2003), "Environmental Benefits Mapping and Analysis Program (BenMAP), User's Manual: <http://www.epa.gov/ttn/ecas/benmodels.html>.
- 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>.
- 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.crao.com/reports/recentstudies00-02/A-37-1%20Final%20Report.pdf>
- Arunachalam, S., Z. Adelman, R. Mathur, and J. M. Vukovich, (2002), "11th Annual Emissions Inventory Conference of the U.S. EPA", <http://www.epa.gov/ttn/chief/conference/ei11/index.html>
- Arunachalam, S., A. Holland, and M. Abraczinskas, (2004), Effects of Grid Resolution on Modeled Attainment Demonstration, In Proceedings of the Annual Conference of the Center for Community Modeling and Analysis System (CMAS) at Chapel Hill, NC.
- 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/Ozone Modeling 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., Masurement-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.
- 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.
- 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).
- 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 NO_x Control Strategies—Scientific and Public Health Findings and Their Regulatory Implications”; EM, July 2003, 27-35.
- 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).
- 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.
- 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.
- Emery, C., and Edward 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
- ENVIRON International Corporation, (2004), "User's Guide to the Comprehensive Air Quality Model with Extensions (CAMx) version 4.10s", Novato, CA, August 2004.
- Environ, (2005). <http://www.camx.com/overview.html>
- Fiore, A.M., D.J. Jacob, H. Liu, R.M. Yantosca, T.D. Fairlie, and Q. Li, "Variability in surface ozone background over the United States: Implications for air quality policy", *J. Geophys. Res.*, 108, 4787, 2003.

- 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.
- 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.
- 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.
- Haney, J.L., and S.G. Douglas, (1996), "Analysis of the Effects of Grid Resolution on UAM-V Simulation Results for the Atlanta Nonattainment Area", Systems Applications International Technical Memorandum SYSAPP-96/68, prepared for Southeast Modeling Center participants, October 21, 1996.
- Henry, R.C., C.W. Lewis, and J.F. Collins, (1994), "Vehicle-related Hydrocarbon Source Compositions from Ambient Data: The GRACE/SAFER Method", *Environmental Science and Technology*, **28**, 823-832.
- Henry, R.C., (1997a), "Receptor Model Applied to Patterns in Space (RMAPS) Part I. Model Description", *J. Air & Waste Mgmt. Assoc.*, **47**, p.216.
- Henry, R.C., (1997b), "Receptor Model Applied to Patterns in Space (RMAPS) Part II. Apportionment of Airborne Particulate Sulfur from Project MOHAVE", *J. Air & Waste Mgmt. Assoc.*, **47**, p.220.
- Henry, R.C., (1997c), "Receptor Modeling Applied to Patterns in Space (RMAPS) Part III. Apportionment of Airborne Particulate Sulfur in Western Washington State", *J. Air & Waste Mgmt. Assoc.*, **47**, p.226.
- 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.

- 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."
- Irwin, J., (2005), "Examination of Model Predictions at Different Horizontal Grid Resolutions", *Environmental Fluid Mechanics*, In Press.
- Jacob, D.J., J. A. Logan, and P.P. Murti, "Effect of Rising Asian Emissions on Surface Ozone in the United States", *Geophys. Res. Lett.*, 26, 2175-2178, 1999.
- Jaffe D., McKendry I., Anderson T., and Price H. "Six 'new' episodes of trans-Pacific transport of air pollutants", *Atmos. Envir.* 37, 391-404, 2003.
- 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, Jennifer M., C. Hogrefe, R. Henry, J. Ku, and G. Sistla, (2005), "An Assessment of the Sensitivity and Reliability of the Relative Reduction 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. Baldrige, B. Eder and B. Cox, (2003), "Spatial-Temporal characterization of Tropospheric Ozone Across the Eastern United States", submitted for publication in *Atmospheric Environment*, **38:26**, 4357-4369, www.sciencedirect.com
- 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.
- 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. Baldrige, 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.
- Morris, R.E., G.M. Wilson, S.B. Shepard, and K. Lee, (1997), "Ozone Source Apportionment Modeling Using the July 1991 OTAG Episode for the Northeast Corridor and Lake Michigan Regions", (DRAFT REPORT), prepared for Mr. Dan Weiss, Cinergy Corporation, Plainfield, IN, NOAA, (1999), <http://www.arl.noaa.gov>.
- 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>.
- 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)).
- Reynolds, S.D., H.M. Michaels, P.M. Roth, T.W. Tesche, D. McNally, L. Gardner and G. Yarwood, (1997), "Alternative Base Cases in Photochemical Modeling: Their Construction, Role and Value", *Atmospheric Environment*, In press.
- Russell, A., and R. Dennis, (2000), "NARSTO critical review of photochemical models and modeling", *Atmospheric Environment*, **34**, 2283-2324.
- 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.
- Seaman, N.L., et al., (1995), "A Multi-Scale Four-Dimensional Data Assimilation System Applied to the San Joaquin Valley During SARMAP. Part I: Modeling Design and Basic Performance Characteristics", *J. Applied Meteorology*, **34**, 1739-1761.
- Seaman, N.L., and D.R. Stauffer, (1996), "SARMAP Meteorological Model. Final Report to San Joaquin Valleywide Air Pollution Study Agency", Technical Support Division, California Air Resources Board, Sacramento, CA., 174 pp.
- Seaman, N.L., et al., (1996b), "Application of the MM5-FDDA Meteorological Model to the Southern California SCAQS-1997 Domain: Preliminary Test Using the SCAQS August 1987 Case", *Ninth Joint Conference on Applications of Air Pollution Meteorology*, American Meteorological Society, Atlanta, GA., January 28-February 2, 1996.

- Seaman, N.L., (1997), "Use of Model-Generated 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.
- Sillman, S., (1995), "The Use of NO_y, H₂O₂, and HNO₃ as Indicators for O₃-NO_x-ROG Sensitivity in Urban Locations", *J. Geophys. Res.* **100**, 14,175-14,188.
- Sillman, S., D. He, C. Cardelino, and R.E. Imhoff, (1997a), "The Use of Photochemical Indicators to Evaluate Ozone-NO_x-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 NO_x-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 Reduction Factors in the Demonstration of Attainment of the 8-Hour Ozone National Ambient Air Quality Standard", *JAWMA*, 54 (8), 950-959.
- 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.

- 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/>.
- Tesche, T.W., and D.E. McNally, (1993a), "Operational Evaluation of the SARMAP Meteorological Model (MM5) for Episode 1: 3-6 August 1990", prepared for Valley Air Pollution Study Agency by Alpine Geophysics, Crested Butte, CO.
- Tesche, T.W., and D.E. McNally, (1993b), "Operational Evaluation of the SARMAP Meteorological Model (MM5) for Episode 2: 27-29 July 1990", prepared for the Valley Air Pollution Study Agency by Alpine Geophysics, Crested Butte, CO.
- Tesche, T.W., and D.E. McNally, (1993c), "Operational Evaluation of the CAL-RAMS Meteorological Model for LMOS Episode 1: 26-28 June, 1991", prepared for the Lake Michigan Air Directors Consortium by Alpine Geophysics, Crested Butte, CO.
- Tesche, T.W., and D.E. McNally, (1993d), "Operational Evaluation of the CAL-RAMS Meteorological Model for LMOS Episode 2: 17-19 July, 1991", prepared for the Lake Michigan Air Directors Consortium by Alpine Geophysics, Crested Butte, CO.
- Tesche, T.W., and D.E. McNally, (1993e), "Operational Evaluation of the CAL-RAMS Meteorological Model for LMOS Episode 3: 25-26 August, 1991", prepared for the Lake Michigan Air Directors Consortium by Alpine Geophysics, Crested Butte, CO.
- Tesche, T.W., and D.E. McNally, (1993f), "Operational Evaluation of the CAL-RAMS Meteorological Model for LMOS Episode 4: 20-21 June, 1991", prepared for the Lake Michigan Air Directors Consortium by Alpine Geophysics, Crested Butte, CO.
- 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)

- Timin, B., (2005a), “An Analysis of Recent Ambient Design Value Data (1993-2004) at 471 Ozone Monitoring Sites”, <http://www.epa.gov/ttn/scram/reportsindex.htm>.
- Timin, B., and W. Cox (2005b), “Analysis of the Variability of Mean Relative Reduction Factors”, <http://www.epa.gov/ttn/scram/reportsindex.htm>.
- 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, (1996a), “Photochemical Assessment Monitoring Stations - 1996 Data Analysis Results Report”, EPA-454/R-96-006.
- U.S. EPA, (1996b), “Review of the national ambient air quality standards for ozone: assessment of scientific and technical information”, OAQPS staff paper, Research Triangle Park, NC: Office of Air Quality Planning and Standards, EPA report no. EPA-452/R-96-007. Available from NTIS, Springfield, VA; PB96-203435.
- U.S. EPA, (1996c), “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, (1997), “EIIIP Volume I, Introduction and Use of EIIIP 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, (1998c), "Implementation Plan PM_{2.5} Monitoring Program", <http://www.epa.gov/ttn/amtic/files/ambient/pm25/pmplan3.pdf>.
- U.S. EPA, (1998d), "National Air Pollutant Emission Trends, Procedures Document 1900-1996", EPA-454/R-98-008, http://www.epa.gov/ttn/chief/ei_datahtml#ETDP.
- 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), "Guidance for Demonstrating Attainment of Air Quality Goals for PM_{2.5} and Regional Haze", http://www.epa.gov/ttn/scram/guidance_sip.htm, Modeling Guidance, DRAFT-PM
- U.S. EPA, (2002), "Guidance for Quality Assurance Project Plans for Modeling (EPA QA/G-5M), EPA/240/R-02/002, December, 2002, Office of Environmental Information, Washington DC.
- U.S. EPA, (2003), 40CFR, Part 51, Appendix W, Revision to the Guideline on Air Quality Models, 68 FR 18440, April 15, 2003.
- U.S. EPA, (2004a), Emission Inventory Improvement Program (EIIP), series of technical documents, <http://www.epa.gov/ttn/chief/eiip>.
- U.S. EPA, (2004b), "Developing Spatially Interpolated Surfaces and Estimating Uncertainty", EPA-454-R-04-0004, <http://hill.nccr.epa.gov/oar/oaqps/pm25/docs/dsisurfaces.pdf>.
- U.S. EPA, (2004c), "The Ozone Report: Measuring Progress Through 2003." EPA 454/k-04-001. <http://www.epa.gov/airtrends/ozone.html>
- 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), "Evaluating Ozone Control Programs in the Eastern United States: Focus on the NO_x Budget Trading Program, 2004", <http://www.epa.gov/airtrends/2005/ozonenbp.pdf>.

- U.S. EPA, (2005c), "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_aug2005.pdf.
- 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.
- 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*, In press.
- 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.

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 additional screening 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 8-hour daily maximum ozone concentrations averaged over several days and multiplies this ratio times the site-specific monitored design value at each monitoring location. If the product is < 84 ppb near all monitoring sites, the test is passed.

Modeling system - This is a group of models used to predict ambient ozone 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 reduction factor (RRF) - The ratio of predicted 8-hour daily maximum ozone averaged over multiple days near a monitoring site with future emissions to corresponding predictions obtained with current emissions.

Unmonitored Area Analysis - An analysis used to ensure that a proposed control strategy will be effective in reducing ozone 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)

$$OBS = \frac{1}{N} \sum_{i=1}^N Obs_{x,t}^i$$

Mean Prediction: The time-average mean predicted value (in ppb) paired in time and space with the observations.

$$PRED = \frac{1}{N} \sum_{i=1}^N Pred_{x,t}^i$$

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.

$$RATIO = \frac{1}{N} \sum_{i=1}^N \frac{Pred_{x,t}^i}{Obs_{x,t}^i}$$

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.

$$BIAS = \frac{1}{N} \sum_{i=1}^N (Pred_{x,t}^i - Obs_{x,t}^i)$$

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.

$$NMB = \frac{\sum_{i=1}^N (Pred_{x,t}^i - Obs_{x,t}^i)}{\sum_{i=1}^N (Obs_{x,t}^i)} * 1$$

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.

$$FBIAS = \frac{2}{N} \sum_{i=1}^N \frac{(Pred_{x,t}^i - Obs_{x,t}^i)}{(Pred_{x,t}^i + Obs_{x,t}^i)}$$

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.

$$ERR = \frac{1}{N} \sum_{i=1}^N |Pred_{x,t}^i - Obs_{x,t}^i|$$

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.

$$NME = \frac{\sum_{i=1}^N |Pred_{x,t}^i - Obs_{x,t}^i|}{\sum_{i=1}^N (Obs_{x,t}^i)} * 100$$

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.

$$FERROR = \frac{2}{N} \sum_{i=1}^N \frac{|Pred_{x,t}^i - Obs_{x,t}^i|}{Pred_{x,t}^i + Obs_{x,t}^i}$$

Correlation Coefficient (R²): 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.

$$CORRCOEFF = \frac{\sum_{i=1}^N (Pred_i - \bar{P})}{\sqrt{\sum_{i=1}^N (Pred_i - \bar{P})^2}}$$

This page is intentionally left blank.

United States
Environmental Protection
Agency

Office of Air Quality Planning and Standards
Emissions, Monitoring and Analysis Division
Research Triangle Park, NC

Publication No. EPA-454/R-05-002
October 2005
