USE OF MODELS AND OTHER ANALYSES

IN ATTAINMENT DEMONSTRATIONS FOR THE

8-HOUR OZONE NAAQS

Draft

October 5, 1998

U.S. EPA, OAQPS, EMAD, AQMG
Research Triangle Park, NC 27711

DISCLAIMER

This document presents recommendations for using air quality models and other analyses to demonstrate attainment of national ambient air quality standards for ozone which were promulgated in 1997. The content of this document is subject to further review within the U.S. Environmental Protection Agency and by external reviewers, and thus is subject to change. Any mention of trade names or commercial products is not intended to constitute endorsement or recommendation for use.

Table of Contents

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

List Of Figures

List Of Tables

Executive Summary

This document identifies how to determine if results of an air quality model and other analyses "demonstrate attainment" of the 8-hour national ambient air quality standard (NAAQS) for ozone. The document also describes how to apply air quality models to *obtain* the results later used to see if attainment is demonstrated. Guidance in this document applies to all "traditional" nonattainment areas. This includes locations that violate the 8-hour ozone NAAQS and may or may not violate the 1-hour NAAQS.

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

ES 1.0. What Is An Attainment Demonstration?

An 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 measures which will result in the required emissions reductions. This guidance describes how to use air quality models and other analyses to determine if results of a simulated control strategy indicate attainment would be likely. Determining necessary emission reductions may be done by relying exclusively on results obtained with air quality simulation models. These include the outcome of a modeled attainment test plus a screening test to estimate whether a proposed emission reduction suffices to meet the NAAQS. Other analyses, including trends analyses, observational models, etc. may be used to supplement the modeled attainment and screening tests

A modeled attainment test compares ozone concentration predictions with the form of the ozone NAAQS. The ozone NAAQS is met if the 4th highest 8-hour daily maximum ozone concentration, averaged over 3 consecutive years, is $\leq$ 0.08 ppm. The average monitored 4th highest 8-hour daily maximum concentration is called the "design value" for ozone. The modeled attainment test is passed if predicted design values in the vicinity of all monitoring sites are $\leq$ 84 ppb.

Provided the modeled attainment test and a supplementary screening test are passed or close to being passed, States may use a broader set of analyses to estimate if attainment is likely. This is called a "weight of evidence determination". A weight of evidence determination includes results obtained with air quality simulation models plus conclusions drawn from analyzing monitored air quality data, emissions estimates and meteorological data. Results of each analysis are considered in concert to determine whether or not attainment is likely.

ES 2.0. What Is The Recommended Modeled Attainment Test?

The recommended modeled attainment test uses monitored design values in concert with model-generated data. The test uses model results in a "relative" rather than "absolute" sense. For a given modeled day, the test considers the ratio of the model's predicted future to predicted current 8-hour daily maximum concentration near a monitor. The ratio is called the "relative reduction factor", or RRF. The test consists of 4 steps.

1. Compute a current site-specific design value from *monitored* data.

2. Using modeled data, estimate a relative reduction factor near the monitor for each modeled day. Calculate the arithmetic mean relative reduction factor for the modeled days to get a typical response to the simulated control strategy.

3. Multiply the mean relative reduction factor obtained in step 2 times the site-specific design value in step 1. The result is a predicted site-specific future design value. If this value is ≤ 84 ppb, the test is passed near the monitor site being evaluated.

4. Repeat steps 1 through 3 for each monitoring site where the current monitored design value approaches or exceeds 84 ppb. If predicted future site-specific design values are ≤ 84 ppb at each site, the test is passed.

The modeled attainment test does not address future air quality at locations where there is no nearby ozone monitor. If the air quality model *consistently* predicts 8-hour daily maximum ozone concentrations at a particular, unmonitored location which are substantially higher than any predicted near a monitoring site, this could be cause for concern. Thus, a State should perform an additional screening analysis under these circumstances. Attainment can be demonstrated by passing the modeled attainment test, as well as a screening test applied at any grid location with consistently higher predictions than any near a monitoring site. The screening test is to multiply the area wide monitored design value (i.e., the highest of the site-specific design values) times the mean relative reduction factor(s) predicted at the suspect receptor location(s). If the resulting estimated future design value(s) is ≤ 84 ppb, the outcome is not inconsistent with attainment of the NAAQS. The modeled attainment test and screening analysis are discussed in Section 3.0.

ES 3.0. What Is A Weight Of Evidence Determination?

A weight of evidence determination is a diverse set of technical analyses performed to corroborate findings of the modeled attainment and screening tests and to help assess adequacy of a proposed strategy when the outcome of one or both modeled tests is ambiguous. If a weight of evidence analysis is used, a State should consider a recommended core set of analyses consisting of (1) a set of air quality simulation model results which includes the previously described tests plus spatial analyses of estimated concentrations, (2) an analysis of observed air quality and estimated emissions trends, and (3) an analysis of outcomes produced by observational models.

We identify factors which enhance credibility of evidence produced by each of the core analyses, as well as outcomes which would be consistent with the likelihood that a strategy demonstrates attainment. This is illustrated in Table ES.1 for the 3 recommended core analyses.

Table ES.1. Recommended Core Analyses For A Weight Of Evidence Determination

(1) Type of Analysis	(2) Factors Increasing Credibility of the Analysis	(3) Outcomes Consistent with Hypothesis That a Candidate Strategy will Lead to Attainment
Air Quality Simulation Models	-good model performance -extensive observational data base available -short projection periods -carefully QA'd inventory -confidence in meteorological inputs -good ability to pose and address questions about a strategy's adequacy -other analyses tend to corroborate conclusions -selected episode days have observations near the design value	-the modeled attainment test is passed, or nearly passed -screened estimates for future design values at sites w/o monitors are ≤ 84 ppb -commitment is made to deploy monitors at locations not passing the screening analysis -substantial modeled improvement in air quality is predicted using several measures described in Section 4.1.1. -similar conclusions are reached with other peer reviewed models

Analysis of Air Quality and Emissions Trends	-extensive monitoring network exists -both ozone and precursor trends are available -statistical model used to normalize trend for meteorological differences explains much variance -short projection periods used in the analysis -a pronounced, statistically significant downward trend is apparent -similar conclusions are reached using several trend parameters -continued, comparable relative reductions in emissions are provided for	-a pronounced downward normalized trend exists in the site-specific design value at all sites with design values greater than 84 ppb. -Using projected emissions to extrapolate the air quality trend line to the required attainment date indicates an 8-hour daily maximum concentration $\leq$ 84 ppb. -Other observed air quality trend parameters also show a substantial decline.
Use of Observational Models	-an extensive monitoring network exists -precursor and indicator species are measured using instruments with appropriate sensitivity -monitoring sites appear spatially representative -data have been QA'd, and results are self-consistent -plausible physical explanations exist for findings	-Findings indicate sources controlled in the candidate strategy are important causes of observed high ozone -Analysis of indicator species suggests the direction of the strategy (e.g., emphasis on VOC or NOx) is appropriate.

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

(1) a State should discuss why the proposed analysis is relevant for assessing likely attainment,
(2) a State should identify the procedure to be used and the data base available to support it, and
(3) a State should identify (in advance) outcomes which would be consistent with a hypothesis that a proposed strategy demonstrates attainment.

Weight of evidence and its use is discussed in Section 4.0.

ES 4.0. Why Do We Recommend This Modeled Attainment Test And Offer An Option To Perform A Weight Of Evidence Determination?

Test results can be readily related to the form of the NAAQS. The NAAQS is concerned with the 4th highest 8-hour daily maximum concentration averaged over 3 consecutive years. It is difficult to tell whether a modeled exceedance in a particular episode is or is not inconsistent with meeting the NAAQS. Thus, using a model by itself to rigorously assess whether the NAAQS is met would require modeling a substantial number of days in three different years. Further, we believe relatively resource-intensive models are needed to simulate effects of reducing precursor emissions on ozone. Thus, the test uses *observed* design values to "anchor" model predictions to the form of the NAAQS. Design values are, by definition, calculated consistently with the form of the NAAQS, and their use allows a State to apply a resource-intensive model to see how they might be changed by a control strategy.

Using models in a "relative" rather than "absolute" sense reduces uncertainty in the predictions. Problems in interpreting model results posed by uncertainty in the predictions may be greater for addressing the 8-hour NAAQS than was true for the 1-hour standard. The 8-hour NAAQS is closer to continental background values. Further, design values tend to be closer to the specified level of the NAAQS than is true for the 1-hour standard. As a result, the signal (i.e., the change we wish to effect in the design value) to noise (i.e., uncertainty in predictions) ratio is likely to be smaller than heretofore. The recommended test reduces uncertainty (i.e., "noise") in two important ways. First, monitored data (i.e., current design values) are incorporated directly into the test. These data are likely measured with greater accuracy than an absolute model prediction, and precision of the measurements is known. Second, the ratio of future to current modeled ozone concentrations likely has less associated uncertainty than an absolute model prediction. This follows because for each model application some of the sources of uncertainty (e.g, meteorological inputs) remain unchanged. Taking the ratios of two estimates with similar sources of error should lead to smaller uncertainty since some of the error in the numerator and denominator cancels.

The attainment demonstration no longer relies on a single outcome or results on a single day. The modeled attainment test is accompanied by a screening test. Further, if the outcome of these tests is doubtful, a weight of evidence determination may be used to see whether other model outputs and other types of analyses provide corroborative evidence for conclusions drawn from the modeled attainment and screening tests. The outcome of the tests themselves is based on a composite set of calculations from several modeled days rather than a single day. This reduces the risk of choosing an inappropriate strategy on the basis of a single outcome which is subject to uncertainty.

ES 5.0. What Documentation Is Needed To Support An Attainment Demonstration?

A State should address 9 subject areas in its documentation. These are enumerated in Table ES.2. Documentation should be accompanied by an executive summary which addresses each of the 9 areas shown in the table. Documentation requirements are addressed in Section 6.0.

Table ES.2. Recommended Documentation For Demonstrating Attainment Of The 8-hour Ozone NAAQS

Subject Area	Purpose of Documentation	Issues Included
Modeling/Analysis Protocol	Communicate scope of the analysis and document stakeholder involvement	Names of stakeholders participating in preparing and implementing the protocol; Types of analyses performed; Steps followed in each type of analyses; Days and domain considered.
Emissions Preparations and Results	Assurance of valid, consistent emissions data base. Appropriate procedures are used to derive day-specific emission estimates needed for air quality simulation 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.
Air Quality/Meteorology Preparations and Results	Assurance that representative air quality and meteorological inputs are used in analyses	Extent of data base and procedures used to derive & QA inputs for analyses used in the weight of evidence determination; Departures from guidance and their underlying rationale.

Performance Evaluation for Air Quality Simulation Model (and Other Analyses)	Show decision makers and the public how well the model (or other analyses) reproduced observations or otherwise performed on the days selected for analysis	Summary of observational data base available for comparison; Identification of performance tests used and their results; Ability to reproduce observed temporal and spatial patterns; Overall assessment of what the performance evaluation implies.
Diagnostic Tests	Ensure rationale used to adjust model inputs or to discount certain results is physically justified and the remaining results make sense.	Results from application prior to adjustments; Consistency with scientific understanding and expectations; Tests performed, changes made and accompanying justification; Short summary of final predictions.
Description of the Strategy Demonstrating Attainment	Provide the EPA and the public an overview of the plan selected in the attainment demonstration.	Qualitative description of the attainment strategy; Reductions in VOC, NOx, and/or CO emissions from each major source category for each State/county from current (identify) base case emission levels; CAA mandated reductions and other reductions; Identification of authority for implementing emission reductions in the attainment strategy. Show predicted 8-hr site-specific future design values for the selected control scenario and identify any location which fails the screening test described in Section 3.4;

Data Access	Enable the EPA or other interested parties to replicate model performance and attainment simulation results, as well as results obtained with other analyses.	Assurance that data files are archived and that provision has been made to maintain them; Technical procedures for accessing input and output files; Computer on which files were generated and can be read; Identification of contact person, means for downloading files and administrative procedures which need to be satisfied to access the files.
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 within the required time.	Description of the modeled attainment test and observational data base used; Identification of air quality simulation model used; Identification of other analyses performed; Outcome of each analysis, including the modeled attainment test; Assessment of the credibility associated with each type of analysis in this application; Narrative describing process used to conclude the overall weight of available evidence supports a hypothesis that the selected strategy is adequate to attain the NAAQS.
Review Procedures Used	Provide assurance to the EPA and the public that analyses performed in the attainment demonstration reflect sound practice	Scope of technical review performed by those implementing the protocol; Assurance that methods used for analysis were peer reviewed by outside experts; Conclusions reached in the reviews and the response thereto.

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

Part II of the guidance describes how to apply air quality models to obtain the results used in an attainment demonstration. The procedure we recommend has 7 steps.

1. Develop a modeling/analysis protocol.
2. Select an appropriate model to support the demonstration.
3. Select appropriate meteorological episodes to model.
4. Generate emissions inputs to the air quality model.
5. Choose an appropriate area to model with appropriate horizontal/vertical resolution.
6. Generate meteorological and air quality inputs to the air quality model.
7. Evaluate performance of the air quality model and perform diagnostic tests to obtain further insights.

Many of these steps require substantial effort. States should work closely with the appropriate U.S. EPA Regional Office(s) in executing each step. This will increase the likelihood of approval of the demonstration at the end of the process.

ES 6.0. What Does A Modeling/Analysis Protocol Do And What Should It Contain?

A modeling/analysis protocol is a document which identifies methods and procedures to be used in the analyses. The protocol also identifies ground rules to be followed in undertaking analyses to estimate emission reductions needed to meet the NAAQS. Ground rules include a description of how affected stakeholders in the modeling/analysis process will be encouraged to participate, the process by which decisions will be made, means used for communicating issues and decisions, and the methods, data bases and procedures to be used to obtain results. As the name implies, the protocol should address use of other analyses as well as air quality modeling. The document is usually prepared by the State/local agency(is) having lead responsibility for the modeling/analysis, in consultation with stakeholders. Specific topics which should be included in the protocol are identified in Section 9.0.

ES 7.0. What Should I Consider In Choosing An Air Quality Model?

We do not identify a "preferred" or "Guideline" urban scale or nested regional model to use in attainment demonstrations for ozone. The U.S. EPA has spent much effort to develop a user friendly, yet scientifically rigorous, model and modeling system. This is called CMAQ/MODELS3. We believe the CMAQ model and the MODELS3 modeling system will be able to address ozone and integrate ozone-related simulations with simulations related to meeting goals for $PM_{2.5}$ and regional haze. The U.S. EPA will provide support for CMAQ/MODELS3 by furnishing codes, user's manuals, training and limited troubleshooting. However, at the present time, CMAQ/MODELS3 has not been shown to perform better than alternatives, nor has it been suitably demonstrated. There are several models for ozone which are available, have been applied in several locations and have performed in an acceptable manner. Therefore, we do not believe that it is possible at this time to identify a "preferred" model. To qualify for use in an attainment demonstration, a model should meet requirements for "alternative models, which are identified in 40CFR Part 51, Appendix W (i.e., the *Model Guideline*). These requirements are listed below.

1. The model has received a scientific peer review.

2. The model can be demonstrated to be applicable to the problem on a theoretical basis.

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

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

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

6. The model is available to users at a reasonable cost and is not proprietary.

To select a model for a particular application, States should first determine what attributes are needed for a qualifying model to address the nonattainment area's ozone problem, and then choose among models possessing these attributes. Five factors should be considered in selecting an air quality model for a specific application. A State may use these factors to demonstrate that use of an "alternative" model is preferable for a specific application. Choice of a model should be concurred with by the appropriate U.S. EPA Regional Office and U.S. EPA Model Clearinghouse. The five factors are listed in approximate order of importance.

1. Nature of the air quality problem leading to nonattainment of the ozone NAAQS should first be assessed, and the selected model should have attributes and capabilities consistent with the perceived nature of the problem.

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

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

4. Time and resource constraints may be considered.

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

Prior to using model results in a specific attainment demonstration, a State should show that the model performs adequately in replicating base case observations available for that demonstration.

Further discussion of model selection is found in Section 10.0.

ES 8.0. How Do I Choose Meteorological Episodes?

Because we do not want modeled results to be dependent on the outcome of any single modeled day, we recommend that at least 10 primary episode days (i.e., excluding model "ramp-up" days) be modeled. Three primary criteria should be considered in choosing episodes to model for supporting ozone attainment demonstrations.

1. Choose episodes containing days with observed 8-hour daily maximum ozone concentrations close to (± 10 ppb) site-specific design values monitored during the 3-year period straddling the episode.

2. Choose a mix of episodes which represents a variety of meteorological conditions observed to correspond to monitored ozone close to site-specific design values.

3. Choose episodes containing days with intensive data bases.

There may sometimes be conflicts among the three primary criteria. Several additional factors may also be considered to help choose episodes. These are: (1) give preference to previously modeled episodes, (2) choose episodes drawn from the period used to calculate the current design value (e.g., 1997-99), and (3) for regional applications, choose episodes which are of interest in as many nonattainment areas as possible. Episode selection is discussed in Section 11.0.

ES 9.0. How Do I Produce Needed Emissions Inputs?

Producing needed emissions inputs requires several steps. First, compile Statewide and then countywide estimates for VOC, NOx and CO emissions for point, area, mobile and biogenic emissions. Second, quality assure the outputs. Third, convert the resulting estimates into speciated, gridded hourly emissions using emission models. Fourth, once again, quality assure the results. Finally, project gridded, speciated hourly emission estimates to a future year which corresponds to when the NAAQS must be met.

The U.S. EPA has prepared a series of guidelines relating to these steps as a part of the Emission Inventory Improvement Program (EIIP). States should be familiar with these guidelines. The National Emissions Trends Inventory (NET) compiled for 1999 is the preferred source of information for Statewide estimates in portions of the modeling domain for which States who are stakeholders have no better information.

Different means are used to obtain emissions information in the form needed by air quality models. Ideally, location and daily/weekly emission patterns should be directly available for point sources. Emission models are needed to characterize emissions from other point sources, stationary area, mobile and biogenic sources. We identify default choices for modeling each of these components. Defaults are EMS95 (stationary sources), MOBILE5A (mobile source emission factors outside of California) and BEIS2 (biogenics). Alternative models may be approved if they have received a scientific peer review.

Quality assurance of emission estimates is necessary for an attainment demonstration to be credible. We recommend that it be performed during several stages of the process needed to derive required emission inputs to an air quality model. Use of a Delphi approach, computer graphics and comparisons with available, speciated air quality data are useful means for quality assuring emission estimates. Generating emission inputs is addressed more fully in Section 12.0.

ES 10.0. How Do I Select A Modeling Domain And Its Horizontal/Vertical Resolution?

States should review available air quality, meteorological and emissions data to help select a domain size which is consistent with a nonattainment area's problem but which is not unnecessarily resource intensive. We suggest a procedure for comparing regional (upwind) observations with local design values which may be useful in choosing between regional and urban scale domains. A typical urban domain may be about 300 km on a side. A typical regional domain is likely to exceed 1000 km on a side.

Choice of horizontal/vertical resolution presents a conflict between resources and data base management vs. scientific rigor. Sensitivity of results to grid resolution is likely to vary on a case by case basis. For urban scale analyses and for the fine portion of a nested regional application, we recommend horizontal resolution of 12 km or finer. Coarse portions of regional applications may use 36-km grid cells. States may initially assume a vertical resolution consisting of 5-7 layers throughout the planetary boundary layer. These default assumptions may be used to initially identify an emission control strategy which appears to demonstrate attainment. However, if topography or other factors appear to warrant it, a State should perform a diagnostic test with finer resolution to confirm that the strategy is likely to demonstrate attainment.

More detailed suggestions for selecting domain size and grid resolution are contained in Section 13.0.

ES 11.0. How Do I Produce Meteorological And Air Quality Inputs Required By An Air Quality Model?

We recommend that States use a dynamic meteorological model with four dimensional data assimilation (FDDA) as the principal means for generating meteorological inputs required by air quality models used in ozone attainment demonstrations. Any such meteorological model which has received a scientific peer review may be used. As with the output from emissions models, it is critical that results of meteorological models be quality assured. We identify four potentially useful means for doing so: (1) compare results for the same day obtained with different dynamic models, (2) use computer graphics to identify wind vector patterns and compare with available observations, (3) compare spatial air quality patterns obtained with an air quality model using the input with observed patterns and with past model applications using other means for generating meteorological inputs, and (4) when applicable, compare calculated dimensionless parameters with ranges believed to be appropriate.

Applying meteorological models over extensive domains with fine scales (i.e., < 12 km) can be very resource intensive and present data base management problems. We suggest means for reducing such problems. These and other issues related to meteorological models are discussed in Section 14.0.

Air quality inputs are needed for initial conditions and for boundary values at the edges of a modeling domain. There is no satisfactory way to use available air quality observations to specify initial conditions. Thus, States should diminish their importance by beginning a simulation one or more days prior to the period of interest for urban applications and two or more days earlier for regional applications. Nested regional models are the usual preferred means for generating boundary conditions to a portion of the large regional domain which is the focus of an attainment demonstration. If an urban scale model is used, the domain should be large enough so that emissions occurring in the center of the domain at 8 am remain within the domain at least until 8 pm of the same day.

ES 12.0. How Do I Evaluate Model Performance And Make Use Of Diagnostic Analyses?

In Section 15.0, we describe 6 means for evaluating model performance.

1. Use computer-generated graphics.

2. Use “ozone metrics” in statistical comparisons.

3. Compare predicted and observed precursor or species concentrations.

4. Compare observed and predicted ratios of indicator species.

5. Compare predicted source category contribution factors with estimates obtained using observational models.

6. Use retrospective analyses in which air quality differences predicted with models are compared with observed trends.

Most of these approaches are only able to address how well a model replicates a past set of observations. While this is useful, the key question is, “how well is a model likely to forecast *changes* in ozone accompanying changes in precursor emissions?”. The 4th (use of ratios of indicator species) and 6th (use of retrospective analyses) approaches have the potential to address this key question. However, both require additional efforts to make certain measurements or to perform additional analyses. We discuss these additional efforts in Sections 5.0 and 15.0.

All of the identified means for evaluating model performance have strengths and weaknesses. Thus, we recommend that as many of these approaches as feasible be used to evaluate model performance. Assessment of whether or not performance is adequate is most properly done by considering evidence produced by all of these evaluation techniques--much the same way as a weight of evidence approach is used in an attainment demonstration.

Diagnostic tests are useful for at least four purposes: (1) to test the robustness of a control strategy (i.e., do we reach more or less the same conclusions?), (2) as a means for confirming that simplifying assumptions do not alter conclusions regarding a strategy, (3) as a means for prioritizing additional data gathering efforts, and (4) to help identify control strategies warranting further evaluation. In designing and evaluating results of diagnostic tests, States should be aware of how models are to be used to support an attainment demonstration. That is, models should be used in a *relative* sense. Thus, diagnostic tests should consider how relative reduction factors (RRF) or other predicted changes in ozone are affected by various factors.

1.0 Introduction

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 or other analyses support a conclusion that attainment of the national ambient air quality standard (NAAQS) for 8-hour daily maximum ozone concentrations is likely in locations where a State implementation plan (SIP) revision is required. The second purpose is to describe how to apply an air quality simulation model to produce results needed to support a required attainment demonstration. Part I of this document provides guidance for using results to help demonstrate attainment. Part II provides guidance on how to apply models to produce these results.

1.2 Does The Guidance In This Document Apply To Me?

This guidance applies to all locations for which 1997-1999 ozone data reported to the US EPA's AIRS data base show a violation of the 8-hour NAAQS for ozone and which do not qualify or have elected not to be treated as a "transitional" nonattainment area. Qualifications for a "transitional" nonattainment area are defined in a 1997 Presidential Directive (Clinton, 1997). Guidance for performing attainment demonstrations in "transitional" nonattainment areas is presented elsewhere (U.S. EPA, 1998). The term, "traditional nonattainment areas", has been coined to identify areas for which the guidance in this document applies.

SIP revisions designed to correct problems meeting the 8-hour NAAQS in traditional nonattainment areas are due within 3 years after an area is designated "nonattainment" (e.g., July 18, 2003, assuming designation occurs on July 18, 2000) . This means that attainment demonstrations supporting these revisions should be completed by 2002, and work underlying these demonstrations needs to begin no later than 1999.

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 each receptor location, neither now nor in the future. Thus, there will be a distribution of differences between predictions and observations resulting from comparisons on different days at each receptor. Uncertainty arises for a variety of reasons, including limits in the model's formulation which reflect incomplete understanding or a need to make computations tractable, data base limitations and uncertainty in forecasting future determinants of emissions. States should recognize these limitations when preparing their modeled attainment demonstrations, as should those reviewing the demonstrations.

We recommend several qualitative means for recognizing model limitations and resulting uncertainties when preparing an attainment demonstration. First, we recommend using models in a relative sense in concert with observed air quality data (i.e., taking the ratio of future to present predicted air quality and multiplying it times a monitored design value). As described later, we believe this approach should reduce some of the uncertainty attendant with using absolute model predictions directly. Second, we offer the option for States to use several model outputs, as well as other analyses besides models to provide corroborative evidence concerning adequacy of a proposed strategy for meeting the NAAQS. Outcomes of modeling and corroboratory analyses are 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 could undertake, if they so choose, to better apply corroborative approaches and to better understand whether subsequent air quality observations are consistent with the expectations underlying a SIP. A State has the responsibility to prepare a subsequent SIP revision, if the U.S. EPA finds that the SIP is substantially inadequate to achieve the NAAQS.

Premise 2. Resource intensive approaches are often likely to 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 regional reductions in NOx emissions should lead to attainment of the 8-hour NAAQS in much of the eastern U.S., concentrations approaching the .08 ppm level specified in the NAAQS will affect local strategies needed to attain the NAAQS in the remaining nonattainment areas.

If regionality is a problem, 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 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 which can “specialize” in that kind of analysis. Another example might be formation of a regional center of some kind to perform analyses as directed by its client group of States.

Premise 3. There will be a widespread need to use nested regional models. 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. Near nonattainment areas, it is likely that more detailed attention will need to be paid to atmospheric mixing of nearby emissions than is necessary for emissions in locations which are more remote from the nonattainment areas. This is consistent with use of nested regional models.

This guidance identifies a model which the US EPA will support (CMAQ) (U.S. EPA, 1998a). The model is embedded within the MODELS3 system. Support includes documentation, ready access to the code, training, updates and limited troubleshooting. However, we believe it is premature to identify CMAQ or any other nested regional model as the "guideline" model for ozone. States may use CMAQ or an alternate model provided certain criteria, identified in this guidance, are met. These criteria apply equally to CMAQ and alternative model(s). The guidance also provides recommendations for developing meteorological, air quality and emissions inputs for nested regional models, and makes suggestions for quality assuring inputs and evaluating model performance.

Premise 4. Problems posed by high ozone, fine particulates and regional haze share several commonalities. Ozone formation and formation of secondary particulates result from several common reactions and reactants. Often similar sources contribute precursors to both ozone and fine particulate problems. In some regions of the U.S., high regional ozone and secondary particulates have been observed under common sets of meteorological conditions. Reducing fine particulate matter is the principal controllable means for improving regional visibility. U.S. EPA policy is to encourage "integration" of programs to reduce ozone, fine particulates and regional haze to ensure they do not work at cross purposes and to foster maximum total benefit for lower costs.

Integration of strategies to reduce ozone, fine particulates and regional haze is complicated by different required dates for SIP revisions (2003 for ozone, circa 2005-2006 for fine particulates and regional haze). One reason for a subsequent review of a strategy selected to meet the ozone NAAQS is to check its compatibility with plans (whose details become known later) to meet goals for fine particulates and regional haze. This guidance identifies activities which could yield useful information for a subsequent review to help States prepare for such a check if they so choose. We recommend use of models, such as CMAQ, which will also have the capability of later considering effects of control strategies on fine particulates and regional haze.

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 6 sections (i.e., Sections 2.0-7.0). Part II consists of 9 sections (Sections 8.0-16.0).

Section 2.0 contains an overview of the procedure we recommend for using results to help demonstrate attainment of the 8-hour ozone NAAQS. The recommended approach is to first use an air quality simulation model to estimate current and future ozone concentrations. Next, use the predicted relative changes in ozone in concert with measured data to estimate future ozone concentrations. We refer to this exercise as a "modeled attainment test". If the test is passed and a similar screening test, applied at selected locations without monitors, is also passed, these outcomes can be used to conclude that attainment is likely if the simulated control strategy is adopted. If the attainment and screening tests are passed or close to being passed, a suite of model predictions as well as several additional data analyses may be used for corroboration. Corroborative analyses use air quality and emissions data plus additional interpretation of model results. Results of the modeled attainment test and corroboratory analyses are considered together in a "weight of evidence determination" to assess whether or not a proposed control strategy is likely to be successful in meeting the NAAQS. "Weight of evidence" may be used either to require more or to require fewer control measures than the modeled attainment test suggests is necessary.

Section 3.0 describes the recommended modeled attainment test and screening test in detail. The Section includes an example illustrating use of the modeled attainment test we recommend.

Section 4.0 describes how a weight of evidence determination should be performed, if a State chooses to demonstrate attainment using this approach. A weight of evidence determination consists of a series of required corroborative analyses plus additional, optional analyses. Each required analysis is identified, along with conditions which lend credibility to its outcome. Outcomes which would be consistent with concluding that a proposed control strategy will work are also identified. Several examples of optional analyses are also provided, along with recommendations for accompanying documentation.

Section 5.0 identifies several data gathering activities and analyses which States could undertake to enhance corroborative analyses and support subsequent reviews. None of these activities is required for the current ozone SIP revision to be approved. However, they would increase credibility of corroborative analyses. This, in turn, could strengthen their potential to serve as deciding factors when model results are close to passing or close to failing the test. Further, the activities could strengthen subsequent reviews. A subsequent review may be desirable to diagnose why a strategy is or isn't working, or to relate the chosen strategy to others which are later considered to reduce $PM_{2.5}$ or regional haze.

Section 6.0 identifies the necessary documentation describing the analyses used to demonstrate attainment of the ozone NAAQS.

Section 7.0 lists the references cited in Part I and in this introduction (Section 1.0).

Part II ("How should I apply air quality models to produce results needed to help demonstrate attainment?") begins in Section 8.0 with an overview of the topics to be covered.

Section 9.0 identifies the need for a modeling/analysis protocol. We also discuss the protocol's function as well as what subjects should be addressed in the protocol.

Section 10.0 addresses what should be considered in choosing a model to use to support the attainment demonstration of the ozone NAAQS. Several criteria are identified for accepting use of a model for this purpose.

Section 11.0 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 12.0 discusses how to develop appropriate emissions estimates for use in the selected air quality model. Topics include use of available inventory estimates, quality assurance, application of emissions models and estimating projected emissions.

Section 13.0 identifies factors which should be considered in choosing a model domain and horizontal and vertical resolution for the model application.

Section 14.0 addresses how to develop meteorological inputs as well as initial and boundary air quality data for use in a modeling exercise supporting an attainment demonstration. Topics covered include use of dynamic meteorological models, four dimensional data assimilation, relationship to domain size and the need for "ramp-up" days.

Section 15.0 treats the topics of model performance evaluation and use of diagnostic analyses.

The guidance concludes with Section 16.0, which lists references cited in Part II.

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 measures which will result in the required emissions reductions. As noted in Section 1.0, this guidance focuses on the first component of an attainment demonstration--interpretation and conduct of analyses to estimate the amount of emission reduction which "demonstrates attainment". Emission reduction strategies should be simulated by reducing emissions from specific source categories rather than exclusively through broad "across-the-board" emission reductions from all sources.

States may estimate the amount of emission reduction needed to demonstrate attainment by passing a "modeled attainment test" plus passing a screening test at selected locations without an ozone monitor. If the results of the tests are ambiguous, a State may consider a broader set of model results plus perform a set of additional corroboratory analyses to determine whether the "weight of evidence" indicates that 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 simulation model is used to simulate current and future air quality. If future estimates of an ozone design value are $\leq$ 84 ppb, the test is passed. 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 predictions at a number of receptor locations. We call each of these site-specific ratios "relative reduction factors". Future ozone design values are estimated near existing monitoring sites by multiplying a relative reduction factor at a receptor "near" each monitor times the monitor-specific ozone design value. The resulting predicted site-specific "future design value" is compared to 84 ppb. If all such future site-specific design values are $\leq$ 84 ppb, the test is passed.

The modeled attainment test we recommend predicts whether or not all observed future design values will be less than or equal to the concentration level specified in the NAAQS for ozone under meteorological conditions similar to those which have been simulated. By itself, the test makes no statement about future ozone at locations where there is no nearby monitor. Thus, even if the test is passed, we require a supplementary screening analysis to identify other locations where passing the test might be problematic if monitoring data were available. One purpose of the supplementary analysis is to help determine whether a more extensive weight of evidence determination is a viable alternative for demonstrating attainment. Like the test itself, this supplementary screening analysis is described more fully in Section 3.0. Briefly however, it entails the following:

-identification of areas in the modeling domain where "absolute" predicted 8-hour daily maxima are consistently greater than those predicted in the vicinity of any monitor site, and

-computation of relative reduction factors for each identified unmonitored area with high predicted ozone. These factors are then multiplied by the area wide design value to obtain an estimated future design value for each such location.

The outcomes of the modeled attainment and screening tests determine whether using a weight of evidence determination is a viable alternative for demonstrating attainment. If results obtained from one or both of these tests are far removed from passing, we do not believe the more qualitative arguments made in a weight of evidence determination can be convincing. Table 2.2 illustrates the recommended ground rules for determining whether weight of evidence may be invoked to demonstrate attainment.

Table 2.2. Ground Rules For Using Weight Of Evidence In An Attainment Demonstration

Results of Modeled Attainment Test*	May Weight of Evidence Be Used?
Design Value $\leq$ 84 ppb, all sites	**Yes (To check strategy)**
Design Value 85 - 89 ppb at one or more sites	**Yes (To see if revised strategy is needed)**
Design Value $\geq$ 90 ppb, at one or more sites	**No (Need to revise strategy)**

***** Includes calculations at screening sites, if applicable

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

In a weight of evidence determination, States should perform and/or review results from several diverse types of analyses. States should next note whether or not results from each of these analyses support a conclusion that the proposed strategy will meet the air quality goal. States should then weigh each type of analysis according to its credibility as well as its ability to address the question being posed (i.e., is the strategy adequate for meeting the ozone NAAQS by a defined deadline?). Next, conclusions derived in the two preceding steps are combined to make an overall assessment of whether meeting the air quality goal is likely. This last step is a qualitative one involving some subjectivity. As with the option of using modeled attainment and screening tests results alone, 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 should provide a written rationale documenting how and why the final conclusion is reached regarding adequacy of the final selected strategy.

Results obtained with air quality simulation models should ordinarily be weighted heavily in the weight of evidence determination described in the preceding paragraph. This follows from including *ability to address the question being posed* as one of two criteria for weighing results from different analyses. If the modeled attainment and screening test is passed, this supports a hypothesis that the strategy is adequate. This information is included as one of several elements in a weight of evidence determination to assess the strategy's adequacy. The further model results are from passing the modeled attainment or screening test, the more compelling results from other analyses have to be for a control strategy to demonstrate attainment. If either the modeled attainment test or screening test produces one or more estimated site-specific future design values $\geq$ 90 ppb, the use of a weight of evidence analysis is not appropriate. States should consider a revised control strategy.

2.3 Why Do We Recommend Using Models In A "Relative" Sense And Offer An Option To Perform A Weight Of Evidence Determination?

The modeled attainment demonstration we recommend differs from that in past guidance for ozone in two major respects (U.S.EPA, 1996). 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, results can now be used as a basis for requiring emission reductions greater than those implied by the modeled attainment test as well as a rationale for concluding that not all of the modeled emission reductions are necessary. 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 4th highest 8-hour daily maximum ozone concentration, averaged over 3 consecutive years, to be $\leq$ 0.08 ppm at each monitoring site[1]. The feature of the NAAQS requiring averaging over 3 years presents difficulties using the resource-intensive episodic 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 circumvented by using the *monitored* design value, calculated consistently with the form of the NAAQS, as an inherent part of the modeled attainment test.

---

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

2. Design values for the 8-hour NAAQS are likely to be closer to the concentration specified in the NAAQS than is true for the 1-hour NAAQS for ozone. A review of 1994-96 ozone data reported in the U.S. EPA's AIRS data base suggests that most traditional nonattainment areas are likely to have design values that are within about 40 ppb of the concentration specified in the 8-hour NAAQS. Thus, it is likely that the "signal to noise ratio" in model applications related to the 8-hour NAAQS is lower than is the case for the 1-hour NAAQS, if we continue to use absolute model predictions as the basis for the modeled attainment test. This follows since the ozone concentration we are trying to reduce to the level of the NAAQS is closer to that level than is true for the 1-hour NAAQS. Thus, difficulties in choosing an appropriate control strategy which result from model uncertainty are likely to be more pronounced for the 8-hour NAAQS.

3. Uncertainty is reduced if model results are used in a "relative" sense, as we recommend. This follows for at least two reasons. First, the form of test we recommend eliminates much of the uncertainty from the "starting point" of an attainment test. That is, uncertainty associated with measured design values is likely to be considerably less than that associated with a modeled current "design value"[2]. Second, in computing the relative reduction factors used in the test, we keep meteorological inputs constant. Thus, any error resulting from erroneous or uncertain meteorological inputs or from the way the model addresses meteorological conditions is likely to be a systematic one which is present in both current and future predictions. As illustrated by Yang, et al., (1995), taking the ratio of model predictions rather than relying on the absolute predictions reduces some of the associated uncertainty. This likely results from some of the systematic error in the current and future scenarios canceling each other.

4. Focusing the modeled attainment test on receptors near monitoring sites could result in control targets which are too low if the monitoring network is limited or poorly designed. This is why we believe it may sometimes be appropriate to override the modeled attainment test when it is passed. We recommend using a test for selected sites without monitors. Analyses of regional modeling results suggest that relative reduction factors tend to have lower values (i.e., greater reductions occur) if the starting concentrations are higher (ECR, 1998, Meyer, et al., 1997). Lefohn, et al., (1998) report similar findings in their review of recent trend data. Thus, this exercise provides a screening test for identifying a possible need for more controls despite passing the modeled attainment test.

Provided neither the attainment nor screening test yields future design values $\geq$ 90 ppb, a weight of evidence analysis may be used in determining the likelihood of attainment. A 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 modeled 1-hour concentrations or 8-hour daily maxima on the sample of days selected for

---

[2]U.S. EPA quality assurance guidelines specify that precision of ozone measurements shall be within 15%. See 40CFR Part 50, National Ambient Air Quality Standard for Ozone; Final Rule, p.28. **(http://www.epa.gov/ttnamti1/files/cfr/recent/o3naaqs.pdf)**. Precision of averaged numbers (like the design value) in a distribution of measurements would be much greater.

modeling.

Recommendations. States may demonstrate attainment using model results (which includes a modeled attainment test and a screening test applied at selected sites without monitors) corroborated by a weight of evidence determination, which includes a more diverse set of model results plus other analyses. The modeled attainment test should use model predictions in a relative sense to compute relative reduction factors associated with a strategy. These factors should be multiplied by monitored design values at different monitors to estimate future design values to compare with 84 ppb. The modeled attainment test should be supplemented with a screening test which applies similar procedures using the area wide design value at selected sites without monitors.

A weight of evidence determination may be made, unless the attainment test or screening test is failed by a wide margin. A weight of evidence determination includes the modeled attainment and screening test results, plus results of additional model outputs plus other analyses of ambient and emissions data. A weight of evidence analysis may be used either to increase or decrease control targets implied by a modeled attainment test.

3.0 What Is The Recommended Modeled Attainment Test?

In Section 2.0, 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 these issues regarding selection of inputs to the test, and recommend solutions. We next describe how to apply the test and illustrate this with an example. We then identify some implications resulting from the test. We conclude with a further discussion of a screening test recommended for locations without monitors for which predictions are consistently higher than any near a monitoring site.

Equation 3.1 describes the recommended modeled attainment test, applied in the vicinity of monitoring site I.

$$(DVF)_I = (RRF)_I \, (DVC)_I \tag{3.1}$$

where

$(DVC)_I$ = the current design value (e.g., 1997-99) at site I, ppb

$(RRF)_I$ = the relative reduction factor, calculated in the vicinity of site I, unitless. The relative reduction factor is the ratio of the future predicted 8-hour daily maximum concentration to the current predicted 8-hour daily maximum concentration in the vicinity of the monitor, and

$(DVF)_I$ = 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" current design value ($(DVC)_I$) calculated and how is this related to the design value for the traditional nonattainment area?

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

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

(4) On any given modeled day, meteorological conditions may not be similar to those leading to high concentrations (i.e., near the site-specific design value) at a particular monitor. If ozone predicted near a monitor on a particular day is very much less than the design value, the $(RRF)_I$ computed for that day could be unresponsive to controls. Using equation (3.1) could then lead to an erroneously high projection of the future design value. How should a State reduce the likelihood of such a result?

The preceding questions can be lumped under a single question, "how do I select appropriate inputs for the modeled attainment test?"

3.1 How Do I Select Appropriate Inputs For The Modeled Attainment Test?

Calculating the current site-specific design value $(DVC)_I$. The modeled attainment test is linked to the form of the 8-hour NAAQS for ozone through use of a current monitored design value, calculated consistently with the form of the NAAQS. The 8-hour NAAQS is met in an area if, over 3 consecutive years, the average 4th highest 8-hour daily maximum ozone concentration observed at each monitor is $\leq$ 0.08 ppm (i.e., the modeled design value is $\leq$ 84 ppb). In the modeled attainment test, the "average 4th highest daily maximum concentration" is calculated at each monitor by noting the 4th highest daily maximum concentration in each of 3 consecutive years. The arithmetic mean of these 3 values is then computed. If the mean is $\leq$ 84 ppb at all of the monitors, the NAAQS is attained. Table 3.1 illustrates the procedure.

Table 3.1. Example Monitoring Data for Nonattainment Area "A"

(1) Monitoring Site Number I	(2) 4th High 8-hr Daily Max., Year 1	(3) 4th High 8-hr Daily Max., Year 2	(4) 4th High 8-hr Daily Max., Year 3	(5) Average 4th High 8-hr Daily Max.
Monitor 1	92 ppb	87 ppb	91 ppb	90 ppb*
Monitor 2	89 ppb	81 ppb	82 ppb	84 ppb
Monitor 3	94 ppb	83 ppb	81 ppb	86 ppb*

The site-specific current design values ($(DVC)_I$) used in the recommended modeled attainment test are shown in column (5) of the table.

Two of the three monitoring sites in Nonattainment Area "A" have average 4th high daily maxima > 0.08 ppm, so Area "A" is not attaining the NAAQS. The area's design value corresponds to the highest of the average 4th high daily maximum ozone concentrations. Thus, the design value for Nonattainment Area "A" is 90 ppb.

Identifying surface grid cells in the vicinity of a monitoring site. There are three reasons why we believe it is appropriate, in the modeled attainment test, to consider cells "in the vicinity of" 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 model's formulation and 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 reckon grid cell location from there. This is often, and indeed should be, many kilometers from monitoring sites in a modeled traditional nonattainment area. Considering several cells "in the vicinity of" a monitor rather than the single cell containing the monitor diminishes the likelihood of quirks in the test's results resulting from geometry of the superimposed grid system.

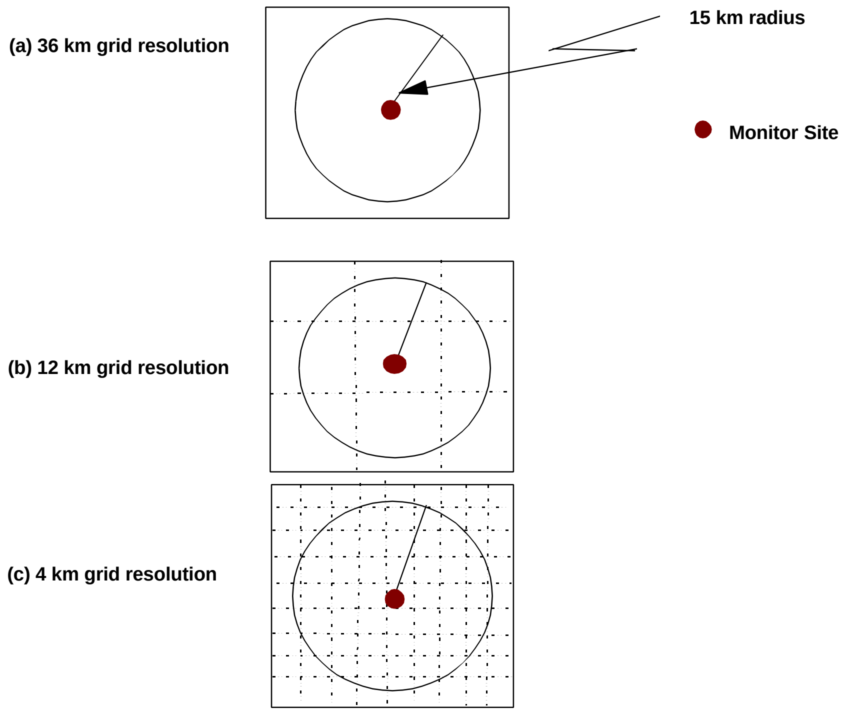
For purposes of the modeled attainment test, if any part of a surface grid cell is located within 15 km of a monitoring site, it should be considered "in the vicinity of" the site. States may consider presence of topographic features or demonstrated mesoscale flow patterns (e.g., land/sea, land/lake interfaces) to refine our definition of "in the vicinity of", provided the justification for doing so is documented. As shown in Figure 3.1, the number of cells considered is a function of the grid's horizontal resolution. A grid system with cells having 12 km on a side would typically include about 9 cells. A system with 4 km resolution would typically include about 45 cells.

Choosing model predictions to calculate a relative reduction factor ($(RRF)_I$) near a monitor. Two decisions need to be made. First, 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. Thus, we recommend calculating an $(RRF)_I$ for each day. The $(RRF)_I$ is the ratio of an 8-hour daily maximum concentration predicted with the future control scenario to an 8-hour daily maximum concentration predicted with current emissions. The 8-hour daily maxima should be used, because they are closest to the form of concentration specified in the NAAQS.

The second decision that needs to be made is, "which one(s) of the 8-hour daily maxima predicted in the vicinity of a monitor should we use to calculate the RRF?" This could be done in any number of ways, but two leading approaches emerge. The first is to choose the nearby grid cell with the highest predicted 8-hour daily maximum concentration with current emissions and the grid cell with the highest predicted 8-hour daily maximum concentration with the future control scenario. Compute the $(RRF)_I$ for each day by taking the ratio of the highest 8-hour daily maximum for the future control scenario to the highest 8-hour daily maximum for current emissions. Note that the grid cell chosen for the future control scenario need not be the same as the one chosen for the scenario with current emissions. Note also that the grid cells chosen to calculate the $(RRF)_I$ may differ on each of the modeled days.

A second approach is to compute an average 8-hour daily maximum concentration for all surface grid cells for the current and future control scenarios on each modeled day. Estimate the $(RRF)_I$ for each day by taking the arithmetic mean of the 8-hour daily maxima with the future control scenario, and dividing it by the arithmetic mean of the 8-hour daily maxima predicted for current emissions.

Figure 3.1. Relationship Between Grid Resolution and Grid Cells Considered to be in the Vicinity of a Monitor

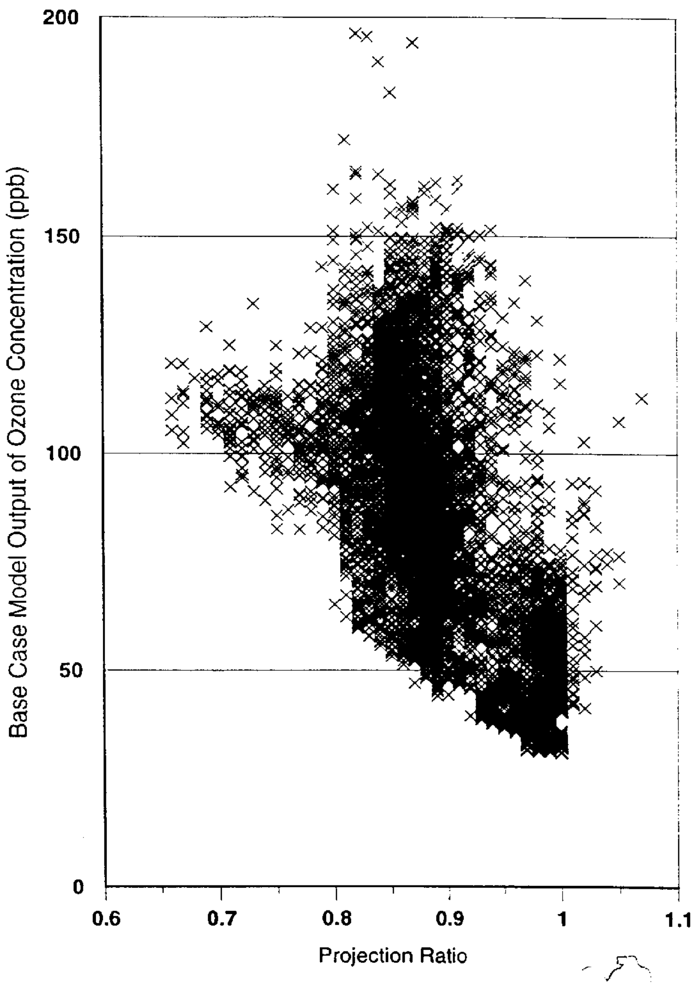


We believe the approach which takes the ratio of the maximum 8-hour daily maxima on each day is preferable for several reasons. First, it is more likely to reflect any phenomena which cause a plume to migrate as a result of implementing controls. Second, it is likely to take better advantage of data produced by a finely resolved modeling analysis. Third, it is somewhat simpler, in that it is not necessary to compute averages for a number of cells (i.e., 45 cells for a grid with 4 km resolution).

Limiting modeled 8-hour daily maxima chosen to calculate RRF. To pass the modeled attainment test, a State needs to show that the future design value $(DVF)_I$ will be less than or equal to 84 ppb for all monitoring sites in the traditional nonattainment area. Figure 3.2 shows that as modeled daily maxima move below the level of the 8-hour NAAQS, there is a tendency for the RRF to have higher values (i.e., the predictions are less responsive to the simulated controls) (ECR, 1998). The scatter in Figure 3.2 likely reflects use of a modeling domain encompassing the eastern half of the United States. Nevertheless there is a general tendency of the RRF to decrease with high base case predictions. Although the episode selection procedure described in Section 11.0 should help focus modeling on days with concentrations near a traditional nonattainment area's design value, there will inevitably be some modeled days where the predicted (as well as observed) ozone concentrations near a monitoring site could be well below the site-specific design value. To illustrate with a simple example, consider a city with 2 monitors, one north of the city and one south. We would expect the site north of the city to observe high ozone, at or near the design value on the selected days with wind blowing to the north. However, on days when the wind is to the south, the northern site may see little benefit from the control strategy. If local emissions are influential in affecting observed concentrations, we would expect to see concentrations well below the northern site's design value on a day with winds to the south. Presumably, there would be several such modeled days, since the analysis needs to provide assurance that a strategy will suffice to meet the NAAQS at all sites in the nonattainment area, including the site south of the city.

As a rule of thumb to avoid likely overestimates of future design values, we recommend excluding some modeled days from consideration when the modeled attainment test is applied to receptors near a specific monitoring site. More specifically, for a site with an observed current design value (DVC) of $\geq$ 85 ppb, States should not consider any day for which the predicted current maximum 8-hour daily maximum concentration at a nearby grid cell is < 70 ppb. For sites with observed current design values (DCV's) $\leq$ 84 ppb, States should not consider any day for which the predicted current maximum 8-hour daily maximum concentration at a nearby grid cell is more than 10 ppb below the site's current (observed) design value.

Figure 3.2. **Base Case Model Output Ozone Concentrations Compared to Projection Ratio**



<u>Recommendations</u>. States should estimate current design values (DVC) for each monitoring site in the nonattainment area using the procedure illustrated in Table 3.1. States should consider modeled 8-hour daily maxima from all surface grid cells for which any part of the cell is within 15 km of the monitoring site. Site-specific relative reduction factors $(RRF)_I$ should be estimated by taking the highest nearby modeled 8-hour daily maximum ozone concentration for the future control scenario and dividing it by the highest nearby modeled 8-hour daily maximum concentration for the base case. Relative reduction factors (RRF) should ordinarily be calculated for each modeled day at each monitoring site. However, certain days may be excluded from consideration at a site if one of two conditions is met:

(a) the site-specific monitored design value is ≥ 85 ppb and the maximum modeled current 8-hour daily maximum ozone concentration on a day is < 70 ppb, or
(b) the site-specific monitored design value is ≤ 84 ppb and the maximum modeled current 8-hour daily maximum ozone concentration on a day is more than 10 ppb below the monitored design value.

3.2 How Do I Apply The Recommended Modeled Attainment Test?

At a minimum, States should apply the modeled attainment test at each monitoring site within a traditional nonattainment area observing a design value of 75 ppb or greater during the 3 consecutive years serving as the basis for a required SIP revision (e.g., 1997-1999). Inputs described in Section 3.1 are applied in Equation (3.1) to estimate a future design value at all such sites. This process is repeated for each monitor in the test. If all resulting predicted future design values (DVF) are ≤ 84 ppb, the test is passed. The modeled attainment test is applied using 4 steps.

Step 1. Compute a site-specific current design value (DVC) from *observed* data.

This is done as illustrated in Table 3.1. The values in column (5) of the table are site-specific design values.

Step 2. Use air quality modeling results to estimate a site-specific relative reduction factor for each modeled day and determine the mean relative reduction factor for that site.

In this step, we begin the process of seeing whether the modeled improvement accompanying a control strategy is consistent with meeting the NAAQS at the site. The modeled relative reduction factor (RRF) in the vicinity of a monitoring site (I) is computed for each modeled day used in the attainment test. This is done by dividing the predicted 8-hour daily maximum near the monitor corresponding with the future control strategy by the predicted 8-hour daily maximum predicted nearby for the current period (e.g., 1997-1999). Equation (3.1) is used.

$$(DVF)_I = (RRF)_I \, (DVC)_I \qquad (3.1)$$

where the terms are as explained previously.

The arithmetic mean value for the RRF at site I is then estimated as shown in Table 3.2. For the illustration shown in Table 3.2, we have assumed that there are 5 modeled days considered in the attainment test. This is inconsistent with the suggestions made in Section 11.0 to model at least 10 days. However, it is done for convenience of illustration. The purpose of recommending 10 or more days be modeled is so that there will be a sample of several days with observed concentrations near the design value at many of the monitoring sites.

Table 3.2. Example Calculation of a Site-Specific Relative Reduction Factor

(1) Day	(2) Observed Current Site-specific Design Value, ppb	(3) Max. Current Predicted 8-hr Daily max.conc near Monitor, ppb	(4) Max.Future Predicted 8-hr Daily max.conc near Monitor, ppb	(5) Modeled Rel. Red. Factor (RRF), Col.(4) / Col.(3)	(6) Mean Relative Reduction Factor, SUM(Col.5) / (#Qualifying Days)
1	95	115	92	0.800	-
2	95	103	90	0.874	-
3	95	95	82	0.863	-
4	95	65	75	Not Estimated	-
5	95	102	87	0.853	-
Mean	-	-	-	-	0.848

Note that on day 4, the predicted current modeled maximum 8-hour daily maximum ozone concentration was very low (< 70 ppb). As discussed in Section 3.1, this day should not be considered at this site.

Step 3. Multiply the observed current design value obtained in Step 1 times the mean relative reduction factor obtained in Step 2. If the projected air quality estimate is ≤ 84 ppb, the test is passed at the monitoring site being evaluated.

In Table 3.2 we see (column (2)) that the current observed design value at Monitor I is 95 ppb. Using equation (3.1), the predicted future design value for site I is,

$$(DVF)_I = (95 \text{ ppb})\ (0.848) = 81 \text{ ppb}$$

Note that in this example the modeled attainment test is passed near monitor I.

Step 4. Repeat steps 1-3 for all ozone monitoring sites with current design values ≥ 75 ppb during the 3 years used as the basis for the current monitored design value. If the test is passed at each site, the modeled attainment test is passed.

<u>Recommendations</u>. The modeled attainment test is applied in 4 steps.

1. Compute a site-specific current design value from *observed* data.

2. Use air quality modeling results to estimate a site-specific relative reduction factor for each modeled day. Use the resulting information to compute a site-specific mean relative reduction factor.

3. Multiply the mean relative reduction factor obtained in step 2 times the corresponding site-specific monitored design value obtained in step 1. The result is an estimated site-specific future design value. If this value is ≤ 84 ppb, the test is passed at the monitor site being evaluated.

4. Repeat steps 1-3 for all ozone monitoring sites with current design values ≥ 75 ppb during the 3 years used to compute the current monitored design value. If the test is passed at each site, the modeled attainment test is passed.

3.3 What Are Key Implications Of The Recommended Modeled Attainment Test?

The recommended modeled attainment test for the 8-hr ozone NAAQS raises some implications which warrant further discussion.

1. The attainment test focuses on monitoring sites. In this sense, the modeled attainment test is identical to the monitored test, used to define whether or not attainment occurs. One shortcoming of this approach is that there are usually only a small number of monitoring sites compared to the area which is modeled. This could result in the model predicting 8-hour daily maximum concentrations in one or more locations which are higher than any predicted near a monitor. If this result occurs consistently (e.g., predicted current 8-hour daily maxima at a location is > 5% above any predicted near monitored locations on 50% or more of the days modeled), it should be investigated further. As described in Section 3.4, a screening test should be completed if there is one or more locations with predictions which are consistently higher than any near a monitoring site.

If the attainment test and screening test are passed or close to being passed, a State may choose to perform a weight of evidence analysis to assess whether attainment is likely. As discussed in Section 4.0, a weight of evidence determination includes an assessment of other model outputs which are less dependent on monitored information. Another important consideration in the weight of evidence assessment would be evidence that the State plans to deploy an ozone monitoring site in a location(s) where the modeling suggests there could be future problems meeting the NAAQS. The weight of evidence approach for demonstrating attainment is discussed in Section 4.0.

2. The attainment test makes assumptions about the shape of a distribution of average 8-hour daily maxima after controls are in place. What is the basis for these assumptions? The attainment test assumes that the distribution of 8-hour daily maxima for each site will flatten after controls are established (e.g., the difference between the 90th percentile concentration and the median concentration diminishes after controls are in place), but that a day's ranking in the tail of a precontrol distribution is likely to be similar to its ranking in the post-control distribution. The first of these assumptions is common sense. Concentrations which are close to background levels already are unlikely to be greatly affected by control measures. This assertion is also borne out by results of a studies by Pacific Environmental Services, Inc. (1997) showing trends in observed distributions of 8-hour daily maxima in 9 cities between 1980 and 1995. The PES study also shows that relative change observed in typical 95th and 99th percentile 8-hour daily maximum ozone observations between 1983 and 1995 is generally similar. A modeling study by Meyer, et al. (1997) lends further support by showing that a day's ranked severity in the pre-control distribution is very similar to its ranking in the post-control distribution of predicted 8-hour daily maxima for numerous cities.

3. The test adjusts observed concentrations during a current period (e.g., 1997-1999) to a future period (e.g., 2010) using model-derived "relative reduction factors". It is important that emissions for the base period used in the test correspond with the period reflected by the current design value (e.g., 1997-1999). Failure to observe this constraint may diminish credibility of the relative reduction factors. For example, if 1990 emissions were compared to emissions projected for 2010, the contrast would likely be greater than if 1996 or 1999 emissions were compared to the 2010 estimates. Presumably, this would lead to larger predicted relative differences in ozone between 1990 and 2010 vs. between 1996 or 1999 and 2010. If the resulting smaller relative reduction factors were applied to 1997-1999 observed design values, a State would likely underestimate 2010 ozone concentrations.

Unfortunately, the constraint described in the preceding paragraph may introduce confusion about just what we mean by "emissions for the base period". That is, the term "base case" is commonly understood to mean emissions corresponding to the episode we are modeling.

For example, if we were modeling a 1991 episode, "base case" emissions would be 1991 emissions. As described in Section 15.0, it is essential to use "base case" emissions *together with* meteorology occurring in the modeled episode in order to evaluate model performance. However, 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 the current observed design value (e.g., 1997-1999) and the period corresponding to the statutory attainment date (e.g., 2010). In this guidance, we refer to the former as the "current" emissions, and continue to refer to emissions used for model performance evaluation as "base case" emissions.

It may often be desirable to model meteorological episodes occurring during the period reflected by the current design value (e.g., 1997-1999). This avoids the need to derive a "current" inventory which differs from the "base case" inventory. It also avoids the need for an additional air quality model simulation corresponding to a "current" inventory which differs from the "base case" inventory. However, episodes need not be selected from the period corresponding to the current design value, provided they are representative of meteorological conditions which commonly occur when exceedances of 0.08 ppm occur. The idea is to use selected representative episodes to capture sensitivity of predicted ozone to changes in emissions during such commonly occurring conditions. There are at least two reasons why using episodes outside the period with the current design value may be desirable: (1) availability of air quality and meteorological data from an intensive field study, and (2) availability of a past modeling analysis in which the model performed well and with which the State is satisfied.

Recommendations. States should review absolute projected future model predictions for 8-hour daily maxima to identify locations with consistently higher predictions than any near a monitoring site. Further investigation is warranted if such locations are identified. To apply the recommended modeled attainment test, States should use emission estimates which correspond (a) with the period represented by the current design value (e.g., 1997-1999), and (b) with the period by which attainment of the NAAQS is required.

3.4 What Additional Model Outputs Should I Review, And Why Is This Necessary?

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 AIRS. The results of this review are intended to address concerns about limiting the modeled attainment test to a small number of monitoring sites.

The additional review is in the form of a screening test which should: (1) identify areas in the domain where absolute predicted 8-hour daily maximum ozone concentrations are *consistently* greater than any predicted in the vicinity of a monitoring site, and (2) for each identified area, multiply a location-specific relative reduction factor times the

current design value for the nonattainment area to estimate a "future design value". If the resulting estimates are less than or equal to 84 ppb at all screened locations, the screening test is passed.

In the first part of the screening test, the word "consistently" is important. An occasional prediction which exceeds any near a monitor is not necessarily indicative of violating a NAAQS which focuses on the 4th highest daily maximum concentration, averaged over 3 consecutive years. Interpretation of "consistently" is discretionary for those implementing the modeling protocol. However, in the absence of any stronger rationale, we recommend the following as a criterion:

-predicted 8-hour daily maxima at the location in question is > 5% higher than any near a monitored location on 50% or more of the modeled days.

The "5%" difference is consistent with the size of rounding differences at 0.08 ppm and is also consistent with previous guidance the EPA has prepared for "transitional" nonattainment areas, which allows some additional latitude in demonstrating attainment when estimates come within 5 ppb of the level specified in the NAAQS (U.S. EPA, 1998). Occurrence of such a difference on 50% or more of the modeled days increases the likelihood that a difference might show up in a design value averaging observations over 3 years.

What do we mean by "areas in the domain" in the first part of the screening test? For each modeled day, States should consider individual surface grid cells with predictions more than 5% greater than any within 15 km of any monitoring site. A 15 km circle should be drawn around the center of each such cell. As a result, several cells may be identified for each modeled day. If any surface cell shows up within these circles on 50% or more of the modeled days, a future design value should be estimated for that cell using screening procedures described in the following paragraphs.

Once one or more locations is identified with current predictions consistently exceeding those near any monitor, we recommend applying a screening method to estimate future design values for such locations. The screening method applies an equation similar to Equation (3.1).

For location j,

$$(DVF_{est})_j = (RRF)_j \, (DVC_{area}) \qquad (3.2)$$

where

DVF_{est} = the estimated future design value obtained with the screening equation (3.2), ppb;

$(RRF)_j$ = the relative reduction factor at location j, computed as recommended in Section

3.1, unitless. In this case, "nearby" surface grid cells should be identified by placing a circle with a 15 km radius at the center of the grid cell with the highest predictions and considering all grid cells with any portion within this circle to be "nearby" cells.

(DVC_{area}) = the design value for the traditional nonattainment area, ppb. This is the highest of the site-specific monitored design values. Thus, for the example shown in Table 3.1, (DVC_{area}) would be 90 ppb.

<u>Recommendations</u>. In addition to applying the modeled attainment test, States should apply a screening estimate for a future design value at all locations with consistently higher current predicted 8-hour daily maxima than those near any monitoring site. The screening test is passed if the predicted future design value is $\leq$ 84 ppb at all flagged locations.

4.0 If I Choose To Use A Weight Of Evidence Determination, What Does This Entail?

As noted in Table 2.2, if the modeled attainment and screening tests are passed or close to being passed, States may use a weight of evidence determination to estimate whether a simulated strategy is likely to meet the 8-hour NAAQS for ozone. A weight of evidence determination examines results from a diverse set of analyses. Each type of analysis has an identified outcome or set of outcomes consistent with a hypothesis that a proposed control strategy is sufficient to meet the NAAQS within the required time frame. If such an outcome occurs, then results of that analysis support the hypothesis that the proposed strategy is adequate. Each analysis is given a weight, depending on the ability of the analysis to address adequacy of a strategy and on the credibility of the analysis. If most (i.e., overall weight of) evidence produced by the diverse analyses supports the hypothesis, then attainment of the NAAQS is demonstrated with the proposed strategy.

Each weight of evidence determination will be subject to area-specific conditions and data availability. Area-specific factors will likely 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 to weigh each needs to be done on a case by case basis by those implementing the modeling/analysis protocol. In Section 4.1, we identify several recommended core analyses which should be used to corroborate one another in a weight of evidence determination. It is appropriate to require considering a core set of analyses to reassure reviewers and the public that a selective set of analyses, supporting a particular viewpoint, has not been chosen. If it is not feasible to perform one or more of the core analyses, a State should document why not. In Section 4.2, we note that additional, optional corroborative analyses may be performed. We provide several examples of such analyses, and identify conditions which should be met for them to be considered in a weight of evidence determination.

4.1 What Analyses Should I Consider In A Weight Of Evidence Determination?

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

Table 4.1 addresses each of the 3 recommended core analyses. In the table, we identify factors which might cause those implementing the modeling/analysis protocol to weigh a particular set of results more heavily (column (2)). We also identify outcomes for each analysis consistent with a hypothesis that emission reductions implied by a strategy are adequate to demonstrate attainment (column (3)).

Table 4.1. Recommended Core Analyses for a Weight of Evidence Determination, Factors Affecting Their Credibility and Outcomes Consistent with Meeting the NAAQS

(1) Type of Analysis	(2) Factors Increasing Credibility of the Analysis	(3) Outcomes Consistent with Hypothesis That a Candidate Strategy will Lead to Attainment
Air Quality Simulation Models	**-good model performance** **-extensive observational data base available** **-short projection periods** **-carefully QA'd inventory** **-confidence in meteorological inputs** **-good ability to pose and address questions about a strategy's adequacy** **-other analyses tend to corroborate conclusions** **-selected episode days have observations near the design value**	**-the modeled attainment test is passed, or nearly passed** **-screened estimates for future design values at sites w/o monitors are ≤ 84 ppb** **-commitment is made to deploy monitors at locations not passing the screening analysis** **-substantial modeled improvement in air quality is predicted using several measures described in Section 4.1.1** **-similar conclusions are reached with other peer reviewed models**

Analysis of Air Quality and Emissions Trends	**-extensive monitoring network exists** **-both ozone and precursor trends are available** **-statistical model used to normalize trend for meteorological differences explains much variance** **-short projection periods used in the analysis** **-a pronounced, statistically significant downward trend is apparent** **-similar conclusions are reached using several trend parameters** **-continued, comparable relative reductions in emissions are provided for**	**-a pronounced downward normalized trend exists in the site-specific design value at all sites with design values greater than 84 ppb.** **-Using projected emissions to extrapolate the air quality trend line to the required attainment date indicates an 8-hour daily maximum concentration ≤ 84 ppb.** **-Other observed air quality trend parameters also show a substantial decline.**
Use of Observational Models	**-an extensive monitoring network exists** **-precursor and indicator species are measured using instruments with appropriate sensitivity** **-monitoring sites appear spatially representative** **-data have been QA'd, and results are self-consistent** **-plausible physical explanations exist for findings**	**-Findings indicate sources controlled in the candidate strategy are important causes of observed high ozone** **-Analysis of indicator species suggests the direction of the strategy (e.g., emphasis on VOC or NOx) is appropriate.**

We discuss each of the recommended core corroboratory analyses in the following subsections.

4.1.1 Air Quality Simulation Models

Weight given to results obtained with air quality simulation models depends on how good the model performance is as well as the rigor with which the performance has been tested. As shown in Figure 3.2, relative reduction factors depend to some extent on the magnitude of currently predicted 8-hour daily maxima. Thus, model results have higher credibility if nearby predicted 8-hour daily maxima agree within about 20% of observations on most or all of the days used to compute the mean relative reduction factors near each monitor. Performance evaluation is discussed further in Section 15.0.

Model applications for which an extensive observational data base exists can be weighed more heavily, especially if the data base includes monitored values of indicator species and precursor data. For ozone, one of the most uncertain inputs to a modeling analysis is the emission projections which must be made to a future year(s) of interest. This uncertainty is reduced if the projection period is short. Hence, weight of evidence provided by modeling is increased with short projection periods. If rigorous quality assurance and review is provided for the model's emissions and meteorological inputs, this may increase confidence that the model is yielding correct answers for the right reasons. Thus, rigor used in preparing model inputs also increases credibility given to the results. Considering days with concentrations near site-specific design values increases confidence that relative reduction factors, developed for use in the tests, are correct.

Selecting episode days which have occurred recently is also advantageous. This follows if there is a long term downward trend in observed ozone. Selection of a "severe" episode (compared to recent observations) from an earlier year may not actually be so severe. In Section 11.0, we suggest a method for characterizing severity which circumvents this problem.

Proper selection of episode days increases confidence in the results of a modeled attainment test. Weight given to the quantitative results of a modeled attainment test is greater if corroborative, more qualitative, analyses yield supporting conclusions about appropriateness of a strategy. Finally, of the analyses available, modeling has the greatest capability for addressing questions about adequacy of a strategy to meet air quality goals in the future. Thus, modeling results should be weighted heavily in a weight of evidence determination.

The outcome from modeling supports use of a proposed strategy for attainment if the modeled attainment and screening tests, described in Section 3.0, are passed. If the attainment test's outcome is a "pass" or "near pass" and the screening test for future design values at receptors without monitors is not passed, a commitment to deploy ozone monitors at such locations should be an important consideration in approving an attainment demonstration. In general, the closer modeled output is to producing unequivocal results, the easier it is for other analyses to produce evidence which supports a conclusion that attainment is likely. Other indicators that a proposed strategy may be adequate are (1) a screening analysis at selected receptor sites without monitors suggests future design values will be less than or equal to 84 ppb, (2) model predictions show major improvements in air quality using a variety of measures, and (3) other peer reviewed atmospheric simulation models show similar results.

We recommend that at least 3 additional model outputs be examined in weight of evidence determinations to provide assurance that passing or nearly passing the recommended modeled attainment and screening tests is indicative of likely attainment. Like the tests, and for similar reasons, each of these additional outputs reflects *relative* changes in predicted air quality.

1. Compute the relative change in surface grid-hours > 84 ppb in the nonattainment area.

This output reflects the frequency with which predicted *hourly* concentrations exceed the concentration specified in the 8-hour NAAQS. Such a measure is not directly related to the form of the NAAQS. Further, if current and future predictions are subject to a systematic bias, this output could be misleading. However, if modeled episodes are chosen to represent a variety of meteorological conditions under which the NAAQS is exceeded at one or more monitoring sites, and the model performs well using the measures described in Section 15.0, a large reduction in the frequency of predicted hourly concentrations of 85 ppb or more is consistent with a conclusion that a proposed strategy would meet the NAAQS.

The decision about what constitutes a "large" reduction in the predicted frequency of hourly concentrations > 84 ppb is subjective, since it is difficult to relate this measure to the NAAQS. As a default assumption, we offer an "80%" reduction. However, those implementing the modeling protocol may use a different criterion if a consensus is reached and its basis is documented.

2. Compute the relative change in the number of grid cells in the nonattainment area with predicted 8-hr daily maxima > 84 ppb.

This output estimates reduction in the pervasiveness of estimated hourly concentrations in excess of the 8-hour concentration specified in the NAAQS. It is subject to the same caveats as the preceding output. One additional complication may occur if there are not many surface grid cells in which current emissions lead to 8-hour daily maximum ozone estimates > 84 ppb. Unless those implementing the modeling protocol agree on a different criterion which is more suitable for their particular application, we offer an "80%" reduction as a default estimate for a "large" reduction.

3. Compute the relative change in the total difference (ppb-hr) of hourly predictions > 84 ppb in the nonattainment area.

Although not the same, this output is similar in concept to the change in the "dosage" to concentrations greater than 84 ppb. Since we are interested in estimating the likelihood that a strategy will lead to attainment rather than estimating reduction in total dosage, this output should be calculated differently from the procedure ordinarily used to calculate dosage metrics, like SUM06. We recommend using Equation (3.3).

$$\text{Relative Difference} = \frac{\text{SUM (Conc. - 84)}_{\text{future}}}{\text{SUM (Conc. - 84)}_{\text{current}}} \qquad (3.3)$$

Where

Conc. = any hourly concentration > 84 ppb

SUM is the sum of the difference for all hours in all surface grid cells within the nonattainment area.

This metric is subject to the same caveats as the preceding two metrics. As with the other two we suggest an 80% reduction as a default estimate for "large reduction".

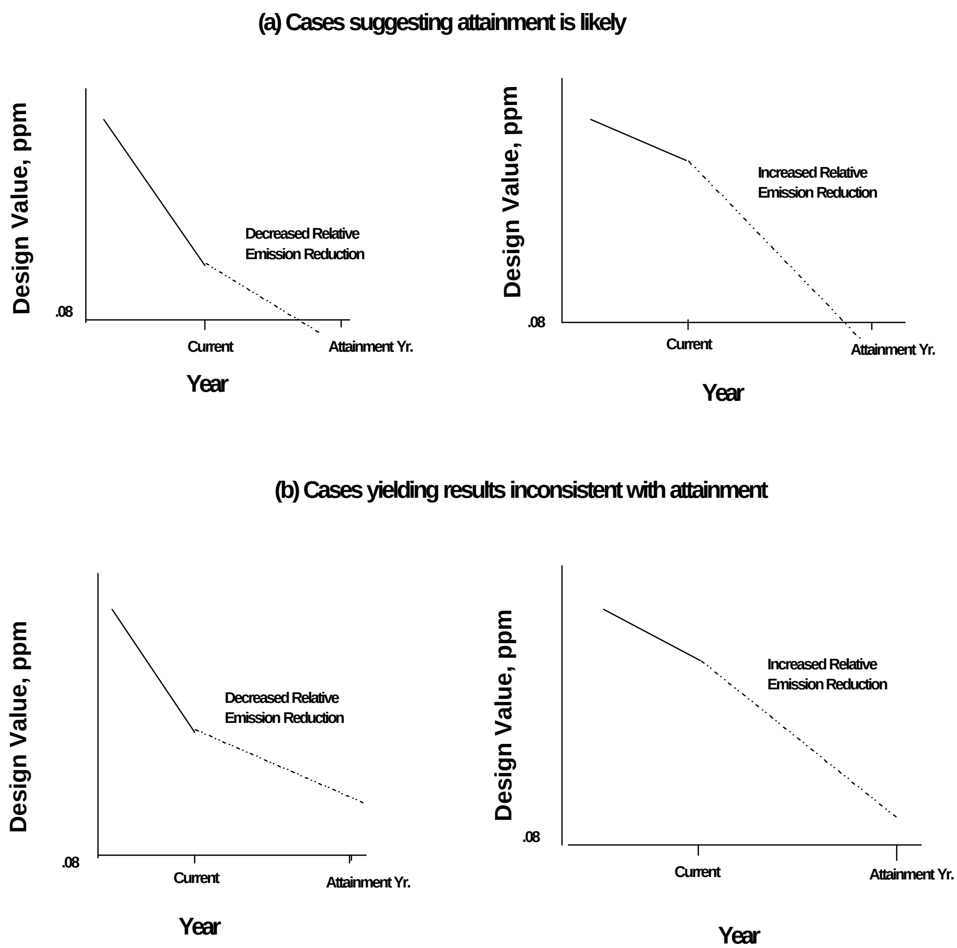
4.1.2 Analysis Of Air Quality And Emissions Trends

This approach is to normalize air quality trends observed over a period for meteorological differences occurring from year to year. The Cox/Chu approach, used extensively in U.S.EPA (1996), is one example of how air quality trends can be normalized (Cox, et al., 1993, 1996). Other procedures can also be used. A curve is fit through the normalized trend and extrapolated to the year in which the air quality goal is to be met. Extrapolations are made by considering past trends as well as past and projected emission reductions. If the trend is statistically significant, the extrapolated value for the attainment year is at or below the air quality goal, and projected relative emission reductions are comparable or greater than reductions occurring during the period for which the trend is constructed, results of the trend analysis suggest a strategy will be adequate. This procedure is illustrated in Figure 4.1 and by the following example.

> **Example.** Estimate the relative reduction in emissions (VOC, NOx or both) occurring during the period corresponding to the observed normalized trend in the average 4th highest 8-hour daily maximum in the nonattainment area. Use the estimated emission trend in concert with the normalized air quality trend to determine an "emission reduction sensitivity factor" (e.g., (ppb)/(percent emission reduction)). Multiply the sensitivity factor times the percent reduction in emissions projected between the current period and the required attainment date. Subtract the result from the current design value to get a projected design value. If the projected design value is $\leq$ 84 ppb, the trend analysis supports a hypothesis that a proposed control strategy will suffice to reach attainment by the required date.

Weight given to trend analyses may depend on several factors. The more air quality data available and the greater variety of trend parameters which show major improvements, the more heavily the results should be weighted. In the case of ozone, availability of trends in ambient precursor data which are consistent with the ozone trends also lend weight to the results. Weight of evidence given to trend results is higher if the procedure used to normalize the trend for meteorological differences explains much of the variability attributable to these differences. Finally, trend analysis is more believable if it is not necessary to extrapolate very far into the future.

Figure 4.1. Examples Showing Use of Trend Analysis in Attainment Demonstrations



4.1.3 Use Of Observational Models

Observational models take advantage of monitored data to draw conclusions about the relative importance of different types of VOC and/or NOx emissions as factors contributing to observed ozone. There are at least 4 approaches with potential for doing this: receptor models (Watson (1997), Henry, et al. (1994) and Henry (1997, 1997a, 1997b)), indicator species approach (Sillman (1997, 1998)), smog produced algorithm approach (Blanchard, et al., (1997, 1997a)) and a relative incremental reactivity approach (Cardelino, et al., (1995)).

Observational models are potentially useful for assessing whether a proposed strategy is oriented toward source categories whose emissions appear to be associated with current observed high ozone. However, their ability to estimate *how much* control is needed is limited, unless one can justify assuming an approximately linear relationship between precursor emissions and observed 8-hour daily maximum ozone. Thus, observational approaches are ideally suited to corroborate results obtained for ozone with more quantitative techniques, like air quality simulation models. Like air quality simulation models, observational models are also subject to uncertainties. Thus, results which are ambiguous should not be considered to be at odds with conclusions reached with an air quality simulation model.

Observational models which rely on use of indicator species can be used to show whether or not ozone is likely to be sensitive to the types of precursors (i.e., VOC or NOx) reduced by a particular control strategy. Receptor models, like the chemical mass balance approach, may be useful for confirming whether a strategy is reducing the right sorts of sources. Observational models can be used to examine days which have not been modeled with an air quality simulation model, as well as days which have been modeled. The resulting information may be useful for drawing conclusions about the representativeness of the responses simulated with the air quality simulation model for a limited sample of days.

Summarizing, if conclusions drawn with one or more observational models suggest that the types of sources to be controlled under a proposed strategy are those that appear associated with high ozone and/or are those to which observed ozone is sensitive, this supports a hypothesis that the strategy is directionally correct.

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.

Recommendations. Weight of evidence determinations are best performed using a variety of diverse analyses in a corroborative fashion. Prior to its application, each

selected analysis should have identified outcomes consistent with concluding a proposed strategy is adequate. At a minimum, States should consider the following 3 types of corroboratory analyses in a weight of evidence determination: (1) output from air quality simulation model(s) (i.e., modeled attainment test results plus other indicators), (2) air quality and emission trend analysis, and (3) interpretation of results obtained with observational models. If it is not feasible to include one or more of the recommended corroboratory analyses, the reasons why not should be documented.

4.2 What If I Want To Consider Additional Corroborative Analyses?

The list of analyses in Section 4.1 is not an exhaustive one. Other types of analyses may supplement the core set recommended for a weight of evidence determination. To have another type of analysis considered, a State should identify why it believes the analysis will produce information which has a bearing on likely attainment of the NAAQS. In addition, the procedure to be used in applying the method and the extent of the data base available to support it should be identified. Finally, prior to application of the method, a State should identify outcomes which would be consistent with a hypothesis that a proposed emission reduction strategy will lead to attainment.

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

Use air quality observations to estimate design values at locations not in the vicinity of a monitoring site. This sort of analysis might be useful, if the screening analysis recommended in Section 3.4 is not passed. Estimates could be made using kriging to interpolate design values at locations of interest. Kriging of urban scale ozone observations to locations without monitors is illustrated by Mulholland, et al. (1998). The objective of such an analysis would be to provide a better estimate of a site-specific current design value than that obtained with the nonattainment area's current overall design value. Equation (3.2) (see Section 3.4) could then be reapplied with this better estimate of "DVC".

If used, a kriging analysis would have to be accompanied by a description of the ground rules used to perform the interpolations. For the results to be credible, they would have to reflect precautions taken against interpolating observations across topographical barriers or making interpolations which are inconsistent with mesoscale flow patterns and spatial patterns of emissions. Assuming proper precautions are taken, an outcome which results in a screening analysis producing estimates of future site-specific design values $\leq$ 84 ppb supports a hypothesis that a control strategy would lead to attainment.

Quantifying uncertainty associated with air quality simulation model estimates. In this guidance, we recommend that "uncertainty" be accounted for using a modeled attainment test which uses models in a "relative" sense and by recognizing that use of corroboratory analyses may be desirable in a weight of evidence determination. Thus, we account for uncertainty in a qualitative way, without actually estimating it. However, several approaches for estimating uncertainty have been proposed in the literature (Gao, et al., (1996), Reynolds, et al., (1997), Yang, et al., (1995)). Many of these approaches assess sensitivity of model predictions to uncertainties in input variables. For outcomes to be most relevant to the way we recommend models be applied in attainment demonstrations, it is preferable that such studies focus on sensitivity of estimated relative reduction factors (RRF) to variations in inputs or model formulations.

In our opinion, uncertainty estimates may, themselves, be uncertain. Thus, if a State chooses to quantify uncertainty, we recommend that the results be used qualitatively as a factor in a weight of evidence determination. Interpretation of the results should be agreed upon on a case by case basis by those implementing the modeling/analysis protocol.

Recommendations. Optional analyses may be considered in addition to the 3 recommended analyses identified in Section 4.1. To use an optional analysis in a weight of evidence determination, a State should (1) explain the rationale for the analysis, (2) identify the data base underlying the analysis, (3) describe the methodology to be used in applying the analysis, and (4) identify outcomes which would be consistent with a hypothesis that a proposed control strategy will suffice to attain the NAAQS.

5.0 How Can I Improve Weight Of Evidence Analyses And Subsequent Reviews Of My SIP Revision?

In Section 4.0, we identified a set of analyses which should be considered when performing a weight of evidence assessment of whether a proposed control strategy will lead to attainment of the 8-hour ozone NAAQS. To be most credible, many of these analyses rely on presence of good ambient data bases. Although commitment to undertake subsequent review of a SIP revision is not a prerequisite for approval of the revision, States should anticipate a need at the required time of attainment to confirm that the NAAQS has indeed been met or a need to diagnose why not. In this Section, we identify measurements and activities which may support weight of evidence analyses and may improve future capabilities to review reasons for attainment or non-attainment of the NAAQS. We conclude by outlining needs to perform future analyses which integrate strategies to meet the 8-hour NAAQS with those to meet goals for the 1-hr ozone NAAQS, $PM_{2.5}$ NAAQS and reasonable progress goals to reduce regional haze.

5.1 What Data Gathering Or Analyses Might Be Helpful To Support Weight Of Evidence Analyses Or Subsequent Reviews?

Efforts to improve the monitored air quality data base and to update and improve emission inventory estimates should both lead to improved weight of evidence analyses and improved subsequent reviews. In this subsection, we identify types of monitoring which may prove helpful. We then briefly discuss efforts to improve the inventory. We conclude by identifying ways in which this improved information might ultimately be used.

Deploying additional air quality monitors. One type of additional monitoring which should be considered has already been mentioned in Sections 3.0 and 4.0. This is to deploy additional ozone monitors in locations which a screening test, described in Section 3.4, suggests may have future design values $\geq$ 85 ppb. This would allow a better future assessment of whether the NAAQS is being met at locations where the model now consistently predicts concentrations higher than any near existing monitoring sites. Further, presence of more dense monitoring networks might improve the reliability with which optional weight of evidence analyses, like kriging, could be used to estimate current design values at other locations without nearby monitors.

Measurement of "indicator species" is a potentially valuable means for assessing which precursor category (VOC or NOx) limits further production of ozone near the monitor's location at various times of day and under various sets of meteorological conditions (some of which may not have been previously considered with an air quality simulation model). Sillman (1998) and Blanchard, et al., (1997) identify several sets of indicator species which can be compared to suggest whether monitored ozone is limited by availability of VOC or NOx. Comparisons are done by looking at ratios of these species. The following appear to be the most feasible for use in the field by a regulatory agency: O_3/NOy, O_3/(NOy - NOx) and O_3/HNO_3. Generally, high values for the ratios suggest ozone is limited by availability of NOx 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 NOy, NOx and/or HNO_3 with high precision (i.e., greater than that often possible with frequently used "routine" NOx measurements). Thus, realizing the potential of the "indicator species method" as a tool in a subsequent review diagnosing why observed ozone concentrations do or do not meet previous expectations may depend on deploying additional monitors. States should consult the Sillman (1998) and Blanchard, et al. (1997) references for further details on measurement requirements and interpretation of observed indicator ratios.

Receptor models are another class of observational approaches which is potentially useful for corroborating assumptions made in air quality simulation models or for diagnosing reasons for unexpected air quality observations in a subsequent review. For use in ozone-related applications, receptor models require observations of VOC species, such as those made in the PAMS network. Receptor models work by noting a combination of speciated source profiles which best explains speciated air quality observations on a day (chemical mass balance approach) or by noting a limited number of species which track each other well from day to day (multi variate statistical approach). Both approaches are limited by collinearity of many of the VOC species. This prevents many distinctive source categories from being identified or leads to inconclusive results concerning which source categories are contributing to observed air quality. Measuring more species is a potential means for reducing the collinearity limitation. An opportunity for doing this may exist as a result of the U.S. EPA's implementation plan for the $PM_{2.5}$ monitoring program (U.S. EPA, 1998b). This plan provides resources for measuring PM species (including some organic particulates) at approximately 300 locations. A State could collocate monitors collecting gaseous organic species at some of these sites. Availability of sites with collocated gas and aerosol phase organic measurements could increase the power of receptor models as diagnostic tools for explaining observations in subsequent reviews. In Section 5.2, we note that "organic carbon" is one of the key components of $PM_{2.5}$. This component is not so well understood as several of the others. This is true, in part, because of the volatility of some species of organic particulate. Well designed studies which measure both gaseous and aerosol phase organics at a site may lead to a better understanding of sources of organic particulates as well as ozone.

Making measurements aloft. Almost all measured ambient air quality data are collected within 10 meters of the earth's surface. However, the modeling domain generally extends several kilometers above the surface. Further, during certain times of day (e.g., at night) surface measurements are not representative of concentrations aloft. Concentrations aloft can have marked effects when they are mixed with ground-level emissions during daytime. Thus, weight given to modeling results can be increased if good agreement is shown with measurements aloft. The most important measurements aloft are ozone, NOy, NO, NO_2, as well as several relatively stable species like CO and selected VOC species. Measurements of SO_2 may also be helpful for identifying presence of plumes from large combustion sources. Measurements of altitude, temperature, water vapor, winds and pressure are also useful. Highest priority should be given to making measurements near sunrise as well as during midday.

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 the chemical mass balance receptor model as well as to air quality simulation models. Efforts to improve speciation profiles for local sources should thus enhance credibility of several of the procedures recommended for use in a weight of evidence determination.

Projecting emission estimates and comparing these to subsequent emission estimates. States addressing traditional nonattainment areas with lengthy attainment dates may find it worthwhile to project emissions to two future years and retain the resulting data files for use in two subsequent reviews. The first of these is the year by which attainment is the goal (e.g., 2010 for ozone). The second is some intermediate year (e.g., 2005-2007). This intermediate projection could be useful to help diagnose reasons for subsequent observed ozone trends which are inconsistent with earlier expectations obtained with an air quality simulation model. Retention of projected emission data bases would enable States to compare the projected inventory estimates with an inventory which is subsequently updated. These checks would be possible after the inventory update for 2005 becomes available, shortly after that year. Similar comparisons would be possible for projections made in nonattainment areas which have 2005-2007 as the required attainment date.

In Section 5.2, we note the likelihood that it will be necessary to consider emissions and air quality impacts in each of the 4 seasons when addressing strategies to meet goals for $PM_{2.5}$ and visibility. Anticipating the need to integrate ozone strategies with strategies for meeting these goals, States may wish not to focus exclusively on summertime emissions of VOC, NOx and CO for purposes of developing a strategy to meet the ozone NAAQS. Estimating emissions for each of the 4 seasons and then noting effects of proposed ozone strategies on such estimates may be a factor in choosing among alternative strategies for meeting the 8-hour ozone NAAQS.

Future diagnostic analyses using air quality simulation models. To facilitate a subsequent review, States should retain meteorological as well as current and projected emission input files developed to support the 2003 SIP revision. When a model is applied with updated emissions estimates and/or with meteorological inputs indicative of episodes chosen in 2005-2007, several useful comparisons are possible if the old files are retained. That is, a State should be better able to determine whether differences are explained by revised emission estimates, poor choice of meteorological episodes in the initial analysis or by changes which have occurred in the model formulation during the intervening years. Insights from such comparisons should help a State explain why changes in the strategy reflected in its 2003 SIP revision may or may not be necessary.

5.2 Why Is It Desirable To Plan For A Subsequent Review?

Commitment to undertake activities supporting subsequent review of a SIP revision is not a prerequisite for approval of the revision. Thus, "why do a subsequent review, and why worry about it now?" Subsequent reviews will be needed at the time attainment of the ozone NAAQS is required for traditional nonattainment areas. The purpose of such a review is to confirm that a NAAQS has indeed been met, or to diagnose available information to determine why not. The required attainment date for traditional nonattainment areas could be as early as 2005-2007, or might be as late as 2010.

5.2.1 Integration With Attainment Strategies For 1-Hr Ozone NAAQS

A 2010 attainment date for the 8-hour NAAQS is possible in areas classified as "severe" or "extreme" nonattainment areas for the 1-hour NAAQS for ozone. For such areas, priority is given to attaining the 1-hour NAAQS. Additional measures, needed to meet the 8-hour NAAQS, may be implemented after the attainment date for the 1-hour NAAQS. Thus, the projected attainment date for areas with the most serious ozone problems occurs well after 2003, when the SIP revision is due. Further, there is a logical time (at the required time of attainment for the 1-hr NAAQS in "severe" nonattainment areas) to review and diagnose emissions and air quality data so as to refine an initial strategy for meeting the 8-hour NAAQS. Such a review, performed at an interim date, helps to ensure that strategies for meeting the 8-hr and 1-hr NAAQS for ozone are "integrated". That is, an 8-hour strategy builds upon the consequences of a strategy for meeting the 1-hour NAAQS, which are reflected in air quality and emissions data circa 2005-2007. Thus, two subsequent reviews are desirable for nonattainment areas with later attainment dates (e.g., at the time when attainment of the 1-hour NAAQS is required in "severe" nonattainment areas, and at the time the 8-hour NAAQS must be met). For traditional nonattainment areas with more immediate attainment dates (i.e., 2005-2007), only one subsequent review is appropriate within the time frame required for attainment. This would most likely occur at the time attainment is required.

5.2.2 Anticipated Modeling Principles For $PM_{2.5}$ And Visibility, And Integrating Ozone Strategies With Goals For $PM_{2.5}$ And Visibility

The U.S. EPA's policy is to encourage integration of control strategies to reduce ozone with those designed later to meet NAAQS for $PM_{2.5}$ and reasonable progress goals to reduce regional haze. We believe such integration will reduce overall costs of meeting multiple air quality goals. The desire to integrate strategies meeting air quality goals for ozone, $PM_{2.5}$ and regional haze is another reason for subsequent review of an ozone SIP revision submitted in 2003. Much of the data base used as a basis for later $PM_{2.5}$ SIP revisions will be collected during 2000-2002. Thus, the scope of the $PM_{2.5}$ problem, if any, will not be fully known at the time modeling and other analyses must be completed to support the 2003 SIP revision for ozone. We anticipate that SIP revisions for $PM_{2.5}$ will be due about 2005-2007---about the same time as a subsequent review of the sufficiency of the previously selected strategy to meet the 8-hour ozone NAAQS. Periodic review of strategies to improve visibility is also likely in this time frame.

We anticipate completing guidance for demonstrating attainment of $PM_{2.5}$ NAAQS and reasonable progress reducing regional haze about the end of 1999. This guidance will be the subject of intense review. Consequently, our current ideas could change. Nevertheless, we present some modeling/analysis "principles" for PM to help States develop data bases and capabilities for considering joint effects of control strategies for ozone, $PM_{2.5}$ and regional haze in a subsequent review of the 2003 SIP revision for ozone.

1. Emissions and meteorological conditions vary seasonally. Effect of a control strategy on annual $PM_{2.5}$ concentrations should be assessed by estimating effects on mean $PM_{2.5}$ for each season and using the resulting information to estimate annual impacts. Emission estimates for VOC, NOx, primary $PM_{2.5}$, sulfur dioxide and ammonia will be needed.

2. PM is a mixture of component species. Each component may be attributable to causes which differ from those for others. The modeled attainment test should separately estimate effects of a control strategy on major components of the mix. Effect of a strategy on $PM_{2.5}$ can be assessed by noting the net effect of a strategy on each major component of the mix. We will likely recommend the following components for separate consideration:

-mass associated with sulfates;
-mass associated with nitrates;
-mass associated with organic carbon;
-mass associated with elemental carbon;
-mass associated with all other species.

3. The recommended modeled attainment test for the annual $PM_{2.5}$ NAAQS will likely focus on monitoring sites with speciated data. Models will be applied in a relative sense to estimate component- and site-specific relative reduction factors. Relative reduction factors will be used with current speciated design values to estimate future design values. A weight of evidence approach will be identified as an alternative to using the modeled attainment test by itself. Because the period of record for measurements is much less than that for ozone, observational models will likely be relatively more important and trend analysis relatively less so.

4. Ambient air quality data should be reviewed to assess whether exceedances of the concentration specified in the 24-hr NAAQS for $PM_{2.5}$ is a hot spot problem significantly influenced by nearby primary emissions, or a problem which is significantly influenced by more pervasive high concentrations of secondary $PM_{2.5}$. If the problem is a hot spot problem and a model performs well predicting primary $PM_{2.5}$ from the nearby source(s), a modeling approach similar to that followed for PM_{10} may be appropriate. If the problem is more pervasive, with important contributions from secondary components of $PM_{2.5}$, or model performance predicting primary $PM_{2.5}$ is poor, a relative approach similar to the approach for the annual NAAQS is likely.

5. Visibility attenuation estimates will be obtained from estimates made for each of the previously identified major components of $PM_{2.5}$. A modeled test for reasonable progress

will estimate relative reduction factors for the major components of $PM_{2.5}$. These will be used with speciated $PM_{2.5}$ concentrations representative of current days with poor visibility in a Class I area to estimate representative future speciated concentrations of $PM_{2.5}$. Current and future visibility extinction coefficients will then be estimated using procedures described in Sisler (1996). Reasonable progress will be determined by comparing estimates of current and future extinction coefficients to see whether a 10% improvement in visibility is likely.

<u>Recommendations</u>. The following data gathering activities may lead to more informative weight of evidence analyses and better subsequent reviews:

-deploy ozone monitors in areas where models consistently predict ozone greater than any predicted near existing monitor sites;

-collocate sufficiently sensitive monitors to measure NOy, NO_2, HNO_3, and NOx at selected ozone monitoring sites;

-collocate monitors to measure gaseous organic species with selected monitors in the $PM_{2.5}$ monitoring network used to estimate particulate organic species;

-make measurements aloft, especially during early morning hours (near sunrise), as well as during midday;

-improve local speciated VOC emission data bases;

-retain meteorological, current and projected emission files as well as output files used in modeling the strategy reflected in the 2003 SIP revision for possible future diagnostic tests with newer data bases and/or models.

A State need not include plans for a subsequent review of its strategy demonstrating attainment of the 8-hour NAAQS for ozone as part of its 2003 SIP revision. However, a subsequent review will be needed at the time of required attainment to ascertain whether attainment has occurred. States with a protracted attainment date for the 8-hour NAAQS may also wish to consider a subsequent review at the time the 1-hour NAAQS should be met (e.g, 2005-2007). Subsequent reviews may be helpful for "integrating" strategies to meet the 8-hour ozone NAAQS with those used to meet the 1-hour NAAQS and with those addressing air quality goals for $PM_{2.5}$ and regional haze.

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

States should follow guidance on reporting requirements for attainment demonstrations in U.S. EPA (1994). The first 7 subjects in Table 6.1 are similar to those in the 1994 guidance. The 1994 guidance envisions an air quality simulation model as the sole means for demonstrating attainment. However, the current guidance (i.e., this document) identifies a weight of evidence determination as a means for corroborating 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 weight of evidence determination, and identification of reviews to which analyses used in the attainment demonstration have been subject.

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

Subject Area	Purpose of Documentation	Issues Included
Modeling/Analysis Protocol	**Communicate scope of the analysis and document stakeholder involvement**	**Names of stakeholders participating in preparing and implementing the protocol;** **Types of analyses performed; Steps followed in each type of analyses;** **Days and domain considered.**
Emissions Preparations and Results	**Assurance of valid, consistent emissions data base. Appropriate procedures are used to derive day-specific emission estimates needed for air quality simulation 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.**

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

Description of the Strategy Demonstrating Attainment	**Provide the EPA and the public an overview of the plan selected in the attainment demonstration.**	**Qualitative description of the attainment strategy;** **Reductions in VOC, NOx, and/or CO emissions from each major source category for each State/county from current (identify) base case emission levels;** **CAA mandated reductions and other reductions;** **Show predicted 8-hr site-specific future design values for the selected control scenario and identify any location which fails the screening test described in Section 3.4;** **Identification of authority for implementing emission reductions in the attainment strategy.**
Data Access	**Enable the EPA or other interested parties to replicate model performance and attainment simulation results, as well as results obtained with other analyses.**	**Assurance that data files are archived and that provision has been made to maintain them;** **Technical procedures for accessing input and output files;** **Computer on which files were generated and can be read;** **Identification of contact person, means for downloading files and administrative procedures which need to be satisfied to access the files.**

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 within the required time.**	**Description of the modeled attainment test and observational data base used;** **Identification of air quality simulation model used;** **Identification of other analyses performed;** **Outcome of each analysis, including the modeled attainment test;** **Assessment of the credibility associated with each type of analysis in this application;** **Narrative describing process used to conclude the overall weight of available evidence supports a hypothesis that the selected strategy is adequate to attain the NAAQS.**
Review Procedures Used	**Provide assurance to the EPA and the public that analyses performed in the attainment demonstration reflect sound practice**	**Scope of technical review performed by those implementing the protocol;** **Assurance that methods used for analysis were peer reviewed by outside experts;** **Conclusions reached in the reviews and the response thereto.**

<u>Recommendations</u>. States should address the 9 subject areas shown in Table 6.1 in the documentation accompanying an attainment demonstration. The documentation should contain a summary section which addresses issues shown in the table. More detailed information should be included in appendices, as necessary.

7.0 References Cited In Part I And In Section 1.0

Blanchard, C.L. and P.M.Roth, (1997), *User's Guide Ozone M.A.P.P.E.R., Measurement-based Analysis of Preferences in Planned Emission Reductions, Version 1.1*, Final Report prepared for the U.S. EPA pursuant to Contract 68-D6-0064 Work Assignment, W.M.Cox, EPA Work Assignment Manager.

Blanchard, C.L., F.W.Lurmann, P.M.Roth, H.E.Jeffries and M.Korc, (1997a), "The Use of Ambient Data to Corroborate Analyses of Ozone Control Strategies", submitted for publication in *Atmospheric Environment.*

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*, pp.161-181.

Clinton, W.J., (July 16, 1997), Memorandum to the Administrator of the Environmental Protection Agency, Subject: "Implementation of Revised Air Quality Standards for Ozone and Particulate Matter".

Cox, W.M. and S.Chu, (1993), "Meteorologically Adjusted Ozone Trends in Urban Areas: A Probabilistic Approach", *Atmospheric Environment 27B*, (4), pp.425-434.

Cox, W.M. and S.Chu, (1996), "Assessment of Interannual Ozone Variation in Urban Areas from a Climatological Perspective", *Atmospheric Environment 30*, pp.2615-2625.

ECR, (1998), *Analysis of the Transitional Area Attainment Test Projection Ratio*, Draft Report prepared for the U.S. EPA, Ellen Baldridge, EPA Work Assignment Manager.

Gao, D., W.R.Stockwell and J.B.Milford, (1996), "Global Uncertainty Analysis of a Regional Scale Gas-Phase Chemical Mechanism", *J.Geophys.Res. 101*, pp.9107-9119.

Henry, R.C., C.W.Lewis and J.F.Collins, (1994), "Vehicle-related Hydrocarbon Source Compositions from Ambient Data: The GRACE/SAFER Method", *Environmental Science and Technology 28*, pp.823-832.

Henry, R.C., (1997), "Receptor Model Applied to Patterns in Space (RMAPS) Part I. Model Description", *J.AWMA 47*, p.216.

Henry, R.C., (1997a), "Receptor Model Applied to Patterns in Space (RMAPS) Part II. Apportionment of Airborne Particulate Sulfur from Project MOHAVE", *J.AWMA 47*, p.220.

Henry, R.C., (1997b), "Receptor Modeling Applied to Patterns in Space (RMAPS) Part III. Apportionment of Airborne Particulate Sulfur in Western Washington State", *J.AWMA 47*, p.226.

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), pp.276A-282A.

Meyer, E.L., K.W.Baldridge, S.Chu and W.M.Cox, (1997), "Choice of Episodes to Model: Considering Effects of Control Strategies on Ranked Severity of Prospective Episode Days", *Paper 97-MP112.01*, Presented at 97th Annual AWMA Meeting, Toronto, Ontario, (June 1997).

Mulholland, J.A., A.J. Butler, J.G. Wilkinson and A.G. Russell, (1998), "Temporal and Spatial Distributions of Ozone in Atlanta: Regulatory and Epidemiologic Implications", *J.AWMA 48*, p.418.

Pacific Environmental Services, Inc., (1997), *Draft Technical Memorandum Analysis of 8-hour Ozone Values: 1980-1995*, Report prepared under EPA Contract No.68D30032, Work Assignment III-88, Edwin L. Meyer, Jr., Work assignment manager

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.

Sillman, S., (1997), "The Method of Photochemical Indicators as a Basis for Analyzing O3-NOx-ROG Sensitivity", NARSTO critical review paper, to be published in *Atmospheric Environment*

Sillman, S., (1998), *Evaluating the Relation Between Ozone, NO_x and Hydrocarbons: The Method of Photochemical Indicators*, EPA/600R-98/022.

Sisler, J.F. (Editor), (1996), *Spatial and Seasonal Patterns and Long Term Variability of the Composition of the Haze in the United States: An Analysis of Data from the IMPROVE Network*, Cooperative Institute for Research in the Atmosphere Report ISSN: 0737-5352-32, Colorado State University, Fort Collins, CO.

U.S. EPA, (1993), *Volatile Organic Compound (VOC)/Particulate Matter (PM) Speciation Data System (SPECIATE), Version 1.5*, EPA/C-93-013.

U.S. EPA, (1994), *Guidance on Urban Airshed Model (UAM) Reporting Requirements for Attainment Demonstration*, EPA-454/R-93-056.

U.S. EPA, (1996), *Guidance on Use of Modeled Results to Demonstrate Attainment of the Ozone NAAQS*, EPA-454/B-95-007.

U.S. EPA, (1998), *Modeling Guidance for Transitional Areas*, in preparation.

U.S. EPA, (1998a), *EPA Third-Generation Air Quality Modeling System, Models-3 Volume 9b Users Manual*, EPA-600/R-98/069(b).

U.S. EPA, (1998b), *Implementation Plan $PM_{2.5}$ Monitoring Program*, Available from EPA Internet website, ***http://www.epa.gov/ttn/amtic/files/ambient/pm25/pmplan3.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. Available from EPA Internet website, ***http://ttnwww.rtpnc.epa.gov/html/o3pmrh/faca.htm#fa06***, then hit "Documents and Issue Papers" and download file "SAW.ZIP".

Yang, Y.J., W.R. Stockwell, J.B.Milford, (1995), "Uncertainties in Incremental Reactivities of Volatile Organic Compounds", *Environmental Science and Technology 29,* pp.1336-1345.

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

8.0 How Do I Apply Air Quality Simulation Models To Help Demonstrate Attainment?--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 simulation 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 7 steps:

1. develop a modeling/analysis protocol;
2. select an appropriate model to use;
3. select appropriate meteorological episodes to model;
4. generate emissions inputs to the air quality simulation model;
5. choose an appropriate area to model with appropriate horizontal and vertical resolution;
6. generate meteorological and air quality inputs to the air quality simulation model;
7. evaluate performance of the air quality simulation model and perform diagnostic tests.

In this Section (Section 8.0), we briefly describe each of these steps to better illustrate how they are interrelated. Because many of these steps require considerable effort to execute, States should take care to keep the appropriate U.S. EPA Regional Office(s) informed as they proceed. This will increase the likelihood of having an approvable attainment demonstration when the work is completed. Steps outlined in this Section are described in greater depth in Sections 9.0 - 15.0.

1. Develop a modeling/analysis protocol. A protocol describes how modeling will be performed to support a particular attainment demonstration. It outlines the methods and procedures which will be used to perform the subsequent 6 steps needed to generate the modeling results. It does this by: a) identifying those responsible for implementing the modeling, b) identifying those who will review each step as it occurs, c) identifying procedures to be used to consider input/suggestions from those potentially affected by the outcome (i.e., "stakeholders"), and d) outlining how decisions will be made concerning technical analyses needed to complete each step in the modeling procedure. In short, the protocol defines the "game plan" and the "rules of the game".

2. Select an appropriate model for use. This step includes reviewing air quality data to gain insight about the nature of a nonattainment area's ozone problem, reviewing rules established in the *Guideline for Air Quality Models* (U.S. EPA, 1998c), and considering experience/expertise of those performing the modeling. 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.

3. Select appropriate meteorological episodes to model. Like the preceding step, this step requires review of available air quality 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 Sections 3.1 and 3.2. Finally, it requires a review of meteorological conditions which have been observed to accompany monitored exceedances of the concentration specified in the NAAQS (i.e., $\geq$ 85 ppb). The object of these reviews is to select episodes which a) include days with observed concentrations close to site-specific design values at as many sites as possible, 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.

4. Generate emissions inputs to the air quality simulation model. Emissions are the central focus in a modeled attainment demonstration. That is, they are the only input to an air quality model which those implementing the protocol can control. Hence, they are the major input which gets changed between the present and future. Emissions which are input to 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. Emissions which are input to the air quality model are also affected by the latter's horizontal/vertical resolution and, of course, the size of the area modeled. In short, treatment of emissions is a central and complex one which, itself, involves several steps. These include deriving emission inventories, quality assuring results, applying results in an emission model(s), and (again) quality assuring results. Emission inputs may be needed for as many as 3 periods: (1) a "base case period" corresponding to that of the selected episodes, (2) a "current period", corresponding to that represented by the current monitored design value, and (3) a future period, corresponding to the required attainment date.

5. Choose an appropriate area to model with appropriate horizontal and vertical resolution. This step is influenced by the choice of episodes modeled. Meteorological and air quality (i.e., ozone) data corresponding to these episodes and, if applicable, to other, plausible episodes, need to be reviewed prior to choosing size of the area modeled. Presence of topographical features or mesoscale meteorological features (e.g., land/sea breeze) near or in the nonattainment area of principal interest are factors to consider in choosing spatial resolution for that portion of the modeling grid. Another factor affecting the chosen resolution is the available spatial detail in the emissions data used as input to an emissions model. Finally, feasibility of managing large data bases and resources needed to simulate finely resolved air quality (and meteorological input) estimates are factors which cannot be ignored in choosing size and resolution of the modeling domain.

6. Generate meteorological and air quality inputs to the air quality simulation model. Unlike emissions, meteorological inputs remain constant during "base case", "current" and "future" periods simulated with the air quality model. Nevertheless care needs to be taken in specifying these, as they may affect relationships predicted between ozone and emissions. Ozone modeling will likely have to consider large geographical areas in many instances. Further, past modeling has shown that meteorological conditions aloft can have an important effect on predicted ozone. Finally, meteorological monitoring is relatively sparse outside of cities and, especially, aloft. Thus, we recommend that meteorological models ordinarily be used to generate meteorological inputs. Application of meteorological models and choice of model grid resolution in the preceding step are closely related. Meteorological conditions near the area which is the focus of 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 rigor with which it might be most desirable to treat dispersion of nearby emissions.

Air quality inputs consist of initial conditions and boundary conditions to the model domain. Importance of initial conditions should be diminished by beginning a simulation at a time prior to the period which is of interest. Nature of boundary conditions is an important factor in deciding how large to make the size of the area modeled. The most satisfactory way to generate future boundary conditions is through use of an air quality simulation model. Therefore, those implementing the protocol will once again be faced with a tradeoff between cost/feasibility of data base management vs. a desire to limit the importance of an arbitrarily specified input to the modeling exercise.

7. Evaluate performance of the air quality simulation model and perform diagnostic tests. To an important extent, credibility of a modeled attainment test's results and other modeled outputs is affected by how well the model replicates observed air quality. Evaluating model performance and conducting diagnostic tests depend on prior definition of the modeling exercise and specification of model inputs. Hence, this is generally the last step prior to using the model to support an attainment demonstration, as described in Part I.

In the past, 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 true in past guidance. These include use of precursor observations, use of indicator species, use of corroborative analyses with observational models and use of retrospective analyses.

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 should remember how model results are used in the modeled attainment test recommended in Section 3.0. In general, model results are used in a relative rather than absolute sense. In particular, the modeled attainment test requires use of relative reduction factors (RRF), generated by models. Thus, diagnostic tests should be used to consider how RRF, as well as absolute ozone predictions, are affected by changes to model inputs.

Recommendations. States should follow seven steps in applying models to generate information required for use in modeled attainment demonstrations.

1. Develop a modeling/analysis protocol
2. Choose an appropriate model
3. Choose appropriate episodes
4. Generate quality assured emission inputs
5. Choose an appropriately sized and spatially resolved modeling domain
6 .Generate appropriate meteorological and air quality inputs
7. Evaluate model performance and undertake diagnostic tests.

Execution of each step should be performed in accordance with procedures identified in the protocol. Rationale and outcome of the steps should be documented as described in Section 6.0. To increase the likelihood of an approvable demonstration, States should carefully coordinate development and execution of steps with the appropriate U.S. EPA Regional Office(s).

9.0 What Does A Modeling/Analysis Protocol Do, And What Does Developing One Entail?

Developing and implementing a modeling/analysis protocol is a very important part of an acceptable modeled attainment demonstration. Much of the information in U.S. EPA (1991) regarding modeling protocols remains applicable. States should review the 1991 guidance on protocols. In this document, we have revised the name of the protocol to "Modeling/Analysis Protocol" to emphasize that the protocol needs to address all types of analyses considered in a weight of evidence determination, not just modeling.

9.1 What Is The Protocol's Function?

The most important function of a protocol is to serve as a means for planning and communicating how a modeled attainment demonstration will be performed *before* it occurs. This should lead to widespread participation in developing the demonstration. It should also reduce risk of spending time and resources on efforts which the appropriate U.S. EPA Regional Office(s) believes are unproductive or inconsistent with Agency policy.

The protocol also serves several important, more specific functions. First, it identifies who will be helping the State or local air quality agency (generally the lead agency) to undertake or evaluate analyses needed to support a defensible demonstration (i.e., the stakeholders). Second, it identifies how communication will occur among stakeholders to develop consensus on various issues. Third, it identifies methods and procedures used to support the demonstration. Fourth, the protocol describes the review process applied to key steps in the demonstration. Fifth, it describes how changes in methods and procedures or in the protocol itself are agreed upon and communicated with stakeholders and the appropriate U.S. EPA Regional Office(s).

9.2 What Subjects Should Be Addressed In The Protocol?

States should address the following subjects in their modeling/analysis protocol:

1. Stakeholders participating in the process.

2. Management/communication procedures used, including those to amend the protocol.

3. Choice of the air quality simulation model to be used and how it meets requirements in 40CFR51, Appendix W for using "alternative" models.

4. Assurance that proposed modeling procedures have been scientifically peer reviewed and plans for technical review of use of the procedures in the specific application and the outputs.

5. Types of analyses included in the weight of evidence determination, if used.

6. Outcomes for each analysis which will be considered consistent with suggesting a selected strategy will meet the NAAQS.

7. Data base used to support air quality modeling and other types analyses used in a weight of evidence determination.

8. Rationale for choice of air quality and emissions model and choice of method for generating meteorological inputs

9. Methods used to quality assure emissions inputs

10. Domain size and spatial resolution to be used.

11. Criteria/goals in selecting periods to model and process to be used in selecting episodes.

12. Performance evaluation procedures and additional diagnostic tests planned.

13. Outcomes in the modeled attainment and screening tests as well as results of analyses used in a broader weight of evidence determination.

14. Procedures to be used to archive, document and report results.

15. Identification of specific deliverables and schedule for delivery to the appropriate U.S. EPA Regional Office.

<u>Recommendations</u>. States should prepare a modeling/analysis protocol as part of an acceptable demonstration of attainment. Generally, procedures recommended in the 1991 guidance and followed for the 1994 ozone SIP revisions are appropriate. These procedures should be augmented to include a discussion of all analyses to be included in the weight of evidence determination, not just modeling. The protocol should also include provision for review of key parts of the analysis and data base underlying the attainment demonstration.

10.0 What Should I Consider In Choosing An Air Quality Simulation Model?

A model application is often the sum effort of applying three types of models: an emissions model, a meteorological model and an air quality model. The first two (emissions and meteorological) provide information needed by the third. Sometimes choice of an emissions or meteorological model or their features is dictated by the chosen air quality model (i.e., they need to be compatible with the selected air quality model). Thus, choice of 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 a model should meet to qualify for use in any attainment demonstration for the 8-hour ozone NAAQS. We then identify several factors which should help in choosing among qualifying models for a specific application. Choice of emissions and meteorological models is discussed in Sections 12.0 and 14.0, respectively.

10.1 What Prerequisites Should A 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 40CFR Part 51, Appendix W (i.e., the "*Model Guideline*") to be proposed early in 1999. Note that, unlike in previous guidance (U.S. EPA, 1991), 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 or easier to use than several alternatives. Thus, at this time, there is no generally "preferred" model for use in attainment demonstrations of the 8-hour NAAQS for ozone. Using the language in 40CFR Part 51 Appendix W, States should consider nested regional models or urban scale gridded photochemical models as "alternative models" for ozone.

"Alternative models" may be used if they are available to the public for free or at a reasonable cost and are not proprietary. In addition, any one of three conditions must be satisfied: (1) the model produces concentration estimates equivalent to estimates obtained with a "preferred" model, or (2) a comparison with observed air quality data shows that the model performs better than the "preferred" model for the given application, or (3) the model is more appropriate than the preferred model for the given application or there is no preferred model. Condition (3) is generally applicable for alternative models being considered to support attainment demonstrations for the 8-hour ozone NAAQS. For "condition (3)" to be satisfactorily met, an alternative model should: (a) have received a scientific peer review, (b) be applicable to the specific application on a theoretical basis, (c) be used with a data base which is adequate to support its application, (d) have performed in such a way that estimates are not likely to be biased low, and (e) be applied consistently with a protocol on methods and procedures.

It is generally recommended practice to perform side by side comparisons between "preferred" and "alternative" models prior to deciding whether an "alternative" model is preferable for a specific application. However, in accordance with "condition (3)", it may not be necessary to perform a side by side performance evaluation with a “preferred model” in a given attainment demonstration for the 8-hour NAAQS for ozone. If a “preferred model” is ultimately identified by the U.S. EPA, a State would have to show that the selected model is more appropriate than the preferred model for the given application. Criteria described in Section 10.2 may be used to support this showing.

The U.S. EPA has invested considerable effort to develop a nested regional model (CMAQ) within a modeling system called “MODELS3" (U.S. EPA, 1998a). The U.S. EPA will provide support, in the form of documentation, user’s guides, computer codes, updates, training and limited troubleshooting for the CMAQ model. The CMAQ model is designed to address ozone, $PM_{2.5}$ and regional haze-related applications. However, this model has not, as yet, been shown to be clearly superior or easier to use than available alternatives. Thus, use of the CMAQ model is subject to the same review criteria as other models proposed to support an attainment demonstration of the 8-hour ozone NAAQS.

Recommendations. For a model to qualify as a candidate for use in an attainment demonstration of the 8-hour ozone NAAQS, a State needs to show that it meets several general criteria.

1. The model has received a scientific peer review.

2. The model can be demonstrated applicable to the problem on a theoretical basis.

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

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

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

6. The developer of the model must be willing to make the model available to users for free or for a reasonable cost, and the model cannot be proprietary.

10.2 What Factors Affect My Choice of A Model For A Specific Application?

States should consider several factors as criteria for choosing a qualifying air quality model to support an attainment demonstration for the 8-hour ozone NAAQS. These factors are: (1) nature of the observed air quality problem; (2) documentation and past track record of candidate models in similar applications; (3) experience of staff and available contractors; (4) required time and resources vs. available time and resources; (5) in the case of regional applications, consistency with regional models applied in adjacent regions.. The first of these factors is used to identify attributes needed for a model to be chosen. Factors (2)-(5) are used to help choose among candidate models having these attributes. If a candidate model is an "alternative" rather than "preferred" model (as described in 40CFR Part 51, Appendix W), a State should demonstrate that it is suitable for use in the specific application using the preceding and other relevant factors. Finally, before 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.

Nature of the observed air quality problem. This is the most important criterion for selecting an appropriate model. Prior to selecting a model to use in an attainment demonstration, we recommend that those implementing the protocol review available air quality, meteorological and emissions data, and take account of the geographic location of the traditional nonattainment area(s) relative to that of precursor emissions.

States should undertake this review to decide whether it is best to use an urban scale photochemical grid model (e.g., domain size ~ 200-300 km on a side, horizontal resolution $\leq$ 12 km) or a regional photochemical grid model (e.g., domain size ~1000 km or more on a side, horizontal resolution $\geq$ 12 km) with or without nesting. Choice between an urban scale and regional application depends on answers to several questions

1. Is transport of ozone (or precursors) into the nonattainment area a likely major contributor to an area's ozone problem?

2. Are nonattainment areas sufficiently numerous and in relatively close proximity so that it is more efficient to estimate control requirements for several nonattainment areas simultaneously?

3. Is the nonattainment area located near major sources of anthropogenic precursors and/or topographical features requiring fine scale resolution to adequately characterize wind flow?

Answers to the preceding questions require a case by case analysis of available air quality, emissions and meteorological data. Generally however, we anticipate that an urban scale model may suffice for "isolated" nonattainment areas (e.g., in the West, outside of California). Locations subject to transported ozone well above natural background (i.e., 8-hr. daily maximum "natural" background is ~ 40-50 ppb) may need to use a regional model. If there are major concentrations of anthropogenic precursor emissions within about 75 km of the area of concern, an urban scale or nested regional model (incorporating a more finely resolved grid over a limited area) is adviseable. An urban scale or nested regional model is also recommended if receptor sites of interest are located near a major body of water.

Documentation and Past Track Record of Candidate Models. For a model to be used in an attainment demonstration, evidence must be presented that it has been found acceptable for estimating hourly or 8-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.

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 general requirements identified in Section 10.1, a State may use the factors described in this section (Section 10.2) to show that it is appropriate for use in a specific application. Choice of an "alternative model" for use in a specific attainment demonstration of the 8-hour NAAQS for ozone needs to be reviewed by the appropriate U.S. EPA Regional Office and the U.S. EPA Model Clearinghouse.

Satisfactory Model Performance in the Specific Application. Prior to use of a selected model's results in an attainment demonstration, it should be shown to perform adequately in the specific application. Means for evaluating model performance are discussed in Section 15.0.

<u>Recommendations</u>. States should first determine what attributes are needed for a qualifying model to address a nonattainment area's ozone problem, and then choose among models possessing these attributes. Five factors should be considered in selecting an air quality model for a specific application. A State may use these factors to demonstrate that use of an "alternative" model is preferable for a specific application. This demonstration should be concurred with by the appropriate U.S. EPA Regional Office and U.S. EPA Model Clearinghouse. The five factors are listed approximately in order of importance.

1. Nature of the air quality problem leading to nonattainment of the ozone NAAQS should first be assessed, and the selected model should have attributes and capabilities consistent with the perceived nature of the problem;

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

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

4. Time and resource constraints may be considered.

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

Prior to using model results in a specific attainment demonstration, a State should show that the model performs adequately in replicating base case observations available for that demonstration.

11.0 How Do I Decide Which Meteorological Episodes To Model?

The modeled attainment test recommended in Section 3.2 uses mean relative reduction factors, averaged over a set of modeled days. For the mean RRF to be a representative value, the number of modeled days used in the test should be greater than some minimal value. We suggest that the minimum number of modeled days (excluding "ramp-up" days) be $\geq$ 10. This number is intended to reconcile difficulties meeting episode selection criteria and resource constraints with the desire to obtain a representative estimate for the RRF which reflects day to day variations.

At a minimum, three criteria should be used to select episodes which are appropriate to model. First, choose episodes which can be readily related to the form of the NAAQS in the modeled attainment test described in Section 3.0. Second, model periods reflecting a variety of meteorological conditions. Third, model periods for which extensive air quality/meteorological data bases exist. The three criteria may sometimes conflict with one another. For example, there may only be a limited number of days with intensive data bases, and these may not cover all of the meteorological conditions which correspond with monitored ozone concentrations close to site-specific design values during the base period. Thus, tradeoffs among the three primary criteria may be necessary in specific applications.

Those implementing the modeling/analysis protocol may use secondary episode selection criteria on a case by case basis. For example, prior experience modeling an episode, may result in its being chosen over an alternative. Another consideration should be to choose episodes occurring during the 3-year period which serves as the basis for the current monitored design value. As we note in Section 3.4, this could save some modeling resources/effort. A third consideration should be to try to ensure that episodes are chosen so that there are several days with monitored ozone concentrations near the site-specific design value at *each* monitoring site in a nonattainment area. If a State chooses to model several nonattainment areas simultaneously (e.g., with a nested regional model), 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 three identified primary criteria for choosing meteorological episodes to model. We then discuss secondary criteria, which may be important on a case by case basis.

11.1 What Are The Most Important Criteria For Choosing Episodes?

Choose episodes having some days with monitored ozone concentrations comparable to the severity implied by the form of the NAAQS. This criterion is an important one, which is closely related to the modeled attainment test recommended in Section 3.0. In Figure 3.2, we saw that the relative reduction factor (RRF) used in the test is not independent of predicted current ozone concentrations (ECR, 1998). This finding is also consistent with inferences drawn from observed trends (Pacific Environmental Services, 1997, Lefohn, et al., 1998) as well as from other modeling studies (Meyer, et al., 1997). Thus, 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 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 base case period can be used to characterize episode severity. This is done by selecting a 3-year period which "straddles" a modeled episode. For example, if an episode from 1995 were modeled, we recommend looking at measured 8-hour daily maxima at each site in the nonattainment area during 1994-1996. 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 three 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 3-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 3-year 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.

The problem of dealing with longer term variations in meteorological conditions producing high ozone can be reduced by assessing the potential of meteorological conditions to form high ozone in concert with a climatological data base. An example of such an approach is described in Cox, et al., (1996). If such an analysis shows that the 3-year periods straddling each selected episode day and the most recent 3-year period are not an extreme ones, this supports using air quality directly to characterize episode severity.

Note that if the episode is drawn from among the 3 years upon which the nonattainment designation is based, days which are chosen are likely to have monitored observations very close to the current design value. "Close to" could be defined in diagnostic tests in specific studies. In the absence of such information, we define "close to" as "$\pm$ 10 ppb" for purposes of prioritizing choice of episodes. If the base and current periods do not coincide, "close to" is within $\pm$ 10 ppb of the design value during the base period straddling the episode. If it is not feasible to meet this criterion for all monitoring sites, meeting it at sites with *current* design values $\geq$ 85 ppb should receive greatest priority.

Choose a mix of episodes which represents a variety of meteorological conditions observed to correspond to monitored ozone having severity comparable to that implied by the form of the NAAQS. This criterion is important, because we want to be reassured that a control strategy will be effective under a variety of conditions leading to ozone concentrations near the current site-specific design values. We believe the most important indicator of variety is differing wind fields. This affects source/source and source/ receptor orientations and, therefore, most likely the effectiveness of a strategy.

Those implementing the modeling/analysis protocol should describe the rationale for distinguishing among episodes which are modeled. The selection is likely to reflect a number of area specific considerations. Qualitative procedures such as reviewing surface and aloft weather maps, observed or modeled wind patterns may suffice for distinguishing episodes with distinctively different meteorological conditions. More quantitative procedures, such as a CART analysis, 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, et al. (1998).

Choose days with intensive data bases. Preference should be given to days with measurements aloft, available measurements of indicator species (see Section 15.0) 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.

We offer the following 7-step procedure as one which may be useful in combining the three primary criteria for selecting episodes to model.

1. For each episode being considered, States should examine *observed* 8-hour daily maximum concentrations at each (sites with design values < 75 ppb can be excluded) monitoring site during the year of the episode, as well as during the year before and the year after the episode. Thus, if one is examining days in a 1991 episode for suitability in the attainment test, severity of the candidate days should be assessed relative to 1990-92 observations at each selected site.

2. For each of the three years, rank the top ten 8-hour daily maxima observed at each of the monitoring sites selected in step 1.

3. Compute the average 1st high 8-hour daily maximum, the average 2nd high 8-hour daily maximum, etc down to the average 10th high 8-hour daily maximum for each selected monitor.

4. Note a range of concentrations which are ± 10 ppb of the average 4th highest value at each site.

5. Classify qualifying days from step 4 into meteorological regimes, using observed or computed wind fields as the primary criterion for classifying the regimes.

6. Note days in the preceding sample for which intensive data bases exist.

7. Give priority to choosing a mix of episodes containing days with observations ± 10 ppb of the site-specific design values during the base period(s), drawn from a variety of meteorological classes identified in step 5, and for which observations aloft, indicator species and/or precursor measurements are available.

<u>Recommendations</u>. States should include at least 10 primary modeled days in their analysis. Three primary criteria should be used as the basis for selecting meteorological episodes for modeling. Tradeoffs among these may often be necessary. Such tradeoffs need to be resolved on a case by case basis.

1. Choose episodes containing days with observed 8-hour daily maximum ozone concentrations within $\pm$ 10 ppb of the average 4th high daily maximum observed at monitoring sites during a 3-year period straddling the period from which each episode is drawn (i.e., days approximately as severe as implied by the form of the NAAQS).

2. Choose episodes containing days which reflect a variety of wind orientations observed to occur when concentrations approximate design values seen during the base period or during the current period.

3. Choose episodes containing days for which measurements aloft, measurements of indicator species and/or precursor measurements exist.

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 3 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 3 primary criteria are met approximately as well by such episodes as they are by other candidate episodes, a State could likely save a substantial amount of work in evaluating model performance.

Choose episodes which are drawn from the period upon which the current design value is based (e.g., 1997-99). As we note in Section 3.4, fewer emission estimates and fewer air quality model simulations are needed if the "base period", used to evaluate model performance, and the "current period", used in the recommended modeled attainment test, are one in the same. Following this criterion could also make the first primary criterion more straightforward. That is, current design values rather than episode severity estimated for a period several years ago could be used as a basis for choice of episodes.

Choose episodes having observed concentrations "close to" implied severity of the form of 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 first 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.

If a State chooses to model several nonattainment areas simultaneously, choose episodes which meet the other criteria in as many of these nonattainment areas as possible. As discussed in Section 10.0, a State or group of States may decide to apply a regional model or a nested regional model to demonstrate attainment in several nonattainment areas at once. Time and resources needed for this effort could be reduced by choosing episodes which meet the other criteria in several nonattainment areas which are modeled.

Recommendations. States may be able to resolve conflicts among the primary criteria for selecting episodes by considering one or more secondary criteria. The following are identified as secondary criteria. States may identify, document and present the rationale for criteria in addition to these if they choose.

1. Give preference to previously modeled episodes.

2. Give preference to episodes occurring during the period corresponding to the current design value.

3. Give preference to episodes maximizing the number of days and sites observing 8-hour daily maxima close to the level of severity specified in the NAAQS.

4. If applying a regional model, choose episodes meeting the other primary and secondary criteria in as many nonattainment areas as possible.

12.0 How Do I Produce Emission Inputs Needed For An Air Quality Simulation Model?

Developing emissions inputs needed in air quality models requires several steps. First, States need to compile statewide emission estimates for precursors of ozone, as well as information subsequently used to spatially and temporally allocate emissions within each county. Next, an emissions model is used to convert the statewide emission information into hourly, gridded, speciated estimates needed by the air quality model. The emissions model also makes use of meteorological information (e.g., temperatures) to adjust emissions from some sources whose emissions are affected by environmental conditions. Both of these steps need to be accompanied by continual efforts to quality assure estimates as they are being made. The final emissions input needed in the modeling process is gridded, speciated, hourly emissions estimates which have been projected to a future year. The projections reflect net effects of growth and the control strategy which is to be simulated with the air quality model.

In the following Sections, we identify information which needs to be compiled on a statewide basis. We then identify emission models which can be used to convert the statewide estimates to the inputs needed by air quality models. We next describe several approaches useful for quality assuring estimates obtained as the first two steps proceed. We conclude with a short discussion of emission projection methods.

While our discussions focus on air quality model needs, there is more extensive guidance on how to prepare emissions inventories. Much of this emission inventory specific guidance has resulted from a joint State/local agency and U.S. EPA effort called the Emission Inventory Improvement Program (EIIP). A series of seven documents have been issued by EIIP:

- Volume I: Introduction and Use of EIIP Guidance for Emissions Inventory Development
- Volume II: Point Sources Preferred and Alternative Methods
- Volume III: Area Sources Preferred and Alternative Methods
- Volume IV: Mobile Sources preferred and Alternative Methods
- Volume V: Biogenics Sources Preferred and Alternative Methods
- Volume VI: Quality Assurance Procedures
- Volume VII: Data Management Procedures

In addition, guidance exists or is being prepared on emission projections, the National Emission Trends inventory methodology, and temporal allocations, spatial allocations, and chemical speciation of emission inventories (U.S. EPA, 1997a, 1997b, 1997c, 1997d, 1997e, 1997f, 1997g, 1998d, 1998e, 1998f). The EIIP documents are available electronically through the U.S. EPA Internet website at ***http://www.epa.gov/ttn/chief/EIIP/techrep.htm.*** States should consult these documents as they prepare their emission inventories.

US EPA ARCHIVE DOCUMENT

12.1 What Statewide Emission Estimates Are Needed To Support Air Quality Models?

Statewide emissions need to be divided into 4 broad categories: stationary point source emissions, stationary area source emissions, mobile emissions for on road and off road sources and biogenic/geogenic emissions. Point sources should be classified by SCC and have associated location information (e.g., latitude/longitude coordinates) as well as diurnal and weekly operating schedules. Area source emissions should be classified by SCC and reported by county. Surrogate factors, used to spatially allocate emissions from the source category within an air quality model grid superimposed over the county, should be identified for each area source category. Examples of surrogate factors might be such things as population or employment by census tract, land use, etc. If information exists concerning diurnal emission patterns for different area source categories, this information should also accompany the statewide area source emission estimates. Onroad and offroad mobile source emissions should be estimated using the most current version of the U.S. EPA MOBILE model (or, in California, the current version of EMFAC) in concert with activity (i.e., vehicle miles traveled (VMT)). The mobile source emission estimates should be accompanied by recommended surrogates for spatially disaggregating the mobile emissions and by diurnal and weekly activity patterns so that gridded, hourly estimates can be obtained for mobile emission estimates in subsequent steps. Estimates for biogenic emissions can be made using the BEIS2 emissions model (Geron, et al., 1994) or updates approved by the U.S. EPA. A State should report biogenic emissions on a county basis. Information regarding land use is needed for each county.

For model applications addressing the ozone NAAQS, emission estimates for each source category should include statewide VOC, NOx and CO estimates for each month of the year. The VOC estimates should be accompanied by a recommended speciation profile for each source category. We recommend that States rely on local measurements to the maximum extent possible for the speciation profile estimates. However, default information on VOC species profiles is available in U.S. EPA (1993), if needed. These data can be obtained electronically through the U.S. EPA's Internet website at ***www.epa.gov/ttn/chief/software.html#speciate.***

Recommendations. States should be familiar with 1997 and 1998 U.S. EPA guidance describing appropriate procedures for estimating statewide emissions needed to support SIP revisions for ozone. Air quality models require emission estimates from point, area, mobile and biogenic sources. In order to convert this information for use in air quality models, VOC species profiles, rationale for suballocating emissions within a county and for assuming diurnal and weekly variability in emissions is needed for each point source and for each major area source category, as well as for mobile sources. Emission estimates for VOC, NOx and CO are needed for each month of the year to support possible use of regional model applications performed for warm weather cities and future needs to integrate ozone and $PM_{2.5}$ control strategies.

12.2 Can I Use the National Emissions Trends Inventory As A Starting Point?

If there are no previously available modeled inventories to serve as a starting point, we recommend that States derive an inventory suitable for use with models starting from the National Emissions Trends inventory (NET) (U.S. EPA , 1998d). The most recent NET reflects statewide, annual emission estimates for VOC, NOx, and CO for 1996. However, the U.S. EPA plans to have a 1999 NET available during the latter half of 2000. The 1999 NET is the preferred starting point for estimating emissions needed to support modeling underlying the 2003 SIP revision. Statewide emissions, by county, are in the NET and are available electronically through the U.S. EPA Internet website at ***www.epa.gov/oar/oaqps/efig/ei***.

12.3 How Do I Convert Statewide Inventory Information Into Data Used In Air Quality Models?

Air quality simulation models predicting ozone require day specific hourly emission estimates for VOC, NOx and CO for each cell of a grid superimposed over the area modeled. Typically, there are thousands of grid cells in a model application. To utilize atmospheric chemistry in the air quality simulation model, VOC emissions also need to have their component chemical species identified. We recommend that source specific, local information be used for this purpose whenever possible. The U.S. EPA maintains a "VOC/PM Species Manual" which can be used when more source-specific information is lacking (U.S. EPA, 1993). Finally, emission factors for some sources are dependent on meteorological conditions such as temperature. Thus, meteorological conditions need to be known to estimate day specific emissions. Emission models should be used to account for the numerous and diverse factors which need to be considered to derive emissions inputs to air quality simulation models. Currently, separate models are used to prepare estimates from anthropogenic stationary vs. mobile sources and from biogenic sources.

Anthropogenic emissions from stationary sources. Two emissions models have been widely used to convert estimated emissions from stationary sources for use in air quality simulation models for ozone-related applications.. The first of these, EPS2, has been frequently used in past urban scale modeling applications for ozone (typically less than ~400 x 400 km modeling domains) (Causley, et al., 1990). EMS95 is the second emissions model which has had wide use (Alpine Geophysics, Inc., 1995). EMS95 has been used in the modeling underlying the U.S. EPA's rule to reduce regional NOx emissions (U.S. EPA 1997h), as well as in other applications of nested regional air quality models. Of the two models, EMS95 is more compatible for use with nested regional models.

The version of EPS2 described in Causley, et al., 1990 may be used only for urban scale model applications for ozone. There is an updated version which has been used in applications of a regional model for particulate matter (Systems Applications International, 1998). As of 1998, EMS95 is probably the most widely used and available emission model for nested regional applications. Thus, it is generally preferred.

Anthropogenic emissions from mobile sources. MOBILE5A is the most current available model to estimate mobile emission factors from a vehicle fleet representative of any specified year (U.S. EPA, 1994a). States other than California should use MOBILE5A or any update to this model identified as appropriate by the U.S. EPA's Office of Mobile Sources. Resulting emission factors need to be combined with activity levels (e.g., vehicle miles traveled) to estimate emission levels which have been suitably disaggregated spatially and temporally for use as inputs in air quality simulation models..

Biogenic Emissions. The BEIS2 emissions model is the recommended default procedure for estimating biogenic emissions (Geron, et al. 1994 and U.S. EPA (1997e)). This model requires a mix of land uses to be specified for each county, as well as hourly temperature information. The model makes use of stored information regarding geographic distribution of plant species, as well as the provided land use and temperature information, to generate gridded biogenic emissions.

Recommendations. States should use emissions models to convert emission inventory estimates into emissions inputs required by air quality simulation models. Emission models require additional inputs concerning chemical speciation, spatial and temporal disaggregation. The following models are recommended as defaults for developing emission inputs to air quality models:

1. EMS95 for stationary anthropogenic sources;

2. MOBILE5A for mobile source emission factors (outside of California);

3. BEIS2 for biogenic sources.

12.4 What Should I Do To Quality Assure Emissions Estimates?

The most efficient means of quality assuring (QA) emission estimates is to apply QA during the initial emissions estimation process. The previously mentioned EIIP quality assurance document, U.S. EPA (1997f), contains a number of QA procedures that should be used to develop the basic emission inventory. Once the basic emission inventory is ready for modeling, there are three additional quality assurance techniques that may be appropriate. The first is to compare emission estimates to estimates reported elsewhere (i.e., a Delphi approach). States can use results from such comparisons to see whether their estimates are unusual in any way. This focuses attention on portions of the inventory which appear to differ from estimates made for other locations, so that the State can confirm whether or not its initial estimates are appropriate. The NET inventory provided by the U.S. EPA may be useful for this approach.

Displaying emissions estimates graphically is also a useful means for quality assuring them. Emissions models identified in Section 12.3 can produce graphic displays useful for quality assurance. For example, a tile plot of emissions made for a grid superimposed over the area to be modeled is an effective means for identifying misplaced sources and for assuring oneself that spatial patterns of emissions are consistent with where sources are believed to be. Other graphical displays include pie charts and time series plots. Pie charts are useful for assessing whether distribution of emissions among source types or categories is consistent with expectations. Time series displays allow a State to look at estimated diurnal patterns in emissions to see whether these appear logical. They enable comparisons to be made for weekends vs. weekdays to see whether estimated differences appear reasonable.

Comparing emissions with monitored air quality is another means for quality assuring emissions estimates. As we place increased emphasis on measurements of ozone precursor species, comparison with monitored speciated data is likely to become an increasingly important means for quality assuring emissions estimates. Availability of speciated VOC data, such as those in the PAMS network or similar data, makes it possible to use monitored observations to apply source attribution approaches (i.e., "receptor models"). A finding suggesting that air quality observations are the product of a mix of emissions which differs greatly from that inferred from the inventory can point the way toward parts of the inventory which may need greater scrutiny. Receptor models and their uses have been summarized by Watson (1997) as well as in Seigneur et al. (1997). Use of ambient data from the PAMS network to quality assure emissions estimates is described in U.S. EPA (1996a).

Recommendations. Quality assurance of emission estimates is an essential part of the modeling process, and should be performed on a continual, ongoing basis. States should consider the following approaches to quality assurance: primary QA emphasis during the initial development of the basic emission inventory, a Delphi approach, computer graphics depicting emissions model estimates, and comparison with speciated air quality data.

12.5 How Do I Estimate Emissions For Future Years?

Emissions projections for sources within a modeling domain are needed – in conjunction with air quality modeling analyses – to determine if a nonattainment area will meet the ozone NAAQS by the required future attainment date. For ozone , we require States to estimate future precursor emissions for at least one future date--the date of required attainment. As discussed in Section 5.0, when a required attainment date is distant (e.g., 2010) from the 2003 date required for a SIP submittal, a State may wish to consider projecting emissions to an intermediate period (e.g., 2005-2007) as well.

The goal in making projections is to try to account for as many of the important variables that affect future year emissions as possible. Each State is encouraged to incorporate in its analysis the variables that have historically been shown to affect its economy and emissions the most, as well as the changes that are expected to take place over the next 10 to 20 years.

Each State should examine the source types that currently dominate its inventory and each should perform some rough calculations to see if that source distribution is likely to change much in the near future. This should suggest the emphasis that might be placed on projection methods for predominant source categories (if there are any). There is normally a wide range of ozone-precursor and fine particulate-emitting source types. Thus, it is probably only in exceptional cases where there are only one or two major source types that dominate the inventory. Large point-source emitters in ozone nonattainment areas are already subject to Reasonably Available Control Technology (RACT) requirements and, in some cases, control technique guidelines (CTGs), which may be identical. Therefore, there are likely to be many different emitters in State ozone and particulate nonattainment areas whose emissions need to be tracked with time. In cases where there are a few dominant sources, special techniques should be used to ensure that those sources are modeled using more sophisticated techniques than those used for the rest of the inventory.

A State's needs for inputs to a grid-based model are a factor in making projections. As noted previously, grid-based models require source locations (coordinates) as input. Thus, a projection approach that makes its computations at this level is preferred. A less desirable alternative is to assume that all growth and retirement occurs at existing facilities and that there is no variation in growth or control within each source category.

Information detailing the different types of projections that might be required of a State or local air pollution control agency can be found in the EPA publication "Procedures For Preparing Emissions Projections" (U.S. EPA, 1991a). In addition to the necessary types of projections, methods for projecting changes in future air pollution generating activities, quantifying the effects of current and future controls, and combining effects of growth and control are addressed in this document. Although last published in 1991, much of this guidance for estimating future year emissions is still valid. There have been updates to some of the information provided in the 1991 guidance (BEA projection phase-out and EGAS and MOBILE model revisions, etc.) and therefore States should review additional documentation concerning emissions projections in U.S. EPA (1998e).

States may find it useful to examine techniques that have been applied in other areas where control strategy planning has been performed. In the 1991 guidance document, examples of emission projection preparation are recorded in a form suitable for input to a grid-based photochemical model. In the simplest sense, this approach relied on developing a growth factor and a control factor for each major source category.

Recommendations. States should review guidance on emission projections issued by the U.S. EPA in 1991 and additional revisions to this document (U.S. EPA, 1998e).

States should concentrate on making use of dominant source type information and variables historically shown to affect their economy. States should be aware of the uses of their emission projections, and factor relevant information into their estimates. States should review techniques previously used for emission projection efforts to utilize existing projections information. States should QA their emission projections using several methods designed to validate the spatial and temporal allocations, as well as any speciation that may be calculated. States should review past emission projection efforts as part of any subsequent review performed for the reasons identified in Section 5.0.

13.0 What Should I Consider When Selecting A Modeling Domain And Its Horizontal/Vertical Resolution?

A modeling domain identifies the geographical bounds of the area which is modeled. Recommended 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. Choice of domain size is also affected by data base management considerations. Generally, these are less demanding for smaller domains.

Horizontal resolution is measured by the horizontal distance covered by a side of an individual grid cell. For example, a grid with a "horizontal resolution of 12 km" is one whose grid cells are 12 km on a side. Vertical resolution is determined by the number of grid cells considered in the vertical direction. Choice of horizontal/vertical resolution depends on spatial variability in emissions, spatial precision of available emissions data, likelihood that mesoscale or smaller scale meteorological phenomena will have a pronounced effect on precursor/ozone relationships, data base management constraints and computer/cost constraints.

We begin this Section by discussing factors States should consider in choosing domain size. Next, we address choice of horizontal and vertical resolution. We conclude by discussing factors affecting choice of size and resolution of coarse scale and fine scale grids considered in a nested modeling analysis.

13.1 How Do I Choose Between An Urban Scale Or Regional Domain?

States may find it useful to examine the gap between a nonattainment area's design value and the level specified in the NAAQS vs. the gap between observed regional (upwind) concentrations and the level specified in the NAAQS. If the former gap is less than the latter, an urban scale analysis may suffice. To illustrate for the case of ozone, if a nonattainment area had a design value of 95 ppb and regional 8-hour daily maxima were typically 60 ppb, the former gap (11 ppb) is substantially less than the latter gap (24 ppb). Depending on the judgment of those implementing the protocol, the strategy for meeting the NAAQS may thus focus on local control measures. An urban scale domain size may be appropriate. In contrast, if the local design value were 95 ppb but corresponding regional daily maxima were typically 80 ppb, the former gap remains 11 ppb, but the latter is reduced to 5 ppb. Those implementing the protocol may wish to consider using regional as well as local measures in such a case. This would necessitate using a regional modeling domain. In general, if additional regionally implemented control measures are expected to materially affect the amount of additional local controls needed to meet the air quality goal, a regional modeling domain should be used. If not, an urban scale domain should suffice.

What do we mean by "urban scale" and "regional" domains? An urban scale domain is one having horizontal dimensions less than ~ 300 km on a side. Horizontal grid resolution should be 12 km or finer. The domain should be large enough to ensure that emissions occurring at 8 am in the domain's center are still within the domain at 8 pm the same day. If recirculation of the nonattainment area's previous day's emissions is believed to contribute to an observed problem, the urban scale domain should be large enough to characterize this. If recirculation encompasses larger scales than 300 km, an urban scale model is probably not sufficient to address an area's problem.

A regional domain is one having horizontal dimensions typically exceeding 1000 km on a side. Horizontal grid square resolution should be 36 km or finer. Data base management problems are likely to make it infeasible to use identical spatial resolution in urban scale and regional models. Nested regional models are intended to address this problem. A nested regional model is one whose domain typically exceeds 1000 km on a side. However only a portion of that domain (e.g., < 300 km on a side) has spatial resolution similar to that recommended for urban scale models. As noted previously, receptors in the vicinity of all ozone monitors within a nonattainment area should be subject to the modeled attainment test. States should ordinarily have the nonattainment area covered by a fine scale grid.

<u>Recommendations</u>. Selection of a domain size depends on the types of control strategies to be simulated. States should review regional (upwind) design values vs. those occurring in the nonattainment area to determine the emphasis to place on regional vs. local controls. If this review suggests that a regional strategy is an important component of an attainment demonstration, then the domain should be regional (>1000 km) in coverage. Otherwise an urban scale analysis (<~300 km) may suffice.

13.2 What Horizontal Grid Resolution Is Necessary?

For urban scale analyses, we recommend horizontal resolution $\leq$ 12 km. If major sources or topographical features are located within 18 km of one or more receptor sites of interest, diagnostic tests should be performed, if possible, to see whether estimated relative reduction factors differ substantially when 4 km resolution is used. If they do, but using 4-km resolution presents data base management or computer-related problems, States may use a 12 km-resolved grid as a screening approach to tentatively select a strategy. That strategy should then be retested using model results obtained with the 4 km horizontal resolution over as large a portion of the domain as feasible.

For the coarse grid in regional scale analyses, we recommend horizontal resolution $\leq$ 36 km. Horizontal resolution for the fine grid in a nested regional scale model should be determined similarly to the preceding guidance for urban scale applications.

<u>Recommendations</u>. Horizontal grid cell size in regional models should be $\leq$36 km, except in areas used to establish boundary conditions for the regional model (where they can be larger). Urban scale analyses and the fine scale portion of a nested regional model should consist of grid cells $\leq$ 12 km in size. States should perform diagnostic

sensitivity tests to see whether resolution finer than 12 km is warranted in the vicinity of receptor sites of interest. If this is the case a 12 km-resolved grid may be used to tentatively choose a strategy from among several candidates. The chosen strategy should be rerun using a more finely resolved grid over as large an area as feasible.

13.3 What Vertical Resolution Is Appropriate?

As described in Section 14.0, the preferred means for generating meteorological data fields for input to air quality simulation models is to use a dynamic meteorological model with four dimensional data assimilation (FDDA). Such models often consider as many as 20-30 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 so many vertical layers may not be feasible nor cost effective. If this is the case, it is desirable to include a number of vertical layers in the meteorological model which can be divided evenly into the number of layers in the air quality model. This should reduce problems interfacing meteorological model outputs with air quality model inputs. If an even division is not feasible, then priority should be given to having lower layers (i.e., closer to the ground) in the air quality and meteorological models correspond evenly. States should ensure that an air quality model's interface program used to incorporate results from more finely resolved meteorological models works as expected. This may be done by citing past evaluations or by examining output from a single test case.

Recognizing costs and data management difficulties, we recommend States initially assume that 5-7 vertical layers are sufficient to characterize relative reduction factors at receptor sites of interest. These layers should not necessarily be of equal depth. We recommend greatest detail be reflected beneath the maximum afternoon mixing height. Vertical resolution in coarse and fine portions of nested regional models should be identical. States should perform a diagnostic sensitivity test to see whether relative reduction factors are affected by increasing the air quality model's vertical resolution. If the results suggest the factors are sensitive to vertical resolution, the coarser resolution may be used in a screening mode to identify a preferred candidate strategy. The strategy may then be rerun with greater vertical resolution, or the results of the sensitivity test may be considered as a factor in a weight of evidence determination. This should be decided on a case by case basis, depending on the feasibility of performing a model simulation with higher vertical resolution over the size of the modeling domain being considered.

<u>Recommendations</u>. States may initially assume that a vertical resolution of 5-7 layers throughout the planetary boundary layer will be sufficient to characterize relative reduction factors used in the modeled tests for attainment. When nested regional models are used, fine and coarse grids should have similar vertical resolution. Sensitivity tests should be performed to test sensitivity of relative reduction factors to vertical grid resolution. If results show RRF are sensitive to vertical resolution. A decision needs to be made on a case by case basis whether to rerun a chosen strategy with greater vertical resolution or consider this sensitivity as a factor in a weight of

evidence determination.

13.4 What Else Should I Consider In Choosing Finely And Coarsely Resolved Portions Of Nested Regional Models?

Coarse Grid Domain. Size of a coarse grid domain should be consistent with the chemical/physical lifetimes of pollutants to be modeled. It should also reflect the purpose for which regional modeling is undertaken. For example, if a regional analysis is performed to assess effects of a regional strategy simultaneously for a number of nonattainment areas, the domain needs to be larger than if only a limited number of nearby areas were the focus of the study.

Lifetimes vary for ozone and its precursors. Lifetime for NOx (i.e., NO + NO_2) is often likely to be less than a day. Regional analyses performed in the U.S. to date suggest that lifetimes for sulfates and nitrates are two days or less (Dennis, 1994). Sources of VOC are believed to be ubiquitous, due to natural emissions. Many of these natural emissions are relatively reactive, so that multi day transport of stable species of VOC or radical products resulting from oxidation of more reactive species may not be a critical factor for selecting size of a domain for modeling ozone. Lifetime for ozone is notoriously difficult to estimate due to the recycling of this compound with free radicals, concentrations of oxidized species of nitrogen and emissions of fresh NOx and VOC precursors which occur in transit. Given information about the lifetime of nitrates however, it is probably safe to assume a lifetime for ozone which is two days or less. The foregoing information suggests that, ideally, the size of a regional modeling domain should be large enough so that emissions occurring two days prior to the beginning of daylight on a modeled day of interest are included within the domain. Thus, we suggest States focus on their receptor areas of interest, perform some screening analyses with trajectory models to ensure that major source areas within two days' travel time are included in the domain.

Fine Grid Domain. Size of the fine grid domain should be influenced by several factors: (1) proximity of receptor sites to major sources of ozone precursors (especially NOx); (2) presence of topographical features which appear to affect observed air quality, and (3) desire to limit resource intensive efforts needed to use numerical models on a fine scale. The last factor is an important concern for use of nested regional models. Size of a fine grid domain could be smaller than that recommended for an urban scale analysis. This follows, since the coarse domain is available to estimate impacts of sources located at intermediate distances from the receptor area, whereas this information is not available for an isolated urban scale analysis. The issue of how far to extend a fine scale grid is one which may need to be resolved on a case by case basis. We recommend that States examine the issue using diagnostic sensitivity tests (see Section 15.0). For consistency with the modeled attainment test, we recommend that the fine grid should initially extend 15 km (i.e., 3-4 4-km grid cells) beyond any receptor of interest if there are topographic features or major sources likely to influence nearby predicted air quality.

<u>Recommendations</u>. Size of a coarse grid should be large enough to include potentially important sources located two days' travel time from receptor sites of interest. Applications which need to consider numerous receptor sites located some distance apart therefore need to use larger domains than do applications focusing on receptors in close proximity to one another. Extent of a fine grid also depends on the number of receptor sites. States should perform diagnostic analyses to ascertain how far a finely resolved grid needs to extend. As a starting assumption, we recommend extending the finely resolved grid sufficiently to include important topographic features or major sources located within 15 km of all receptor sites of interest.

14.0 How Do I Produce Meteorological and Air Quality Inputs Needed By An Air Quality Simulation Model?

After episodes are selected for modeling, corresponding meteorological inputs need to be generated for use in a air quality simulation model. Although the resulting inputs remain constant, they can affect outcomes of a number of the modeling outputs we have identified for scrutiny in Section 4.1. They may also potentially affect relative reduction factors used in the attainment and screening tests. In contrast to meteorological data, air quality inputs may change between the “current” and “future” period used in the modeled attainment test. This presents a potential problem which needs to be addressed.

In this Section, we describe two approaches for generating meteorological inputs to air quality models, and identify advantages/disadvantages associated with each. We note that using dynamic meteorological models with output “nudged” by observations is usually the preferred approach for generating needed meteorological data. Use of these models for horizontal grid resolutions finer than 12 km may present practical problems. We identify ways to eliminate or diminish these. As is the case for emissions, it is important to quality assure meteorological inputs prior to their being used in an air quality simulation model. We next discuss how these inputs can be checked. We conclude by identifying the role of air quality inputs as initial and boundary conditions for an air quality simulation, and note ways to reduce limitations to the simulation resulting from sparseness of these data.

14.1 What Approaches Are Available For Generating Meteorological Data?

Two approaches have been widely used to generate meteorological data needed in air quality simulation models for ozone. The first of these (diagnostic models) relies primarily on observed data and introduces some additional constraints on wind flow due to terrain features. Observed surface temperatures and sounding data are used to develop other information needed to characterize mixing.

The diagnostic wind model most frequently used with ozone models is described by Douglas, et al. (1990). The main advantage of diagnostic models is that they are relatively easy and inexpensive to apply. Further, they make maximum use of wind observations. There are several disadvantages, however. First, there are never enough observations to adequately define a windfield, particularly aloft. Much of the input to the air quality model is derived through interpolation or subjective methods. Because of the sparseness of observations in many areas, we do not encourage use of diagnostic models for generating inputs to regional scale air quality model applications. A second disadvantage is that the meteorological estimates derived with a diagnostic model are not necessarily physically consistent with one another. In the atmosphere, there is a physical dependency existing between temperature, pressure and windfields. This interdependency is not extensively accounted for in diagnostic models, and the extent to which it is considered depends on the expertise of those applying the model. Nevertheless, if ambient concentrations of a pollutant (e.g., as for one or more components of $PM_{2.5}$) are believed to be primarily affected by winds and urban scale source/receptor orientation, the disadvantages are not serious enough to preclude use of diagnostic models.

The second approach for generating needed meteorological data is to use dynamic meteorological models with four dimensional data assimilation (FDDA). These models attempt to characterize theoretical relationships between meteorological variables and topographical/terrain characteristics. Use is made of relatively sparse observations to help steer (i.e., "nudge") solutions so that they do not diverge from observed meteorological fields. Wind observations are typically used for this purpose. See Seaman, (1997) for a further summary of the attributes of dynamic meteorological models. The MM5 (Grell, et al., (1994) and Seaman, et al., (1996)), RAMS (Pielke, et al., (1992) and Lyons, et al., (1995)) and the SAIMM (Systems Applications International, (1996)) models are among those which have been most widely used with numerical air quality simulation models. The major advantage of dynamic meteorological models is that they provide a way of characterizing meteorological conditions consistent with theory, terrain and each other at times and locations where observations do not exist. Disadvantages have been large required computer resources and considerable expertise needed to apply the approach. Recent advances in computer technology have resulted in increased use of dynamic meteorological models for air pollution applications. The MM5 model is used as the default approach with the CMAQ model in MODELS3. As with emissions models, States need to consider compatibility between candidate meteorological models and the air quality model(s) chosen for use. We believe that use of dynamic meteorological models with FDDA is generally the preferable approach for generating meteorological inputs to air quality models for ozone.

Although improvements in computers have made increased use of dynamic meteorological models possible, we have found that data storage requirements and CPU time increase dramatically as the horizontal resolution required of the meteorological model becomes finer. For example, the CPU time needed to generate meteorological data resolved to 12 x 12 km grid cells is considerably greater than the expected factor of "9" increase in that needed to process meteorology for a domain with 36 x 36 km grid cells. This suggests that States may need to limit the spatial extent of areas and the number of episodes for which dynamic meteorological models are used to process meteorological data for grids with horizontal resolution $\leq$12 km.

Recommendations. States should ordinarily use a peer reviewed dynamic meteorological model with four dimensional data assimilation as the means for generating meteorological inputs to ozone models. Peer reviewed diagnostic models may be used on a case by case basis.

14.2 How Do I Deal With Data Management And Computer-related Constraints When Applying Dynamic Meteorological Models?

We believe practical constraints can be mitigated using one of two approaches. The first approach is to use available results from dynamic models on the next greatest coarse scale (i.e., 36 km for a desired 12 km resolution, 12 km for a desired 4 km resolution) to interpolate more finely resolved fields. An objective approach like bilinear interpolation could be used (U.S. EPA, 1991). This approach would be particularly useful if the major reason for desiring finely resolved meteorological estimates is related to a need to resolve emission estimates more finely. For example, in the case of ozone, fine resolution may be needed to most accurately characterize the apparent detrimental effect of NOx reductions on predicted ozone resulting from titration of ozone by nitric oxide near sources of NOx.

A second approach for circumventing major resource requirements needed to apply dynamic models for finely resolved grids considers topographic information (e.g., presence of land/water interfaces) and measured meteorological data to refine fields coarser fields generated by a dynamic model. This second approach may be preferred if the major reason for desiring finely resolved meteorological inputs has to do with perceived importance of mesoscale features which cannot be adequately considered through an objective interpolation procedure. In essence, the second approach is to apply a diagnostic wind model to the wind field generated by the more coarsely resolved dynamic model.

Finally, consequences of using a coarse grid resolution (e.g., 12 km when 4 km might be more desirable) can be reduced by specifying a land use for each cell that corresponds to usage near the major portion of emissions within a cell. This approach is most applicable at land/water interfaces. By assuming the cell is entirely "land", vertical dispersion of fresh emissions is likely to be better characterized. This might also result in a better characterization of subsequent transport of coastal emissions over adjacent large bodies of water.

Recommendations. Prohibitive computer-related constraints associated with applying a dynamic meteorological model to derive a finely resolved (< 12 km) set of meteorological data can be addressed in one of two ways.

1. Interpolate more coarsely resolved data using objective analysis.

2. Apply a diagnostic wind model using "observations" generated by the dynamic meteorological model for a coarser grid. Assume other variables remain the same as for the coarser grid.

Consequences of coarse resolution may be reduced by assigning a land use factor for each surface cell which corresponds to the location of most emissions within the cell.

14.3 How Do I Quality Assure Results Generated By A Meteorological Model?

Ordinarily, we believe the major focus for quality assurance needs to be on wind velocity and temperature predictions. Fortunately, these are two of the most commonly measured meteorological variables. At least 4 methods are available to quality assure results. The first is a variation of the Delphi approach, described in Section 12.0. It consists of comparing results from two different models for a subset of days being considered. For example, MM5 and RAMS results could be compared to note differences in predicted surface temperatures as well as wind velocities at the surface and aloft. Reasons for major differences would then need to be diagnosed.

Computer graphics also provide useful means for quality assuring meteorological outputs. Examining wind vectors for apparent discontinuities is possible using graphics. It is also possible to construct difference diagrams between observed and predicted temperatures and winds. Locations where agreement is poor may suggest areas needing more finely resolved estimates. Geographical orientation between areas of poor agreement and locations of major sources or observed poor air quality may be plotted to judge potential significance of any disagreement.

A third method is to re-apply an air quality model to a previously simulated episode, keeping all inputs unrelated to meteorology constant, but this time using the meteorological model to be used in future simulations. Examine the resulting air quality predictions. Do they agree reasonably well with air quality observations? If not, how do they differ, and what apparent differences are there between meteorological estimates obtained with the new meteorological model and what had been used previously? Do the differences appear to be physically plausible?

A fourth approach is to apply certain calculations (e.g. divergence calculations, mesoscale dimensionless number calculations) to portions of the modeled grid to ensure that the numbers fall within expected ranges.

<u>Recommendations</u>. States should quality assure results from meteorological models prior to using them in the intended air quality model application. Thereafter, further troubleshooting of model results becomes part of the diagnostic analyses for air quality models, described in Section 15.0. In addition to "nudging" initial solutions with FDDA, four approaches for quality assuring meteorological inputs are identified:

1. Compare results obtained for the same day with different dynamic models;

2. Examine wind vector patterns (or patterns for other variables) using computer graphics;

3. Use resulting meteorological inputs in an air quality model and compare to previous "satisfactory" air quality model results obtained using other methods;

4. Depending on the horizontal and vertical size of the domain simulated, compare calculated dimensionless parameters with ranges considered to be appropriate.

14.4 How Do I Address An Air Quality Simulation Model's Need For Air Quality Inputs?

Air quality inputs are needed in air quality simulation models for two purposes: to specify initial conditions, and to specify boundary conditions. There is no satisfactory way to specify initial conditions in every grid cell. Thus, we recommend beginning a simulation at least one day prior to a period of interest for urban scale applications, and two days prior to periods of interest for regional applications to diminish importance of arbitrary assumptions about initial conditions.

Boundary conditions can be specified in several ways. One way is to nest the area of interest within a much larger domain. This approach is exemplified by using nested regional models, as described previously. The need to diminish importance of boundary conditions is why we recommended in Section 13.0 that States use a large regional domain, with the upwind bound 2 or more days' travel time from the area(s) which is the focus of an analysis. If it is not practical to use a nested regional modeling approach or if boundary conditions are believed to be relatively unimportant, a second approach is to use a large single domain in an urban scale analysis. The domain should be approximately symmetrical about the major local sources affecting local monitoring sites of interest , and should be large enough so that emissions occurring in the center of the domain at 8 am remain within the domain at least until 8 pm on the same day. If recirculation is thought to be part of the problem, the domain size would need to be extended to be able to consider it. A third approach is to make use of monitored data and interpolation to estimate boundary conditions. This approach begs the question about what to assume for *future* boundary conditions. It probably works best where boundary conditions are low and are expected to remain so.

Recommendations. Simulations should begin at least one day prior to the period of interest for urban applications and two days for regional applications. Use of nested regional models is the preferred approach for addressing boundary conditions. Where such an approach is not feasible, States should consider a single domain large enough to ensure that emissions occurring at 8 am in the center of the domain remain within the domain at 8 pm on the same day or that next-day recirculation (if important) can be considered.

15.0 How Do I Assess Model Performance And Make Use Of Diagnostic Analyses?

Results of a model performance evaluation should be considered prior to using model results to support an attainment demonstration. Performance of an air quality model can be evaluated in two ways: (1) how well is the model able to replicate observed concentrations of ozone and/or precursors, and (2) how likely is the model to be accurate in characterizing sensitivity of ozone to changes in emissions? The modeled attainment test recommended in Sections 3.1 and 3.2 uses models to predict sensitivity of predicted ozone to controls and then applies resulting relative reduction factors to *observed* (rather than modeled) ozone. Thus, while both types of performance test are important, the second type is the most important. Unfortunately, it is also more difficult to do.

We recommend diagnostic analyses be performed for several reasons. First, these analyses can be used to better understand why the air quality model predicts what it does. This yields further insight into whether or not the predictions are plausible. Second, diagnostic analyses provide information which helps prioritize efforts to improve/refine model inputs. Third, diagnostic tests can provide insight into which control strategies may be the most effective for meeting the ozone NAAQS. Fourth, diagnostic analyses can be used to assess how "robust" a control strategy is. That is, do I reach the same conclusion regarding adequacy of a strategy when using a variety of assumptions regarding current conditions?

In this section, we first identify methods which may be useful for evaluating model performance. We then discuss each of these methods in greater detail. We next note that there is no single method which offers a panacea for evaluating model performance and suggest that performance is best assessed by considering a variety of methods, much as is done in a weight of evidence determination. We conclude this section by identifying methods for performing diagnostic analyses as well as uses for such analyses.

15.1 How Can I Evaluate Performance Of An Air Quality Simulation Model?

As noted above, model performance can be assessed in one of two broad ways: (1) how accurately does the model predict observed concentrations?, and (2) how accurately does the model predict *responses* of predicted air quality to changes in inputs? An example of the latter type of assessment is, "how accurately does the model predict relative reduction factors (RRF)?"

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 current concentrations under a variety of meteorological conditions, this gives us some confidence that we can correctly characterize future concentrations if meteorological conditions are similar. Computer graphics, ozone metrics, precursor metrics and observational models are all potentially useful for evaluating a model's ability to predict base case air quality.

The second kind of model performance assessment can be made in several ways. One way is by looking at predicted differences on weekends vs. weekdays, provided reliable emissions estimates are available for both, and differences attributable to meteorological differences on weekdays and weekend days available for the comparison can be factored out. Comparing predicted vs. observed differences on weekends vs. weekdays provides a potential means for assessing whether a model accurately characterizes sensitivity of ozone to changes in emissions. A second way is to examine predicted and observed ratios of “indicator species”. These chemical species are labeled “indicators”, because they are indicative of whether the model predicts ozone is most sensitive to VOC or to NOx reductions. If predicted and observed ratios of indicator species agree, we can get a better sense that the model is likely to correctly predict sensitivity of ozone to control strategies. Thus, use of indicator species methods show promise for evaluating model performance in a way which is most closely related to how models will be used in attainment demonstrations. We recommend that greater advantage be taken of these methods in the initial demonstration and in subsequent reviews. A third way for assessing a model’s performance in predicting sensitivity of ozone to changes in emissions is to compare projections after the fact with observed trends.. One reason States should retain data files and output generated in simulating the control strategy selected for the 2003 SIP is to facilitate retrospective analyses. As explained in Section 5.0, these analyses provide potentially useful means for diagnosing why a strategy did or did not work as expected. They also provide an important opportunity to evaluate model performance in a way which is closely related to how models are used to support an attainment demonstration.

States can assess model performance using graphics, ozone metrics, predictions of precursor concentrations, ratios of indicator species, corroborative analyses with observational models and retrospective analyses with observed air quality and emission trends. These methods are described in the following subsections. For the 8-hour ozone NAAQS, States should compare 1-hour observations and predictions as well as observed and predicted 8-hour daily maxima.

15.1.1 How Can Computer-generated Graphics Be Used To Assess Model Performance?

States should refer to guidance in U.S EPA (1991) regarding use of graphics to evaluate model performance. Graphs plot predictions and observations. The 1991 guidance describes the following graphical displays: time series plots, tile plots, scatter plots and quantile-quantile (Q-Q) plots. Each of these graphics can also be used to display *differences* between predictions and their paired observations. Graphics are useful means for understanding *how* predictions and observations differ. For example, time series plots tell whether there is any particular time of day or day(s) of the week when the model performs poorly. Tile plots reveal geographic locations where the model performs poorly. Information from tile plots and time series may provide clues about where to focus quality assurance efforts for model inputs. Scatter plots and Q-Q plots show whether there is any part of the distribution of observations for which the model performs poorly. These plots are also useful for helping to interpret calculations of bias between observations and predictions. For example, they could show large differences between observations and predictions which just happen to balance, producing estimated low bias.

15.1.2 How Can Ozone Metrics Be Used To Assess Model Performance?

Ozone metrics produce numerical comparisons between observations and predictions. Usually the comparisons are between observations and predictions paired in time and space, but they need not be so. Appendix C in U.S. EPA (1991) identifies several metrics, as well as the mathematical formulae for calculating them. With the exception of unpaired 1-hour peak concentrations, we recommend that comparisons of observations and predictions for 1-hour sampling times be used as well as comparisons of 8-hour daily maximum concentrations. One-hour comparisons of metrics provide a much larger data base for assessing model performance than would otherwise be available.

Given the use made of predicted 8-hour daily maxima to develop relative reduction factors (RRF), particular attention should be given to comparisons paired in time and, to some extent in space, for predictions and observations of this variable. For example, a comparison between highest predicted 8-hour daily maxima within 15 km of a monitor and the observed daily maxima would be an especially pertinent one. Looking at the relationship between predicted daily maxima and RRF shown in an area-specific version of Figure 3.2 (e.g., obtained through preliminary simulations of across-the board reductions in VOC and/or NOx) should provide some guide to how close agreement should be for results of the modeled attainment test to be most credible.

The three most widely used metrics for ozone have been unpaired 1-hour daily maximum concentrations, normalized bias and gross error. In past guidance we have identified performance criteria for these three measures. These criteria are based on results obtained in urban model applications (primarily in California) during the 1980's. This information may serve as one input in assessing how well a model performs.

We also recommend enhancements to past guidance. We suggest that comparisons be limited to the hours of 8 am- 8 pm, local civil time. We make this suggestion, because at most monitoring sites nighttime observations are probably only indicative of the first few 10's of meters above the ground, and the values are typically low. Vertical resolution in the model may often be too coarse to characterize such gradients. We do not believe this is a critical failing for purposes of an attainment demonstration, however. This follows, since the NAAQS focuses on daily maximum observed (surface) concentrations. Thus, including nighttime in calculating ozone metrics may suggest misleadingly poor model performance.

We suggest that States make efforts to calculate additional metrics closely related to model results considered in a weight of evidence determination. For example, in Section 4.1 we recommended that the measure, “grid-hours with predicted hourly ozone above 84 ppb”, be used to help demonstrate likelihood of attainment. The conclusion reached from this review is strengthened if the model’s ability to predict spatial and temporal extent of observed ozone greater than 84 ppb is shown to be good. A second example is if the State is primarily interested in showing that a strategy works for meeting the 8-hour NAAQS within or downwind of a nonattainment area, it is useful to subdivide the monitoring sites into “downwind”, “center city” and “upwind” categories on each modeled day. Ozone metrics could then be calculated for each category. Note that the identity of “upwind”, “center city” and “downwind” sites could change from day to day. However, the aggregated metrics for all modeled days would be valid for the three defined *categories* of sites.

15.1.3 How Can I Use Available Precursor Observations To Evaluate Model Performance?

Ozone models have many degrees of freedom. This is another way of saying that you can predict similar ozone concentrations using a variety of combinations of (uncertain) inputs. Thus, a comparison of observed/predicted ozone is not a definitive assessment of model performance. Testing the ability of the model to predict other species, as well as ozone, is one means for increasing confidence in the results.

States should include an assessment of how well a model replicates observed VOC, VOC species treated explicitly in the model’s chemical mechanism, CO, NOy, and NO_2 whenever the data base permits. One concern however about monitored precursor data is, what spatial scale do the monitored data represent? Monitored primary pollutant concentrations, like NO, CO, VOC and VOC species, could greatly depend on proximity of the monitor to one or a small number of sources. Models typically consider horizontal scales of 4 km or larger. Thus, the measurements may not be representative of the spatial scale the model is addressing, and the comparison between monitored and modeled data becomes difficult to interpret. This mismatch between models and monitored data is referred to as “incommensurability”.

Potential problems introduced by incommensurability may be reduced in at least three ways. First, States should focus on secondary species like NO_2, or aggregates including secondary species, like NOx or NOy. The rationale for this strategy is that time is required for these secondary species to form, thus permitting greater mixing on the scales assumed in the model. A second strategy is to consider *ratios* of primary (or secondary) pollutants. For example, observed ratios of one or more selected VOC species to CO are likely to be less variable than concentrations of the individual pollutants. Therefore, the ratios may be more representative of the scales considered in the model. A third way for reducing incommensurability is to use metrics which entail spatial averaging in some manner. Comparisons between spatially averaged observations of VOC at 3 monitoring sites with spatially averaged model predictions “near” the 3 sites is an example of such an approach.

15.1.4 How Do I Use Ratios of Indicator Species To Evaluate Model Performance?

A performance evaluation which includes comparisons between modeled and observed ratios of indicator species carries with it a large potential advantage. Such a comparison may reveal whether the model is predicting sensitivity of ozone to VOC and/or NOx controls correctly. That is, when the model predicts ratios within a certain range, predicted ozone is sensitive to changes in a particular one of its precursors. Outside of the range, the predictions are sensitive to changes in another precursor. If a model predicts *observed* ratios of indicator species accurately, this provides some confidence that the predicted *change* in ozone is predicted accurately. The strength of this evidence of adequate performance, of course, depends on the assumption that the model is accurately characterizing the relationships between indicator species and ozone. The validity of this assumption can be more readily tested in smog chamber experiments than can absolute predictions of ozone.

For ozone modeling applications, uses for ratios of indicator species are described by Sillman (1995, 1997, 1998) and by Sillman, et al. (1997a). The authors of these references have shown several ratios of indicator species to be good indicators of whether peak predicted (i.e., *modeled*) ozone is likely to be most sensitive to reductions in VOC or NOx. Many of the species discussed require measurements beyond those which have been routinely made by most State agencies. Of the ratios discussed, the following involve compounds or mixtures most likely amenable to measurement by State agencies: O_3/NOy, O_3/(NOy-NOx) and O_3/HNO_3. States should review the Sillman and Sillman, et al. references for further details about measurement requirements.

15.1.5 How Can I Use Corroborative Analyses With Observational Models To Help Evaluate Air Quality Simulation Model Performance?

Recently, techniques have been developed to embed procedures within the code of an air quality simulation model which enable users to assess contributions of specific source categories or of specific geographic regions to predicted ozone at specified sites (ENVIRON (1997), Yarwood, et al. (1997, 1997a), Morris, et al., (1997), Yang, et al., (1997, 1997a)). These source attribution procedures characterize what the air quality simulation model *says* are the effects of targeted areas or sources on predicted air quality. Provided speciated VOC data are available at a site, source attributions estimated with these approaches can be compared with those obtained using other models which rely directly on observed air quality data.

The chemical mass balance model (Watson (1997)) is probably the most directly applicable observational approach for this purpose, since it can focus on the same day(s) considered with the air quality simulation model. Cautions raised previously about representativeness of the monitored data continue to apply. Available multi variate statistical models (see, for example, Henry, et al., (1994) and Henry (1997, 1997a, 1997b)) may provide a more qualitative means for assessing an air quality simulation model's performance. Multi variate statistical models work by examining temporal variability in monitored precursor species at a single site or spatial variability on one or a few occasions at many sites. A qualitative comparison is possible if one can contrast observations on days when winds suggest a source contribution is unlikely vs. days when a contribution is likely, or at locations where a source category is important vs. those where it isn't. If the observational approach suggests a major change in a source category contribution, and the air quality simulation model also suggests that category is important or unimportant under similar wind conditions, the observational model lends credence to the air quality simulation model's predictions.

15.1.6 Are Retrospective Analyses Useful As A Means For Evaluating Model Performance?

Retrospective analyses compare past model air quality projections with observed trends in air quality and estimated trends in emissions. The approach is a direct assessment of what we are most interested in---does the model accurately predict changes in air quality? However, it is not as straightforward as it seems. Often, input estimates and assumptions used in past studies are ambiguous and the emissions trends are qualitative. Also, the past studies generally assume constant meteorology, which does not happen. One of the purposes of the reporting requirements described in Section 6.0 is to make it possible for others to replicate modeled analyses at future dates.

In Section 5.0, we noted that a retrospective analysis is an important means for diagnosing why a NAAQS has or has not been attained. Such an analysis provides assurance that improved air quality results from changes in emissions rather than meteorology and/or can identify reasons why satisfactory progress is not being observed. Retrospective analyses will have an ancillary benefit of providing an additional means for evaluating model performance. In order to ensure some planning for subsequent retrospective analyses and to promote some uniformity in the methods used for these analyses, they are probably best performed as part of a subsequent review rather than as supporting evidence in the initial SIP revision.

15.1.7 All Of These Performance Tests Have Shortcomings, So What Do I Do?

There is no single definitive test for evaluating model performance. All tests have strengths and weaknesses. Credence given to model results is increased if a variety of tests is applied and the outcomes either support a conclusion that the model is working well or, at a minimum, are ambiguous. Thus, one can think of a model performance evaluation as a "mini-weight of evidence analysis" focused on the issue of how much weight to give model results in an attainment demonstration.

Finally, we need to address the issue of adjusting model inputs to improve model performance. One of the reasons we recommend a variety of tests for model performance is to reduce the possibility of "getting the right answer for the wrong reason". We recognize however, that many of the inputs to models have associated (often unknown) uncertainties. It is acceptable to adjust inputs within reasonable bounds to improve performance, providing it does not result in poorer performance in any of the several measures of performance which we recommend in Sections 15.1.1 - 15.1.6 . If such an adjustment is made, it should be documented and accompanied by an explanation as to why those implementing the protocol believe it is justified.

<u>Recommendations</u>. States should undertake a variety of performance tests. Results from a diverse set of tests should be documented and weighed to qualitatively assess model performance. Provided suitable data bases are available, greatest weight should be given to tests which assess model capabilities most closely related to how the model is used in an attainment test. A narrative describing overall assessment of model performance should be included among the material submitted to support a recommended SIP revision requiring a demonstration of attainment.

15.2 How Can I Make Good Use of Diagnostic Tests?

Diagnostic tests are performed using one of two broad approaches. The first of these consists of tests in which sensitivity of air quality predictions to perturbations in one or a combination of model inputs is examined. This is the more traditional of the two approaches and has a longer track record. When it is applied, States should recall how model outputs are used in the modeled attainment test recommended in Section 3.0. That is, models are used in a relative sense to provide relative reduction factors. Thus, we recommend that sensitivity of these relative reduction factors should be a principal focus of this type of diagnostic test.

The second type of diagnostic test is one in which means for "tracking" the importance of various phenomena contributing to predicted ozone at a location are embedded within the code of an air quality simulation model. This generally increases running time and should not be used unless a vendor or someone very familiar with the computer code has installed this capability and performed benchmark tests to ensure that the model, with and without the code revisions, yields identical results. The major advantage of this latter type of diagnostic test is that it reduces the number of model simulations needed to obtain insights about what is causing high or unexpected ozone predictions. Thus, once the initial effort to develop this capability has been expended, a better understanding of why predictions are the way they are can be obtained relatively efficiently.

Outcomes of diagnostic tests are useful for several purposes. First, the tests can be used to assess the robustness of a control strategy. For example, States can consider effects of assumed boundary conditions and meteorological assumptions on predicted effectiveness of a control strategy. If the control strategy appears to work for a variety of assumptions, this increases confidence in the model results. Second, models used to support ozone NAAQS attainment demonstrations are resource intensive. We have identified possible ways to reduce required resources. For example, we identified reducing vertical resolution in air quality models to 5-7 layers, diminishing the size of a finely resolved (horizontally) domain and performing modeling analyses using a dynamic meteorological model which is no more finely resolved (horizontally) than 12 km (urban applications) and 36 km (coarse grid portion of regional applications). It may not be feasible to assume greater detail for numerous modeling analyses. However, those implementing the modeling/analysis protocol should perform tests to determine sensitivity of relative reduction factors to vertical resolution, horizontal resolution, domain size of a finely resolved grid and boundary conditions on a limited subset of days. Depending on the outcome, it may be advisable to simulate a tentatively chosen strategy with greater resolutionl for limited times and locations. In any event, the resulting information should prove useful in weighing results obtained with the air quality model. Third, diagnostic analyses may help prioritize additional data gathering efforts so that a better subsequent review/diagnosis can be performed at the time of required attainment. Finally, diagnostic analyses could entail perturbing emissions from certain source categories or in certain geographical areas. Results could be useful in prioritizing control efforts or in noting sensitivity of predictions to uncertainties in the current or future emission inventory.

Occasionally a review of a graphical display, like a tile diagram, may indicate a limited number of locations or incidents which bear further investigation. Diagnostic tests may be used to perform focused analyses on these sites or incidents. These tests entail a more detailed look at a time series of predictions and (if available) observations at or above a site, including chemical species, winds and mixing. The examinations can be done qualitatively. However, more quantification is possible using the second type of diagnostic test described at the beginning of this subsection. A procedure called "process analysis" is an example of the second type of diagnostic test. Process analysis has been used to assess relative importance of various model assumptions as well as simulated physical and chemical phenomena contributing to a predicted ozone concentration at a particular time and location (Jeffries (1994, 1997), Jeffries, et al. (1996), Jang, et al. (1995), Lo, et al. (1997)). Process analysis requires a substantial amount of expertise to be properly installed within the code of an air quality model and to be interpreted to full advantage. If a focused diagnostic analysis, such as one obtained with process analysis, suggests a particular model prediction is likely the artifact of a model assumption rather than a result of real chemical/physical atmospheric processes, States may ignore that prediction in the attainment demonstration.

Recommendations. States should include diagnostic analyses as part of the work leading to selection of a control strategy which demonstrates attainment. These analyses should include sensitivity tests to assess robustness of a proposed strategy and consequences of simplifying assumptions made in the modeling. Additional sensitivity tests may be warranted on a case by case basis. Sensitivity of relative reduction factors to input perturbations should be a prime focus of the tests. Provided capabilities have been properly installed and tested, States may use versions of a model's code which contain capability for tracing importance different phenomena as contributors to predicted ozone concentrations at selected locations.

16.0 References Cited In Part II

Alpine Geophysics, Inc., (1995), *The Emissions Modeling System (EMS-95) User's Guide*, Alpine Geophysics, Inc., Boulder CO.

Benjey, W.G., N.M.Moghari, J.W.Susick and D.E.Tivel, (1997), "Processing Emission Inventory Data in the Models-3 Integrated Distributed Modeling System: A Tutorial Example", *Proceedings, Emission Inventory: Planning for the Future*, AWMA Specialty Conference, October 28-30, 1997, Research Triangle Park, NC, pp.107-122.

Causley, M.C., J.L. Fieber, M. Jimenez and L. Gardner, (1990), *User's Guide for the Urban Airshed Model, Volume IV: User's Manual for the Emisssions Preprocessor System*, EPA-450/4-90-007D, U.S. EPA, Research Triangle Park, NC 27711, (NTIS No.: PB91-131250).

Cox, W.M. and S.Chu, (1996), "Assessment of Interannual Ozone Variation in Urban Areas from a Climatological Perspective", *Atmospheric Environment 30*, pp.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.

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).

ECR, (1998), *Analysis of the Transitional Area Attainment Test Projection Ratio*, Draft Report prepared for the U.S. EPA, Ellen Baldridge, EPA Work Assignment Manager.

ENVIRON International Corporation, (1997), *User's Guide to the Comprehensive Air Quality Model with Extensions (CAMx)*, Novato, CA, April 17, 1997.

Geron, C., A.Guenther and T.Pierce, (1994), "An Improved Model for Estimating Emissions of Volatile Organic Compounds from Forests in the Eastern United States", *J.Geophys.Res. 99*, pp.12,773-12,791.

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.

Henry, R.C., C.W.Lewis and J.F.Collins, (1994), "Vehicle-related Hydrocarbon Source Compositions from Ambient Data: The GRACE/SAFER Method", *Environmental Science and Technology 28*, pp.823-832.

Henry, R.C., (1997), "Receptor Model Applied to Patterns in Space (RMAPS) Part I. Model Description", *J.AWMA 47*, p.216.

Henry, R.C., (1997a), "Receptor Model Applied to Patterns in Space (RMAPS) Part II. Apportionment of Airborne Particulate Sulfur from Project MOHAVE", *J.AWMA 47*, p.220.

Henry, R.C., (1997b), "Receptor Modeling Applied to Patterns in Space (RMAPS) Part III. Apportionment of Airborne Particulate Sulfur in Western Washington State", *J.AWMA 47*, p.226.

Jang, J.C., H.E.Jeffries and S.Tonnesen, (1995), "Sensitivity of Ozone Model to Grid Resolution Part II: Detailed Process Analysis for Ozone Chemistry", *Atmospheric Environment 29* (21), p.3114.

Jeffries, H.E., (1994), "Process Analysis for UAM Episode 287", memo to Brock Nicholson, NC DEHNR, April 8, 1994. Available at the following Internet address ***"ftp://Base_URL/ipradocs/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. Available at the following Internet addresses: ***"ftp://Base_URL/ipradocs/uncgri1.pdf"*** and ***"ftp://Base_URL/ipradocs/uncgri2.pdf"***

Jeffries, H.E., (1997), "Use of Integrated Process Rate Analyses to Perform Source Attribution for Primary and Secondary Pollutants in Eulerian Air Quality Models", Presented at U.S. EPA Source Attribution Workshop, Research Triangle Park, NC, July 16-18, 1997. Available at the following Internet websites, ***www.airsite.unc.edu/pdfs/ese_unc/jeffries***, and ***http://ttnwww.rtpnc.epa.gov/html/o3pmrh/faca.htm#fa06***--hit "documents and issue papers" and then download file, "HEJSRCWK.PDF".

Lo, C.S. and H.E. Jeffries, (1997), "A Quantitative Technique for Assessing Pollutant Source Location and Process Composition in Photochemical Grid Models", Presented at annual AWMA Meeting, Toronto, Ontario (1997). Also available at the following Internet address, ***"fty://Base_URL/ipradocs/pca_awma.pdf"***

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), pp.276A-282A.

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.

Meyer, E.L., K.W.Baldridge, S.Chu and W.M.Cox, (1997), "Choice of Episodes to Model: Considering Effects of Control Strategies on Ranked Severity of Prospective Episode Days", *Paper 97-MP112.01*, Presented at 97th Annual AWMA Meeting, Toronto, Ontario, (June 1997).

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.

Pacific Environmental Services, Inc., (1997), *Draft Technical Memorandum Analysis of 8-hour Ozone Values: 1980-1995*, Report prepared under EPA Contract No.68D30032, Work Assignment III-88, Edwin L. Meyer, Jr., Work assignment manager

Pielke, R.A., et al., (1992), "A Comprehensive Meteorological Modeling System-RAMS", *Meteor.Atmos.Phys., 49* , pp.69-91.

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., (1997), "Use of Model-Generated Windfields to Estimate Potential Relative Mass Contributions from Different Locations", Prepared for *U.S. EPA Source Attribution Workshop*, Research Triangle Park, NC, July 16-18, 1997. Available from EPA Internet Website, ***http://ttnwww.rtpnc.epa.gov/html/o3pmrh/faca.htm#fa06***, then hit "Documents and Issue Papers" and download file "SAW.ZIP".

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 the American Petroleum Institute.

Sillman, S., (1995), "The Use of NOy, H2O2, and HNO3 as Indicators for O3-NOx-ROG Sensitivity in Urban Locations", *J.Geophys. Res. 100*, pp.14175-14188.

Sillman, S., (1997), "The Method of Photochemical Indicators as a Basis for Analyzing O3-NOx-ROG Sensitivity", NARSTO critical review paper, to be published in *Atmospheric Environment*

Sillman, S., D.He, C.Cardelino and R.E. Imhoff, (1997a), "The Use of Photochemical Indicators to Evaluate Ozone-NOx-Hydrocarbon Sensitivity: Case Studies from Atlanta, New York and Los Angeles", *J.Air and Waste Mgmt. Assoc.* , In press.

Sillman, S., (1998), *Evaluating the Relation Between Ozone, NO_x and Hydrocarbons: The Method of Photochemical Indicators*, EPA/600R-98/022.

Systems Applications International, (1996), *Users Guide to the Variable Grid Urban Airshed Model (UAMV)*, SYSAPP 96-95/27r, Documentation available for review on US EPA Internet website, ***www.epa.gov/scram001/t22.htm#uamv*** **(CAUTION: the SAIMM meteorological model is considered part of the UAMV operating system. Non-governmental and all foreign user's must obtain an operating license from Systems Applications International before using this model).**

Systems Applications International, (1998), Personal Communication.

U.S. EPA, (1991), *Guideline for Regulatory Application of the Urban Airshed Model*, EPA-450/4-91-013.

U.S. EPA, (1991a), *Procedures for Preparing Emissions Projections*, EPA-450/4-91-014.

U.S. EPA, (1993), *Volatile Organic Compound (VOC)/Particulate Matter (PM) Speciation Data System (SPECIATE), Version 1.5*, EPA/C-93-013.

U.S. EPA, (1994a), Office of Mobile Sources, *User's Guide for MOBILE5*, EPA-AA-AQAB-94-01.

U.S. EPA (1996a), *Photochemical Air Monitoring System 1996 Data Analysis Results Report*, EPA-454/R-96-006.

U. S. EPA (1997a) *EIIP Volume I, Introduction and Use of EIIP Guidance for Emission Inventory Development*, July 1997, EPA-454/R-97-004a**.**

U. S. EPA (1997b) *Point Sources Preferred and Alternative Methods*, July 1997, EPA-454/R-97-004b**.**

U. S. EPA (1997c) *Area Sources Preferred and Alternative Methods*, July 1997, EPA-454/R-97-004c**.**

U. S. EPA (1997d) *Mobile Sources Preferred and Alternative Methods*, July 1997, EPA-454/R-97-004d.

U. S. EPA (1997e) *Biogenics Sources Preferred and Alternative Methods*, July 1997, EPA-454/R-97-004e.

U. S. EPA (1997f) *Quality Assurance Procedures*, July 1997, EPA-454/R-97-004f.

U. S. EPA (1997g) *Data Management Procedures*, July 1997, EPA-454/R-97-004g.

U.S. EPA, (1997h), *Finding of Significant Contribution and Rulemaking for Certain States in the Ozone Transport Assessment Group Region for Purposes of Reducing Regional Transport of Ozone; Proposed Rule*, 62 Federal Register, p.60317, (November 7, 1997).

U.S. EPA, (1998a), *EPA Third-Generation Air Quality Modeling System, Models-3 Volume 9b Users Manual*, EPA-600/R-98/069(b).

U.S. EPA (1998c), Proposed revision to 40CFR, Part 51, Appendix W.

U.S. EPA, (1998d), *National Air Pollutant Emission Trends, Procedures Document 1900-1996*, EPA-454/R-98-xxx.

U.S. EPA (1998e), Guidance for making emissions projections, in preparation.

U.S. EPA (1998f), Guidance on temporal allocations, spatial allocations and chemical speciation, in preparation.

U.S. EPA (1998g), *Science Policy Handbook: Peer Review*, EPA 100-B-98-001. Available from EPA internet website, ***http://www.epa.gov***

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. Available from EPA Internet website, ***http://ttnwww.rtpnc.epa.gov/html/o3pmrh/faca.htm#fa06***, then hit "Documents and Issue Papers" and download file "SAW.ZIP".

Yang, Y.J., J.G.Wilkinson and A.G.Russell, (1997), "Fast, Direct Sensitivity Analysis of Multidimensional Air Quality Models for Source Impact Quantification and Area-of-Influence Identification", Prepared for the U.S. EPA Source Attribution Workshop, Research Triangle Park, NC, July 16-18, 1997. Available from EPA Internet website, ***http://ttnwww.rtpnc.epa.gov/html/o3pmrh/faca.htm#fa06***, then hit "Documents and Issue Papers" and download file "SAW.ZIP".

Yang, Y.J., J.G.Wilkinson and A.G.Russell, (1997a), "Fast, Direct Sensitivity Analysis of Multidimensional Photochemical Models", *Environmental Science and Technology*, In press.

Yarwood, G. and R.Morris, (1997), "Description of the CAMx Source Attribution Algorithm", Prepared for U.S. EPA Source Attribution Workshop, Research Triangle Park, NC, July 16-18, 1997. Available from EPA Internet website, ***http://ttnwww.rtpnc.epa.gov/html/o3pmrh/faca.htm#fa06***, then hit "Documents and Issue Papers" and download file "SAW.ZIP".

Yarwood, G., G.Wilson, R.E.Morris, M.A.Yocke, (1997a), *User's Guide to the Ozone Tool: Ozone Source Apportionment Technology for UAM-IV*, Prepared for Mr. Thomas Chico, South Coast Air Quality Management District, Diamond Bar, CA, March 28, 1997.