United States Environmental Protection Agency

Office of Air Quality Planning and Standards
Research Triangle Park, NC 27711

EPA-454/B-95-007
June 1996

AIR

EPA

GUIDANCE ON USE OF MODELED RESULTS TO DEMONSTRATE ATTAINMENT OF THE OZONE NAAQS

GUIDANCE ON USE OF MODELED RESULTS

TO DEMONSTRATE ATTAINMENT OF THE OZONE NAAQS

EPA-454/B-95-007

U.S. Environmental Protection Agency
Office of Air Quality Planning and Standards
Research Triangle Park, NC 27711

June 1996

DISCLAIMER

TABLE OF CONTENTS

EXECUTIVE SUMMARY

S.1 Purpose

The purpose of this document is to revise the modeled test for demonstrating attainment of the national ambient air quality standard (NAAQS) for ozone. We also refine guidance for selecting episodes to model so that greater consideration is given to an area's ozone design value and the severity of meteorological conditions accompanying observed exceedances.

The recommended revisions make the modeled attainment test more closely reflect the form of the NAAQS. Like the NAAQS, the test now permits occasional exceedances at any location. The recommended revisions are also intended to take account of uncertainties inherent in available models and in estimating future emissions. These uncertainties have become better appreciated since intensive efforts to apply the Urban Airshed Model (UAM) for regulatory purposes began in 1991.

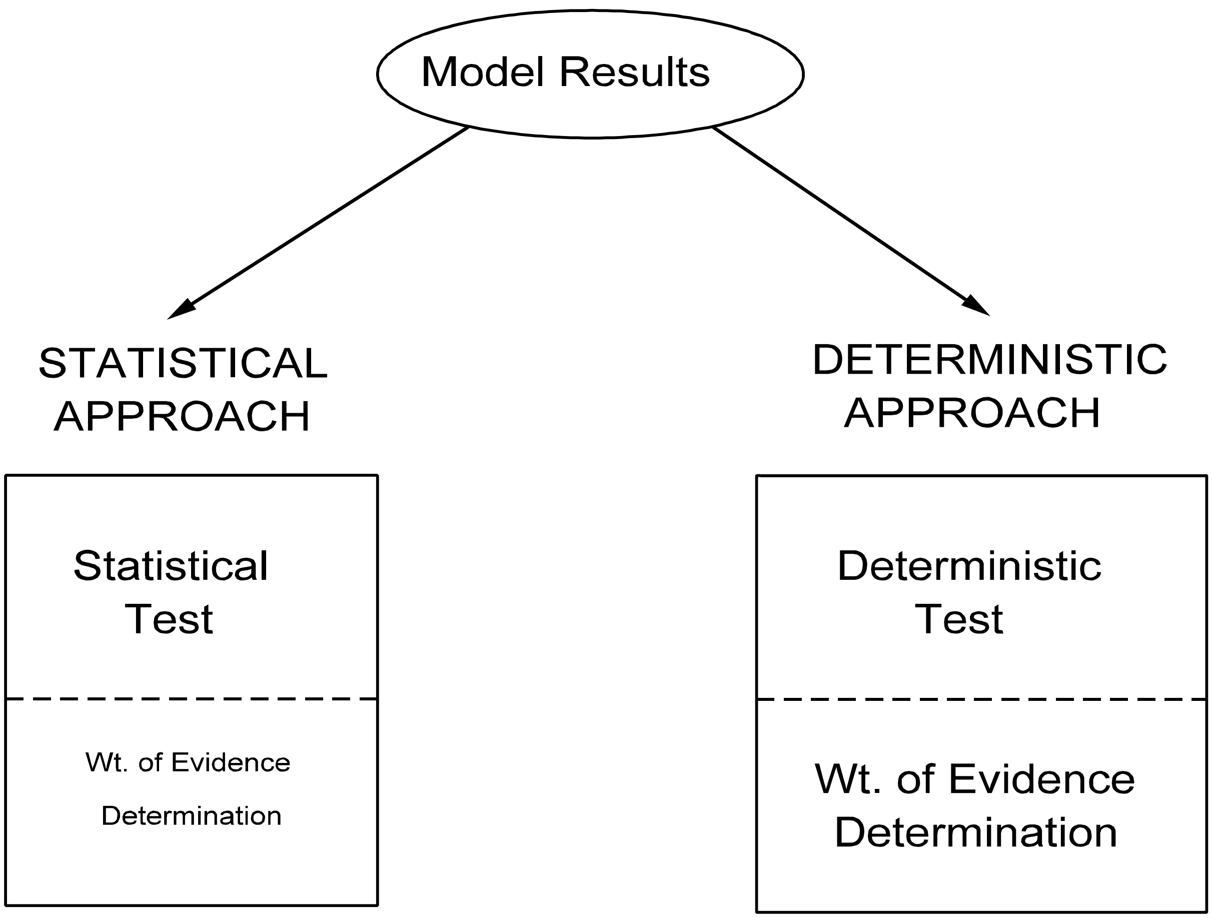
S.2 Overview

Two acceptable approaches are identified for demonstrating attainment of the ozone NAAQS. These are shown in Figure S.1. The first of these is called the "Statistical Approach". This approach contains a "Statistical Test" and a weight of evidence determination. We refer to the second acceptable approach as the "Deterministic Approach". This approach also consists of two parts: a "Deterministic Test" and a weight of evidence determination.

If the test in the selected approach is not passed, there is an option to perform a weight of evidence determination using additional information, such as air quality data. If this leads to compelling evidence that attainment is likely, attainment is demonstrated. In a weight of evidence determination, model results are weighed heavily. The further results are from passing the test, the more difficult it is to develop compelling supplementary evidence that attainment is likely.

Projecting ozone concentrations several years into the future has attendant uncertainties. Uncertainty about growth, location of new sources and effectiveness of future technologies are major components of uncertainty about model predictions. A technically viable attainment demonstration for "severe" and "extreme" nonattainment areas should include provision for at least one mid-course review of air quality, emissions and modeled data. A second review, shortly before the statutory attainment date, is also required. To make the mid-course review as insightful as possible, we recommend that the attainment demonstration include projections to some intermediate year (e.g., ~1999-2000). When combined with a mid-course review of updated air quality,

FIGURE S.1
ACCEPTABLE APPROACHES FOR
DEMONSTRATING ATTAINMENT
OF THE NAAQS FOR OZONE



emissions and modeled data, this should prove helpful in assessing whether refinements are needed in the current control strategy.

Figure S.2 shows the 3-stage process (current review, mid-course review and review at the statutory date) we require. Note that as the attainment date nears, increased reliance on observed data is anticipated. This results from anticipated improvements in the data base as well as from use of shorter projection periods.

S.3 The Statistical Approach for Demonstrating Attainment

The Statistical Approach consists of a test and an optional weight of evidence determination. The statistical test has three parts.

(1) It allows up to 3 exceedances at every location, depending on severity of modeled episodes. Exceedances are only allowed on episode days which are "severe".

(2) It limits the magnitude of each allowed exceedance. Limits depend on severity of each episode, and are calculated so as to be consistent with observed ozone patterns at sites currently attaining the NAAQS.

(3) If the model underpredicts observed ozone, the test requires at least an 80% reduction in the predicted incidence of ozone greater than 124 ppb.

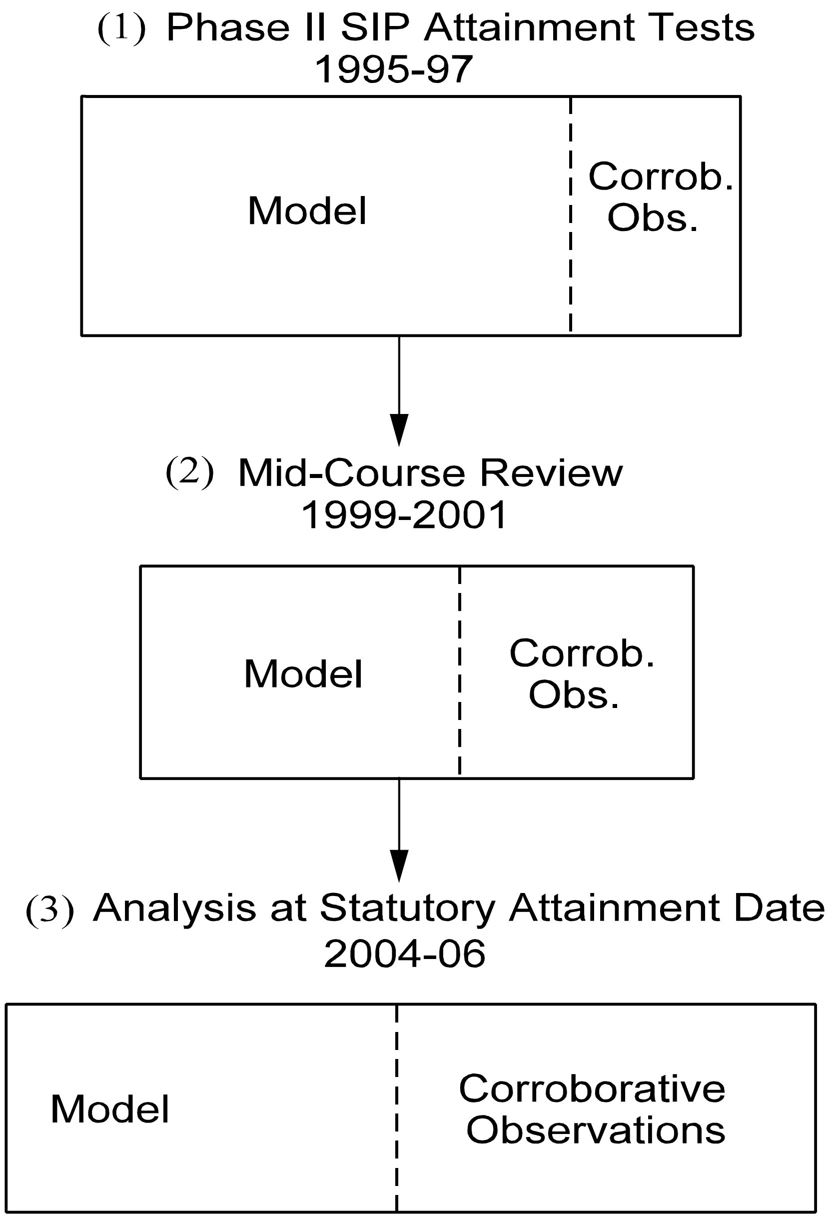
The preceding test depends critically on our ability to rank severity of days selected for modeling. We have developed a default procedure for doing this. In the default procedure, days are ranked using a regression equation relating highest daily maximum ozone concentration to several meteorological variables. If the regression equation is unable to explain at least **65%** of the observed variation in the highest daily maxima, the Statistical Approach should not be used.

A weight of evidence determination is included as a second, optional part of the Statistical Approach. This determination entails use of supplementary analyses to determine whether attainment is likely despite model results which do not pass the statistical test. The further results are from passing the test, the more difficult it becomes to demonstrate attainment through use of a weight of evidence determination. The weight of evidence concept is described more fully in Section S.4.

S.4 The Deterministic Approach for Demonstrating Attainment

The Deterministic Approach consists of a test plus an optional weight of evidence determination. The deterministic test is

FIGURE S.2
MULTI-STAGE NATURE OF SIP DATA ANALYSIS



passed if **daily maximum concentrations predicted in every surface grid cell are ≤ 124 ppb for all primary episode days**. A primary episode day is generally every modeled day except for the first day of each episode.

A weight of evidence determination may be undertaken to demonstrate attainment despite results which do not pass the deterministic test. As with the statistical test, the further results are from passing the deterministic test, the more difficult it becomes to demonstrate attainment through use of a weight of evidence determination. However, the deterministic test is more conservative than the statistical test. Thus, the burden of proof which weight of evidence from supplementary analyses needs to overcome is less in the Deterministic Approach.

In a weight of evidence determination, each of the analyses shown in Table S.1 should be considered and the results should be documented. The degree to which each additional analysis is pursued is decided on a case by case basis. Documentation accompanying weight of evidence results should include explanations of why excluded analyses were not pursued. The middle column in Table S.1 outlines factors which would lead one to assign greater weight to the indicated analysis. The right-hand column describes outcomes which would lend support to concluding that attainment is demonstrated, even though model results do not quite pass the established test.

Table S.1. Factors Affecting Weight of Evidence and Acceptance of Model Results Nearly Passing the Attainment Test

Type of Analysis	Factors Increasing Weight of Evidence	Factors Supporting Deviation from Test Benchmark(s)
Photochemical Grid Model	-good performance -extensive data base -short projection period -confidence in inventories & projections	-overpredictions -major improvement in predicted AQ using a variety of indicators -results come very close to meeting the benchmark(s) -other peer-reviewed grid models predict comparable or better improvement in ozone
Trend Data	-extensive monitoring network -precursor & ozone trends avail. -statistical model normalizing trend explains much variance -little bias in statistically predicted highest ozone -short projection period -pronounced, stat. significant normalized trend -continued, comparable relative reductions in emissions provided for	-pronounced downward normalized trend exceeding that anticipated with grid model

Observational Models	-extensive monitoring network -QA'd, self-consistent results -plausible, physical explanations for findings	-indicates sources other than those in modeled strategies play significant roles
Selected Episodes	-all met.regimes corresponding w. high obs. O3 considered -met.ozone potential of episodes exceeded ~ 1/year	-observed O3 >> design value -Severity of met. conditions expected to be exceeded << 1/yr
Incremental Costs/Benefits	-good documentation for cost estimates -lack of alternatives for reducing emissions -lack of model responsiveness for variety of strategies as benchmark is approached	-lack of model responsiveness accompanied by high incremental costs
Other (optional) Analyses	-rationale documented	

1.0 INTRODUCTION

1.1 Purpose

This document updates portions of the Guideline for Regulatory Application of the Urban Airshed Model (U.S.EPA, 1991) so that guidance better reflects experience gained in model applications since 1991. The updated guidance supports ongoing revisions to State implementation plans (SIP's) to meet the national ambient air quality standard (NAAQS) for ozone. More specifically, guidance described herein will be used along with the 1991 Guideline to implement phase II of the SIP revisions described in a March 2, 1995 U.S. EPA policy memorandum (Nichols, 1995).

Changes described herein focus on the modeled attainment test assessing whether a proposed control scenario will likely lead to attainment of the NAAQS by statutory dates. Thus, guidance described in Section 6.4 of the 1991 Guideline is superseded. We also refine earlier guidance on episode selection (Section 3.1 and Appendix B of the 1991 Guideline). By adding these latter refinements, we do not mean to imply that States should consider replacement episodes in their ongoing SIP revisions. However, if a State elects to simulate a new episode, we recommend that efforts be made to follow the refined guidance.

1.2 Background

As States and the EPA gained experience applying the Urban Airshed Model (UAM), several things became apparent. First, photochemical grid models require a great deal of information. Much of this information is uncertain. Further, model formulation reflects limits imposed by existing scientific knowledge as well as by computational necessities. Uncertainties in model inputs and limitations in model formulation lead to uncertainties in model predictions. This implies that a revised modeled attainment test should provide means to take better account of uncertainty.

A second finding from recent model applications is that controls estimated as necessary to attain the NAAQS can be very high. Despite such estimates, monitored ozone data reflect downward trends in many areas over the past 10 years (U.S.EPA, 1994). Monitored data are the definitive means for classifying an area's attainment status. This has led to a number of redesignations from "nonattainment" to "attainment" status. It has also led to concerns among some that the existing modeled attainment test may be too conservative.

The monitored attainment test allows 1.0 expected exceedance per year of a daily maximum ozone concentration of 124 ppb at all monitoring sites. The previous modeled test required estimated daily maxima to be 120 ppb or less in all surface grid cells for all selected primary episode days. The apparent contrast between the two tests causes concern that the modeled test may lead one to prescribe controls beyond those necessary to pass the monitored test. The guidance described herein provides an opportunity to more closely replicate the monitored test. This is done by considering severity of selected episodes more explicitly and allowing modeled exceedances on “severe” days (i.e., days having meteorological conditions which are unusually conducive to high ozone formation or transport).

2.0 OVERVIEW OF THE ATTAINMENT DEMONSTRATION PROCESS

In the preceding section, we noted two goals for a revised attainment test: (1) it should provide a means to consider uncertainty, and (2) it should replicate the monitored test more closely by allowing an occasional modeled exceedance. These two goals must be consistent with the overriding purpose of an attainment demonstration: to provide a reasonable expectation that the measures and procedures outlined will result in attainment of the NAAQS by statutory dates.

We identify two approaches for meeting these goals. Both approaches need to include documentation that the episodes considered include days with high observed ozone concentrations. Both procedures also include provisions for performing one or more periodic reassessments of observed and predicted air quality data prior to the statutory attainment date. This is to ensure that additional control measures can be invoked before the statutory date should a need be identified later. We refer to these reassessments as "mid-course reviews".

The first acceptable approach is called the "Statistical Approach". It includes a test which allows modeled exceedances on up to 3 days at different locations, depending on the severity of the selected episode days. In addition to limits on allowable number of modeled exceedances, the test includes a limit on the magnitude of daily maximum concentrations exceeding 124 ppb and a minimum required reduction in the occurrence of modeled values in excess of 124 ppb. Thus, the test includes three benchmarks. The first of these limits the number of days with allowed exceedances. The second restricts the magnitude of an allowed exceedance. The third benchmark requires a minimum level of improvement in air quality to be exceeded.

The Statistical Approach uses ranked severity of episode days in a quantitative manner. Ranking severity of episode days introduces additional uncertainties. If the ranking scheme fails to meet defined performance standards, the Statistical Approach may not be used. In this case, it is necessary to revert to the second acceptable approach (i.e., the “Deterministic Approach”, to be discussed later). The Deterministic Approach may consider severity of selected episodes also. However, it does so more qualitatively, in concert with other analyses, in a weight of evidence determination.

If one or more of the statistical test's benchmarks is failed, a weight of evidence determination may be performed using corroborative information. If the corroborative information is consistent with the likelihood that a proposed strategy will lead to attainment of the NAAQS by statutory dates, attainment has been demonstrated. The further model results are from passing the test's benchmarks, the more difficult it is to show that a strategy is, nevertheless, adequate in a weight of evidence determination.

We call the second acceptable approach for demonstrating attainment the "Deterministic Approach". This approach consists of a deterministic test and an optional weight of evidence determination. The deterministic test is passed if predicted daily maximum ozone concentrations are $\leq$ 124 ppb in all surface grid cells on all modeled primary episode days. Thus, the test contains a single benchmark. Exceptions may be considered if modeled exceedances are few, and are likely attributable to an artifact introduced by the model. In addition, if the test is not passed, a weight of evidence determination may be used to show that attainment of the NAAQS is still likely. Because the deterministic test is more conservative than the statistical test, the burden of proof needed in a weight of evidence determination to permit small deviations from the test’s benchmark is less than in the case for the statistical test.

The remainder of this document is organized as follows. In Section 3.0, we offer some supplementary guidance on episode selection which can be used with information in Section 3.1 and Appendix B of Guideline for Regulatory Application of the Urban Airshed Model (U.S.EPA, 1991). In Section 4.0, we describe the Statistical Approach for demonstrating attainment. In Section 5.0, we discuss the Deterministic Approach for demonstrating attainment. Because weight of evidence determinations are potentially more important in the Deterministic Approach, use of weight of evidence is also described in Section 5.0. Section 6.0 addresses the multi-stage nature of attainment assessments. Section 7.0 summarizes key points in this guidance. Finally, we include several appendices. Appendix A contains a complete example illustrating use of this guidance with a set of hypothetical model results. Appendix B is a glossary of new or frequently used terms in this document. Appendix C presents issues related to use of this guidance. These issues are discussed in question/answer format.

3.0 EPISODE SELECTION

In this section, we recommend procedures for selecting new episodes to model. Selection of new episodes for modeling is optional. The procedure we recommend differs from that described in Section 3.1 and Appendix B of the 1991 Guideline in only minor respects. The principal differences are: (1) we now recommend that areas strive to model days with observed daily maxima close to the design value for the most severely classified nonattainment area being modeled; and (2) ranked severity of meteorological conditions accompanying the air quality observations should be considered as a factor in choosing the episode days, where reliable rankings exist. The issue of reliability and performance standards for ranking procedures is discussed further in Section 4.3. The goal of the episode selection process is to choose episodes containing days with observations near but slightly above the design value and meteorological ozone forming potential likely to be exceeded about once per year. Other factors affecting choice of episodes remain as stated in the 1991 Guideline.

The following step by step example illustrates the recommended procedure.

Step 1. Identify Distinctive Meteorological Regimes and List Days Within Each Regime According to the Observed Highest Daily Maximum Ozone Concentration

This step is identical to the current procedure. The primary means of distinguishing meteorological regimes is the source/receptor orientation implied by the windfield. Table 3.1 illustrates the results of this step.

Table 3.1. Example Identifying Candidate Days For Modeling

Met.Regime 1	**Met.Regime 2**	**Met.Regime 3**
C_{11}	C_{12}	C_{13}
C_{21}	C_{22}	C_{23}
C_{31}	C_{32}	No more obs. above 124 ppb
C_{41}	C_{42}	-
C_{51}	No more obs. above 124 ppb	-
C_{61}	-	-
No more obs. above 124 ppb	-	-

where C_{12} is the day with the highest observed daily maximum ozone concentration occurring with meteorological regime 2.

Step 2. Use Data Compiled by the EPA (or Calculated Using an Approved Alternate Procedure) to Assign a Ranking in the Meteorological Ozone Forming Potential for Each Day Identified in Step 1.

Using a default ranking procedure, we have compiled ranked lists of ozone forming potential for days during the ozone season for each day over a 41- year period of record for nearly every modeled area. The ranking is based on meteorological ozone forming potential using a regression model developed by Cox and Chu (Cox and Chu, 1993; Cox and Chu, 1996). Thus, this effort should be minimal for the States if the default procedure is used for the rankings. One merely identifies the assigned rank for each of the days listed in Step 1. Use of alternate ranking procedures is discussed further in Section 4.3.

As a result of Step 2, Table 3.1 becomes Table 3.2.

Table 3.2. Candidate Days with Associated Ranked Ozone Forming Potentials

Met. Regime 1	**Met. Regime 2**	**Met. Regime 3**
C_{11}=190 ppb (1)	C_{12}=162 ppb (33)	C_{13}=145 ppb (150)
C_{21}=177 ppb (5)	C_{22}=155 ppb (57)	C_{23}=135 ppb (55)
C_{31}=167 ppb (79)	C_{32}=134 ppb (110)	-
C_{41}=145 ppb (12)	C_{42}=129 ppb (104)	-
C_{51}=141 ppb (32)	-	-
C_{61}=141 ppb (45)	-	-
-	-	-

Concentrations shown in the table are the highest daily maximum ozone concentration observed by any monitoring site in or downwind of the non-attainment area on the indicated day. Thus, the highest observations noted in the table could well occur at different monitoring sites. Each number in parentheses is the ranking of the corresponding day's meteorological ozone forming potential.

Step 3. Estimate Rank of the Meteorological Ozone Forming Potential Associated with One Observed Exceedance per Year.

Suppose meteorological ozone forming potential has been evaluated for days spanning a 41-year period of record. In this case, the meteorological ozone forming potential associated with days ranked at least as high as "41" (i.e., the ranked number ≤ "41") would be expected to be exceeded < once/year.

Step 4. Select from Among Days in the Table Generated in Step 2 Considering the Design Value and Ranking Data.

One would proceed down each column until an observed ozone concentration approaches the design value and/or a ranking equivalent to 1 expected exceedance (ExEx) per year of the estimated meteorological ozone forming potential is found. For reasons which will be elaborated on in Section 4.3, rankings between "20" and "83" are judged to be equivalent to 1 ExEx with a 41-year period of record. Days which are too extreme could be discarded. For days with comparably ranked meteorological ozone forming potentials, preference is given to days having observed concentrations closest to the design value. If the ranking scheme performs poorly or marginally, preference should be given to selecting days with observations above, but approaching, the design value. (This assumes no other overriding issue like presence of an intensive data base or poor model performance applies--see Section 3.1 in the 1991 Guideline).

Using Table 3.2, let us illustrate the process for selecting candidate episode days. We begin by moving down column 1 for Meteorological Regime 1. In this example, we will assume a design value of 160 ppb. Days C_{11}, C_{21} and C_{41} are probably too extreme. Such meteorological ozone forming potentials would occur with frequencies of once in 41 years, 8 years and ~3.5 years respectively. Days C_{31}, C_{51} and C_{61} have equivalent rankings. Day C_{31} is the choice for Meteorological Regime 1, because it is closest to and slightly above the design value. Going through a similar thought process, Day C_{12} is selected from Meteorological Regime 2. Meteorological Regime 3 presents an interesting case. Here, one day (C_{23}) has a ranking approximately equivalent to 1 ExEx, but also has an observed daily maximum well below the design value. The other day (C_{13}) has a daily maximum closer to the design value, but with an episode less extreme than 1 ExEx. In conflicts such as this, we recommend going with the day closest to the design value. This recommendation stems from the uncertainties attendant with the rankings. (to be discussed in Section 4.3). Thus, we choose Day C_{13} for Meteorological Regime 3. Summarizing, using our recommended guidance for episode selection, we have chosen episodes containing days C_{31}, C_{12} and C_{13} for modeling in this example.

4.0 APPROACH 1: STATISTICAL APPROACH

The Statistical Approach includes a modeled test in which three benchmarks should be passed. First, the number of days with predicted exceedances in defined locations should not be greater than a specified number. The number depends on the severity of the modeled episode days. Second, for episode days in which modeled exceedances are allowed, predicted daily maxima should not exceed a certain value. This value depends on the severity of the selected episode as well as the shape of distributions of observed daily maxima at sites which currently just attain the NAAQS. Third, for each day with an allowed exceedance, improvement in the number of hourly occurrences with predicted ozone greater than 124 ppb should be at least 80%. This third requirement may be waived on any day where the model does not underpredict observed ozone.

Ranking each episode day is a critical part of the Statistical Approach. The default procedure used to develop the rankings for each day's meteorological ozone forming potential is described more fully in Cox and Chu (1993) and in Cox and Chu (1996). We summarize key points of the procedure in Section 4.3. In large modeling domains containing several Metropolitan Statistical Areas (MSA's) for which ranked data are available, which of these data to use becomes an issue. This issue is discussed in Appendix C.

The Statistical Approach allows use of additional analyses to justify deviations from the test's benchmarks. Weight given to these additional analyses depends on the performance of the photochemical grid model, confidence in its inputs and confidence in the procedure for ranking episodes, as well as on the strength of the additional analyses. The three benchmarks and each of these additional analyses are discussed in Sections 4.1-4.2.

4.1 The Statistical Test

4.1.1 Benchmark 1

Benchmark 1: For a Composite of All Episodes, The Number Of Primary Episode Days Where A Modeled Exceedance Occurs Within Any Defined Subregion Of The Model Domain Should Be $\leq$ 3 Or $\leq$ (N-1), Whichever Is Less.

"N" is the number of "severe" primary episode days which are modeled. A day is considered "severe" if its associated "meteorological ozone forming potential" is expected to be

exceeded fewer than two times per year (i.e., ExEx < 2.0/year).

No modeled exceedances are permitted on primary episode days with meteorological ozone forming potential which is expected to be exceeded two or more times per year (i.e., ExEx $\geq$ 2.0/year).

Some clarifications are needed before we can illustrate use of benchmark 1 with an example. First, what do we mean by a **"defined subregion of the model domain"**, and why is this distinction needed? We define a **"subregion"** as an area which approximates a 15 km x 15 km area as closely as permitted by the model's horizontal grid resolution. This is illustrated in Table 4.1.

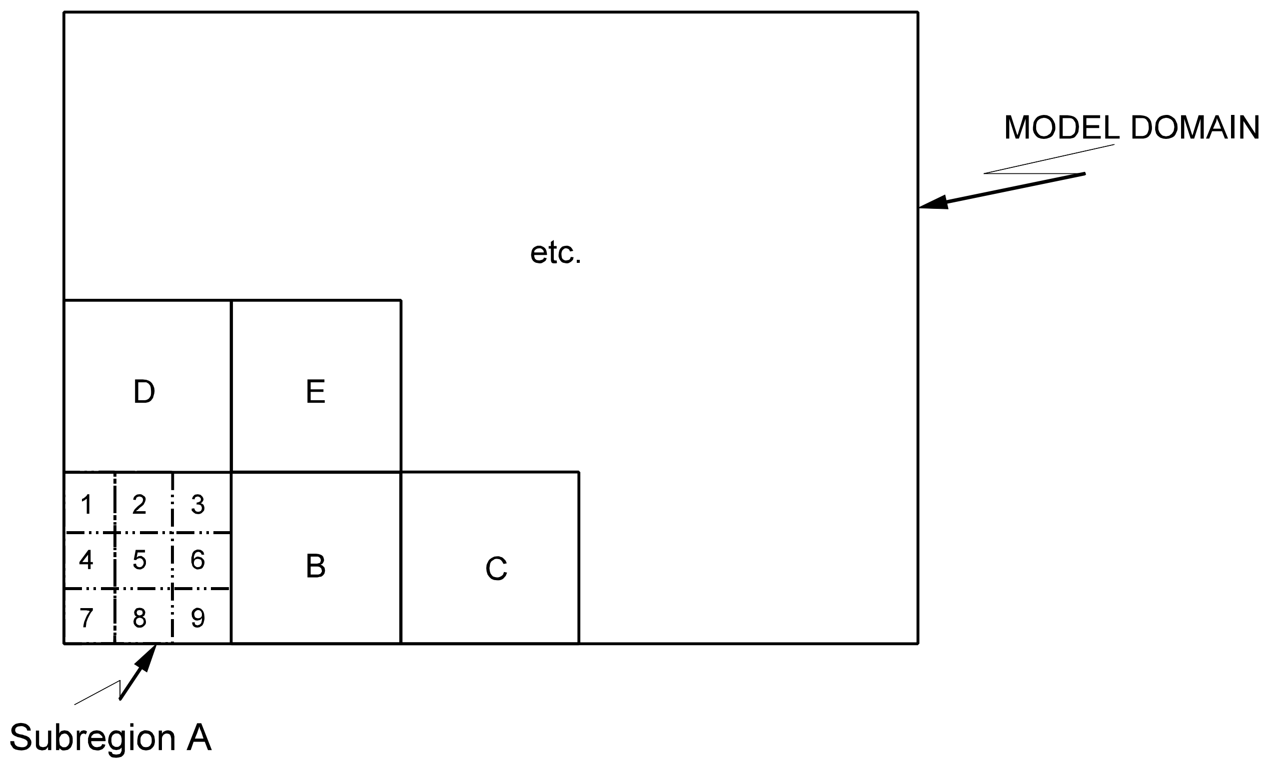
Table 4.1 Definition of Subregions for Grids with Differing Horizontal Resolution

Grid Resolution, km	**Grid Squares in Subregion**	**Area of Subregion, km^2**
2	8 x 8 = 64	256
3	5 x 5 = 25	225
4	4 x 4 = 16	256
5	3 x 3 = 9	225
8	2 x 2 = 4	256
12	1 x 1 = 1	144
16	1	256
18	1	324

We recommend that, for consistency, non-overlapping clusters of grid squares (defining the subregions) originate in the SW corner of the modeling domain. Exceptions can be made on a case by case basis, if warranted. The key requirement is that the subregions are defined consistently for each modeled day. Definition of subregions by clusters of grid squares is illustrated in Figure 4.1 for a grid having 5 km horizontal resolution. A more complete example illustrating identification of subregions is contained in Appendix A.

We believe it is necessary to define subregions within the modeling domain because of uncertainty attendant with wind fields used in photochemical grid models. Winds affect positioning of precursor and ozone plumes. There is no single standardized approach for specifying wind fields, and the exercise requires subjective judgment. Thus, latitude exists for adjusting the wind field within reasonable bounds of uncertainty so as to minimize the likelihood of having more than one predicted exceedance in individual grid cells of the size used in urban applications. For each such subregion, the maximum predicted daily maximum concentration is used in determining consistency of predictions with the benchmark. “Maximum” refers to the highest

FIGURE 4.1. DEFINING SUBREGIONS FOR USE IN STATISTICAL TEST



concentration from among those grid cells that comprise the given subregion. Defining and using subregions in this way provides a means to remain protective of the NAAQS, given uncertainty about the wind field.

The method used to equate a ranking with an expected exceedance frequency also deserves further explanation. Equivalency between a ranking and an ExEx is a function of the number of years for which the meteorological ozone forming potential corresponding with each day is calculated. In the set of default information we will provide to help implement this test, we have used a 41-year period of observed meteorological conditions. Thus, a day with a rank of "42" would be expected to have its meteorological ozone forming potential exceeded exactly once per year, on average. Similarly, a day with a rank of "83" would be expected to be exceeded 2.0 times/year, and a day ranked "21" would be expected to occur just slightly less frequently than 0.5 times/year (i.e., once every 2 years). If one were to use a set of regression equations applied to a 30-year period of record, rankings of "16", "31" and "61" would correspond to expected exceedance frequencies of 0.5, 1.0 and 2.0/year, respectively. In general, the relationship between a ranking and an expected yearly frequency of exceedance (ExEx) is given by:

ExEx = (Rank - 1)/(# of years used for ranking)

Thus, if a day were ranked 30th based on an examination of meteorological observations over a 41-year period, the day's corresponding expected exceedance frequency would be:

ExEx = (30 - 1)/41 = 0.7 times per year.

Examples

Problem 1. A State models 3 episodes containing 6 primary episode days. Modeled exceedances occur in only 1 subregion. These occur on 3 of the days, with meteorological ozone forming potential values expected to be exceeded 0.4, 1.2, and 1.8 times per year. The three days without modeled exceedances in this subregion have ExEx's of 1.5, 1.95 and 8.7 times per year. Is benchmark 1 passed?

First, we see whether there are any subregions in which modeled exceedances occur on more than 3 days. (Since the number of severe primary episode days (N) is "5", the requirement that the number of days with modeled exceedances be no more than "3" is the most restrictive requirement.) There are no subregions having more than 3 days of modeled exceedances.

Finally, we check whether there are any days having an ExEx as frequent as 2.0 times/year or more on which a modeled exceedance occurs. There are none. The benchmark is met.

Problem 2. Suppose conditions are identical to those in problem 1, except now there is also a second subregion in which modeled exceedances occur on the day with 0.4 expected exceedances per year and on the day with 1.95 expected exceedances per year. Does the area still pass the benchmark?

First, we note that there are still no subregions in which modeled exceedances occur on more than 3 days. Note that this milestone is passed, even though there are now 4 days in which a modeled exceedance occurs somewhere in the modeling domain.

Finally, there are still no days having an expected exceedance frequency of 2.0/year or more on which there is a modeled exceedance.

Thus, the area still passes the benchmark.

4.1.2 Benchmark 2

Benchmark 2: Predicted Daily Maxima Corresponding With Each Allowed Modeled Exceedance May Not Be Greater Than A Concentration Derived From A Distribution Of Observed Daily Maxima At Sites Currently Just Attaining The NAAQS.

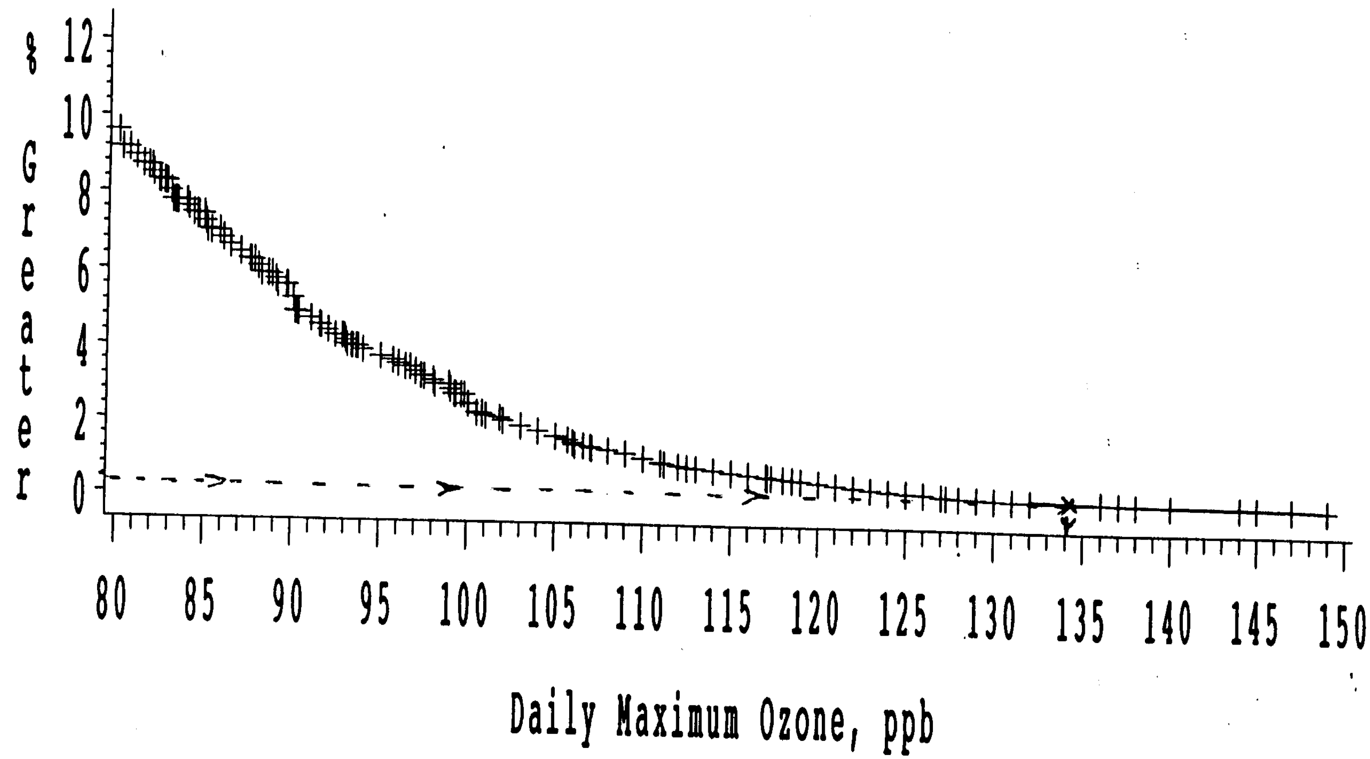
This benchmark also needs further explanation before we can illustrate its use. We have examined the EPA AIRS data base for 1991-1993 to identify ozone monitoring sites which are currently "just attaining" the ozone NAAQS. By "just attaining", we mean that 2-3 exceedances are observed at the site during a 3-year period, and the ozone design value is $\geq$ 115 but $\leq$ 124 ppb. There are 60 sites meeting our criteria for "just attaining". These occur in 21 States at urban, suburban and rural locations, and in areas where the predominant landuse is characterized as "commercial", "agricultural", "industrial", "residential", "forest" and "mobile". Thus, our data include a diverse set of sites. Despite this diversity, we find that the shape of the tail of the distribution of daily maxima is very similar at these sites. This enables us to fit a distribution curve to a large pooled data base whose most extreme observations have an associated probability of exceedance which is very small. Figure 4.2 shows the result of this fitting.

The ability to create the distribution of daily maxima described in the preceding paragraph is important. We are interested in knowing what the distribution of daily maxima might look like at other sites after sufficient controls are implemented to attain the NAAQS. In particular, we are interested in what the distribution should look like at sites which will just attain the NAAQS. Such sites will have exceedances of 124 ppb. We want to know whether the magnitude of a modeled exceedance is consistent with the expected shape of the distribution of daily maxima at a

FIGURE 4,2

Cumulative Frequency

1991 - 1993 Attainment Sites



large number of sites which are already just attaining the NAAQS. This can be estimated as follows:

(1) estimate the probability that the meteorological ozone forming potential is exceeded for each day on which an exceedance is allowed;

(2) for each such day, enter the ordinate in Figure 4.2 at the indicated probability and read off the corresponding daily maximum ozone concentration.

For example, suppose the meteorological ozone forming potential for a modeled day is ranked as the 15th highest for a 41-year period of analysis. The probability that this potential is exceeded is:

$$p = (15/41) \times (1/365) = .001002$$

Thus, there is a 0.1% likelihood that the meteorological ozone forming potential associated with this day is exceeded. Entering Figure 4.2, we see that a daily maximum ozone concentration of 134 ppb has an associated probability of 0.1% of being exceeded. Putting this another way, a daily maximum concentration of 134 ppb on such an extreme day is consistent with meeting the NAAQS.

Calculation of probabilities and use of Figure 4.2 is cumbersome. Table 4.2 may be used to test whether benchmark 2 is met. The columns in the table correspond to the period of record used to rank the ozone forming potential for each day. This information is provided for the convenience of those utilizing ranking procedures other than the default one reflected by the Cox/Chu statistical model. The information is needed if a period of record differing from the default one of 41 years is used with these alternate approaches. Those using the default rankings which we provide should use the column labeled "41" in the table.

Table 4.2. Acceptable Upper Limits for Daily Maximum Ozone (ppb) on Very Severe Days with Allowed Modeled Exceedances

Yrs of Rec Day Rank	10	15	20	25	30	35	**41**	45	50
1	148	151	152	155	158	160	**161**	161	163
2	140	144	148	150	151	151	**152**	153	155
3	137	140	141	145	148	150	**151**	151	151
4	133	137	140	141	144	146	**148**	150	150
5	130	136	138	140	140	143	**145**	146	148
6		133	137	138	140	140	**142**	144	145
7		131	135	137	138	140	**140**	141	143
8		130	133	136	137	139	**140**	140	141
9			132	135	137	138	**139**	140	140
10			130	133	136	137	**138**	139	140
11				132	134	137	**137**	138	139
12				131	133	136	**137**	137	138
13				130	132	134	**137**	137	137
14					131	133	**136**	137	137
15					130	132	**134**	136	137
16						131	**133**	135	136
17						130	**132**	134	136
18							**131**	133	135
19	130						**131**	132	133
20	124						**131**	132	133
21							**130**	131	132
22								130	132
23									131

24									131
25	124	130	130	130	130	130	**130**	130	130

Table 4.2 contains several interesting features. First, a day whose meteorological ozone forming potential is expected to be exceeded at a specified rate (e.g., 0.5 ExEx/year) has an associated daily maximum ozone concentration which remains constant, regardless of the number of years used to create the rankings. This follows from the manner in which the ExEx rate is calculated (ExEx/yr = (rank - 1)/(# of years in data base)). In the table, an ExEx rate of 0.5/year always corresponds to a daily maximum observation of 130 ppb. We use this value as the maximum value permitted for an exceedance corresponding to an ExEx rate between 0.5 and 2.0/year. For reasons identified in Section 4.3, we consider this range equivalent to an ExEx of 1.0/year, given uncertainties in the ranking procedure. **Thus, to pass benchmark 2, a modeled exceedance corresponding to a day ranked such that 0.5/year < ExEx < 2.0/year should not exceed 130 ppb.**

Days ranked as severe or more severe than 0.5 ExEx/year may have modeled exceedances greater than 130 ppb. The acceptable limits are presented in Table 4.2. **Thus, Table 4.2 should be used to establish the acceptable upper bound for a modeled exceedance on very severe episode days.**

Example

<u>Problem 3:</u> Suppose the maximum modeled exceedances on the 4 days with exceedances in Problem 2 are as shown below. Assume ozone forming potential rankings are based on a 41-year record.

<u>Day</u>	**<u>ExEx</u>**	**<u>Corresponding Rank</u>**	**<u>Modeled Max. Daily Maximum O3, ppb</u>**
1	0.4	16	138
2	1.2	49	130
3	1.8	74	126
4	1.95	80	128

Is benchmark 2 met?

First, we note that days 2, 3 and 4 fall in the range where 0.5 <

ExEx < 2.0/year. Benchmark 2 requires modeled daily maxima not to exceed 130 ppb for these days. This milestone is met.

Next, we observe that day 1 is more severe. Entering Table 4.2 in the column labeled "41", we see that the maximum allowed value for the 16th ranked day is only 133 ppb.

Thus, benchmark 2 is not met, since 138 ppb > 133 ppb.

4.1.3 Benchmark 3

Benchmark 3: For Each Day With An Allowed Modeled Exceedance, The Number Of Daytime Hourly Exceedances Of 124 PPB Predicted Throughout The Domain Should Be Reduced By At Least 80%.

This benchmark is included to provide a safeguard against cases where photochemical grid model predictions meet EPA performance criteria but, nevertheless, tend to underpredict observed concentrations. It could be possible in some cases like this to pass the first two benchmarks with few control measures. The type of example we are most concerned about is one in which concentrations in the order of 150 ppb are observed on a day with high ozone forming potential, yet predictions are in the order of 130-135 ppb. We believe a benchmark like this one is necessary to provide assurance that a control strategy passing the statistical test is likely to lead to significant improvements in air quality on occasions when observed concentrations are high.

Unlike benchmarks 1 and 2, benchmark 3 applies to predictions in individual surface grid cells rather than in subregions. It also applies to all daytime hourly predictions (i.e., between 8 AM and 8 PM Local Standard Time (LST)) rather than just daily maxima. The benchmark is passed if the predicted number of "grid cell-hours" with predictions > 124 ppb is reduced by at least 80% on each day for which a modeled exceedance is allowed by benchmark 1. The calculation is made as follows:

$$\frac{\mathrm{SUM}[(\#\text{ of cells})_{base}] - \mathrm{SUM}[(\#\text{ of cells})_{control}]}{\mathrm{SUM}[(\#\text{ of cells})_{base}]} \geq 0.80$$

where each term in the above expression is summed over the 12 hours between 8 AM and 8 PM LST on each day for which a modeled exceedance is allowed. The subscript “base” in the above expression refers to current base case conditions. That is, if a 1991 episode were being simulated, then the required 80% improvement is relative to the 1991 model predictions used in the

UAM performance evaluation.

Since benchmark 3 is included to provide protection in cases where the model underpredicts observed ozone concentrations, it may be waived for any given day if underpredictions do not occur on that day.

For each day with an allowed exceedance, we recommend comparing the maximum peak modeled and maximum peak observed concentrations throughout the domain to determine whether the model is underpredicting observed peak ozone. If the test results do not indicate an underprediction of more than 5% for a given day, benchmark 3 need not be applied on that day.

4.1.4 Summary of the Statistical Test

The statistical test is passed if three benchmarks are passed.

1. **Limits on Number of Modeled Exceedances**. No subregion in the domain may contain modeled exceedances on more than 3 or "N-1" days, whichever is less. "N" is the number of severe days (i.e., days with ExEx < 2,0/year) which are modeled. No exceedances are permitted on a modeled day unless it is severe enough so that its ozone forming potential would be expected to be exceeded fewer than 2 times/year.

2. **Limits on Modeled Concentrations for Allowed Exceedances**. Predicted daily maximum ozone concentrations may not exceed 130 ppb on days with allowed exceedances unless the severity of the episode day is expected to be exceeded $\leq$ 0.5 times per year. For such very severe episodes daily maximum ozone predictions are limited as shown in Table 4.2.

3. **Required Minimum Level of Improvement**. Reduction in the number of (surface) grid cell-hours exceeding 124 ppb during 8 AM to 8 PM local standard time should be at least 80% for each day on which a modeled exceedance is allowed. This benchmark can be waived for a given day if the model does not underpredict observed ozone concentrations.

Appendix A contains a detailed numerical example illustrating application of the statistical test.

4.2 Weight of Evidence Determination

The statistical test described in Section 4.1 already includes substantial effort to consider corroborative evidence. That is, the test considers severity of the modeled episodes and utilizes information developed from observed frequency distributions of monitored daily maxima for ozone. Thus, additional corroborative analyses used in weight of evidence determinations need to be compelling to permit deviations from the benchmarks in the statistical test. The weight of evidence determination provides additional information to enable those reviewing an attainment demonstration to conclude that attainment of the NAAQS is probable even though minor exceptions to the benchmarks exist.

4.2.1 Model Performance and Results

General model performance and focused model performance may be considered in a weight of evidence determination. If the general performance measures (see Appendix C, 1991 Guideline) indicate that model predictions of ozone are not biased low, this supports the notion that small deviations from one or more of the benchmarks may, nevertheless, be consistent with attainment.

Focused model performance evaluation examines ozone production and the model's treatment of meteorological factors affecting ozone predictions in a small number of grid cells. If it can be shown that a high prediction most likely reflects a model artifact rather than chemistry or physics of the atmosphere and the prediction in question affects the outcome of the test, the prediction can be disregarded. Focused performance evaluation is described more fully in Section 5.2.

If model results indicate major improvement in predicted ozone and the episodes considered are very severe (i.e., ExEx $\leq$ 0.5/year), this may also support a small deviation from a benchmark.

4.2.2 Transport

Transport of ozone and precursors into a modeling domain can be an important factor affecting ozone predictions obtained with the UAM and with other photochemical grid models. Thus, transport affects the photochemical grid model's results. As illustrated in Figure S.1 (executive summary), the Statistical and Deterministic Attainment Approaches are applied after model results are obtained. That is, the attainment tests compare model results, obtained with a specified emissions scenario, transport and other assumptions, with the NAAQS. Although it is certainly true that transport can affect the outcome of a test, it does not necessarily have to be a part of the test.

As noted in the preceding paragraph, the primary impact of transport is to affect model results, which are then compared to the NAAQS using a modeled attainment test. There are two ways in which the modeled attainment test itself could more directly reflect transport. First, an area-specific ranking scheme could be developed which contains variables which represent regional transport more directly than is done in the Cox/Chu default procedure which we have used. Some suggestions on how to do this are offered in Appendix C. Days classified as "severe" (i.e., ExEx < 2.0 days per year) using such a scheme would be allowed modeled exceedances. A second way for considering transport in the selected attainment approach is as a part of a weight of evidence determination. This is described further in Section 5.3

and in Appendix C.

Finally, if it is shown that overwhelming transport influences modeled exceedances in all or in a portion of the modeling domain, a policy decision could be made that the modeled attainment tests need only be applied to a portion rather than to the entire domain. Such a decision would need to consider factors outside the scope of this guidance.

4.2.3 Other Analyses

Other analyses need to be considered in a weight of evidence determination to sway conclusions about the adequacy of a control scenario for which the benchmarks are nearly met. These include the following: use of normalized trend data, use of results from observational or other models and consideration of incremental cost/benefit estimates. These additional analyses are described in greater detail when we discuss the Deterministic Approach (see Section 5.3). We believe weight of evidence determinations are of greater importance in this latter approach. Factors affecting weight given to various pieces of evidence and outcomes which are consistent with deviating from one or more of the benchmarks are generally consistent with what is shown in Table 5.1.

4.3 Further Discussion of the Statistical Test

4.3.1 Derivation of the Default Ranking Procedure

The default procedure we have used to rank meteorological ozone forming potential of each day is based on a statistical model by Cox and Chu (Cox and Chu, 1993; Cox and Chu, 1996). To develop this model, Cox/Chu considered the highest daily maximum ozone concentration observed among all monitoring sites each day in each of 32 Metropolitan Statistical Areas/Consolidated Metropolitan Statistical Areas (MSA/CMSA's) and immediate surroundings during 1983-1993. A number of meteorological variables thought to make some contribution to observed ozone were considered for use in regression equations to explain the observed variation in the highest daily maxima. Most important meteorological variables were daily maximum surface temperature, morning average wind speed and direction, afternoon average wind speed and direction, midday relative humidity, total opaque cloud cover, and Julian day (a surrogate for solar intensity). When these variables plus an additional variable depicting underlying trends in ozone during 1983-93 were incorporated into a linear expression used as a scaling factor in a Weibull distribution, good agreement was found with the observed distribution of maxima during the 10-year period. The meteorological variables alone are successful in leading to a good characterization of the variability in observed maxima. With a few exceptions, the equations containing meteorological variables only are able to explain between 60 and 80% of the variance in observed daily maxima. Coefficients for the variables in the linear regression equation vary from city to city.

A linear equation including only meteorological variables was then used in a similar fashion to predict "meteorological ozone forming potential" for each day in the ozone season for each of the 32 MSA's/CMSA's during a 41-year period (1953-1993). We assumed that days outside the ozone season generally have lower ozone forming potential, which is exceeded by most days within the season. Using this procedure, it was possible to rank the ozone forming potential of each day for each MSA/CMSA during the 41-year period.

4.3.2 Sensitivity of the Ranking Procedure

We have performed some sensitivity tests to determine how rankings of days might be affected by changes in the regression equations underlying estimates of meteorological ozone forming potential. On the basis of this, we have concluded rankings of 41 $\pm$ 25 to be approximately equivalent to a 95% level of confidence, with the distribution of error skewed such that larger uncertainty exists for larger ranked numbers (i.e., less severe days).

Thus, one finding from our sensitivity analysis of the estimated rankings is that they become more sensitive to changes in the statistical model as one moves away from the tail of the ranked distribution. This occurs, because as one moves away from the tail, there are increasingly small differences in estimated meteorological ozone forming potential between successively ranked days. Thus, a change in the statistical model leads to occasional large differences in a day's ranking, even though the change in its calculated meteorological ozone forming potential is minuscule. This finding suggests that, the further one moves away from the tail of the ranked distribution, one should give increasingly greater weight to the observed ozone in selecting episodes. For rankings above ~100 (for a 41-year record), ranked data may have broad associated uncertainty. Because differences in calculated meteorological ozone forming potential are greater as one approaches the tail, rankings for the most extreme episodes are likely to be most stable.

We use the preceding findings as the basis for our recommendation that rankings corresponding to ExEx frequencies between 0.5 and 2.0 times per year be considered essentially equivalent. That is, 41 $\pm$ 25 is approximately equivalent to this range.

4.3.3 Performance of Ranking Procedures

We believe that for ranked meteorological ozone forming potential to be used in as quantitative a procedure as the statistical test, there should be a minimum performance requirement for the ranking procedure. We recommend that the variation in the observed highest daily maxima which is explained by the calculated ozone forming potential values should be at least 65% (i.e., adjusted $R^2 \geq 0.65$). This requires that the correlation between observed and predicted daily maxima be greater than 0.8. If this criterion is not met, we recommend that the Deterministic Approach be used in the attainment demonstration. Methods and benchmarks described for the statistical test could be used more qualitatively as part of a weight of evidence determination in the Deterministic Approach.

4.3.4 Use of Alternative Ranking Procedures

As we have previously noted, we are suggesting the Cox/Chu equations as a default procedure for deriving means for ranking the severity of episode days. States may use alternative procedures. Indeed, in some locations we have been unable to meet the performance criterion recommended in Section 4.3.3 using the Cox/Chu equations. Thus, if one wishes to use the Statistical Approach, it will be necessary to develop a better ranking procedure for these areas.

To be accepted for use, an alternative ranking scheme should meet the following criteria:

1) there should be a logical physical/chemical explanation relating to ozone formation and/or transport for a variable to be included in a statistical model;

2) adding an additional independent variable to a statistical model should lead to a statistically significant improvement in the model's performance;

3) ability of a model to explain variance in the observed daily maxima should exceed that of the default procedure; and

4) the adjusted R^2 associated with a model should explain at least 65% of the variance in the observed highest daily maxima.

5.0 APPROACH 2: DETERMINISTIC APPROACH

5.1 The Deterministic Test

After a proposed strategy is modeled, the benchmark in this test is to compare the predicted daily maximum ozone concentration in each surface grid cell with 124 ppb on each of the primary episode days. **If all predicted daily maxima are ≤ 124 ppb, the test is passed.**

By "primary episode day", we mean all days except for the first day of each period selected to model. The first day is called a "ramp-up" day, and is excluded from the test because of dependence of initialization assumptions on poorly known or absent data. Thus, if 3 episodes were selected having 3 days, 2 days and 2 days respectively, all daily maxima predicted in surface grid cells on each of 4 (2 + 1 + 1) primary episode days should be ≤ 124 ppb.

5.2 Exception Through Use of a Focused Model Performance Evaluation

It may happen that the benchmark in this test is very nearly passed, except for isolated modeled daily maxima which exceed 124 ppb. In such cases, attainment can still be satisfactorily demonstrated using this approach. First, let us define what we mean by "isolated". This is subjective, but generally we mean no more than ~2-3 surface grid cells on any primary episode day.

If the predicted exceedances are isolated, then one needs to make the argument that each instance likely reflects inability of the model to properly characterize physical/chemical processes accompanying the modeled exceedance. If the argument is successfully made, the modeled exceedance(s) should be disregarded. This argument is made by reviewing available meteorological and air quality data and comparing temporal and spatial patterns with those predicted by the model during the base case and/or projection period for the incident in question. A relatively new procedure, called "process analysis", is recommended as one means for analyzing model predictions for the purpose of these comparisons whenever feasible (Jeffries et al., 1994, Tonnesen et al., 1994, Jang et al., 1994).

The exception to the test's benchmark that predicted daily maxima be ≤ 124 ppb should not be extended to cases in which there are numerous predicted daily maxima exceeding 124 ppb. The 1991 Guideline provides means for rejecting episodes for which model performance is generally poor. These performance criteria are robust, in that they look at many comparisons between observed and predicted values. We assume that these performance tests provide protection against selecting episodes in which formation of high ozone is widely mischaracterized. It would be quite possible however, for the model to mischaracterize ozone in a few locations and still pass these performance tests. The exceptions to meeting the benchmark in the deterministic test are intended to provide a remedy for this concern.

5.3 Weight of Evidence Determination

If the attainment test is not passed and exceedances cannot be explained as model artifacts, the Deterministic Approach allows use of a weight of evidence determination to assess whether attainment is, nevertheless, likely. The deterministic test is more conservative than the statistical test. This means that control strategies leading to model results which come close to meeting the deterministic test's benchmark are more likely to result in attainment than would be true for results which almost meet the statistical test's benchmarks. Thus, the burden of proof required from corroborative analyses is less for a weight of evidence determination to conclude that results which nearly pass the deterministic test are adequate.

A weight of evidence determination includes a subjective assessment of the confidence one has in the modeled results. This is supplemented with a review of available corroborative information, such as air quality data. The more extensive and credible the corroborative information, the greater influence it could have in permitting deviations from the deterministic test's benchmark. Thus, areas having extensive, monitored information have a greater potential for justifying deviations from the benchmark. In the following subsections, we identify factors affecting weight given to modeled results. We then identify several corroborative analyses to consider in a weight of evidence determination. Each of the types of analyses shown in Table 5.1 should be considered. Arguments supporting conclusions reached in the weight of evidence determination need to be documented. Reasons for not pursuing a particular type of evidence rigorously also need to be documented.

5.3.1 Factors Affecting Confidence in Modeled Results

Weight of evidence given to model results depends on the following factors:

-model performance;

-confidence in the underlying data bases;

-length of the projection period;

-how close results come to meeting the test's benchmark.

The longer the projection period, more sparse the data base and the poorer and less comprehensive the model performance evaluation, the lower the "weight of evidence" given to the grid modeling results relative to the other factors considered in a weight of evidence determination. Generally, the closer results come to meeting the test's benchmark, the less compelling other evidence supporting a deviation from the benchmark needs to be. In addition, model results showing major improvement in predicted ozone with a variety of indicators could be used to support the notion that attainment is likely. Results from other peer reviewed grid models, which have replicated observations well, may also be used to support a weight of evidence determination.

5.3.2 Severity of Modeled Episodes

The more extreme the days selected for modeling, the greater the "weight of evidence" that a control strategy leading to model results nearly meeting the benchmark is sufficient to demonstrate attainment. We have discussed methods for ranking episode severity at length in Sections 4.1 and 4.3. As noted in Section 4.3, if performance of a statistical model underlying a ranking procedure is insufficient, it is not advisable to use this information quantitatively. However, it could still be used qualitatively, in concert with other information, to support deviations from the deterministic test's benchmark. The better the performance of the statistical model underlying the ranking and the more severely ranked a day in a selected episode, the greater the support for deviating from the benchmark. It is not necessary to perform the analyses described in Section 4.0 in a weight of evidence determination. Generally however, the more extensive the analysis of episode severity, the greater the weight of evidence which can be assigned.

If it is not feasible to rank meteorological ozone forming potential as discussed in Section 4.0, it is still possible to gain a qualitative sense of an episode's severity through review of air quality data. If there are several daily maxima in the area which are greater than the area's design value, this provides qualitative information supporting a deviation from the deterministic test's benchmark. We emphasize the importance of looking at more than a single monitoring site. A high value at a single site could reflect an upset or unusual emission event at a source or small group of sources. This does not necessarily imply that the accompanying meteorological conditions are severe. A further precaution is needed in assessing severity of historical episodes on the basis of air quality observations. Lower concentrations in recent years may, in part, reflect the results of emission reductions. Thus, higher concentrations in earlier years do not necessarily imply unusually severe meteorology.

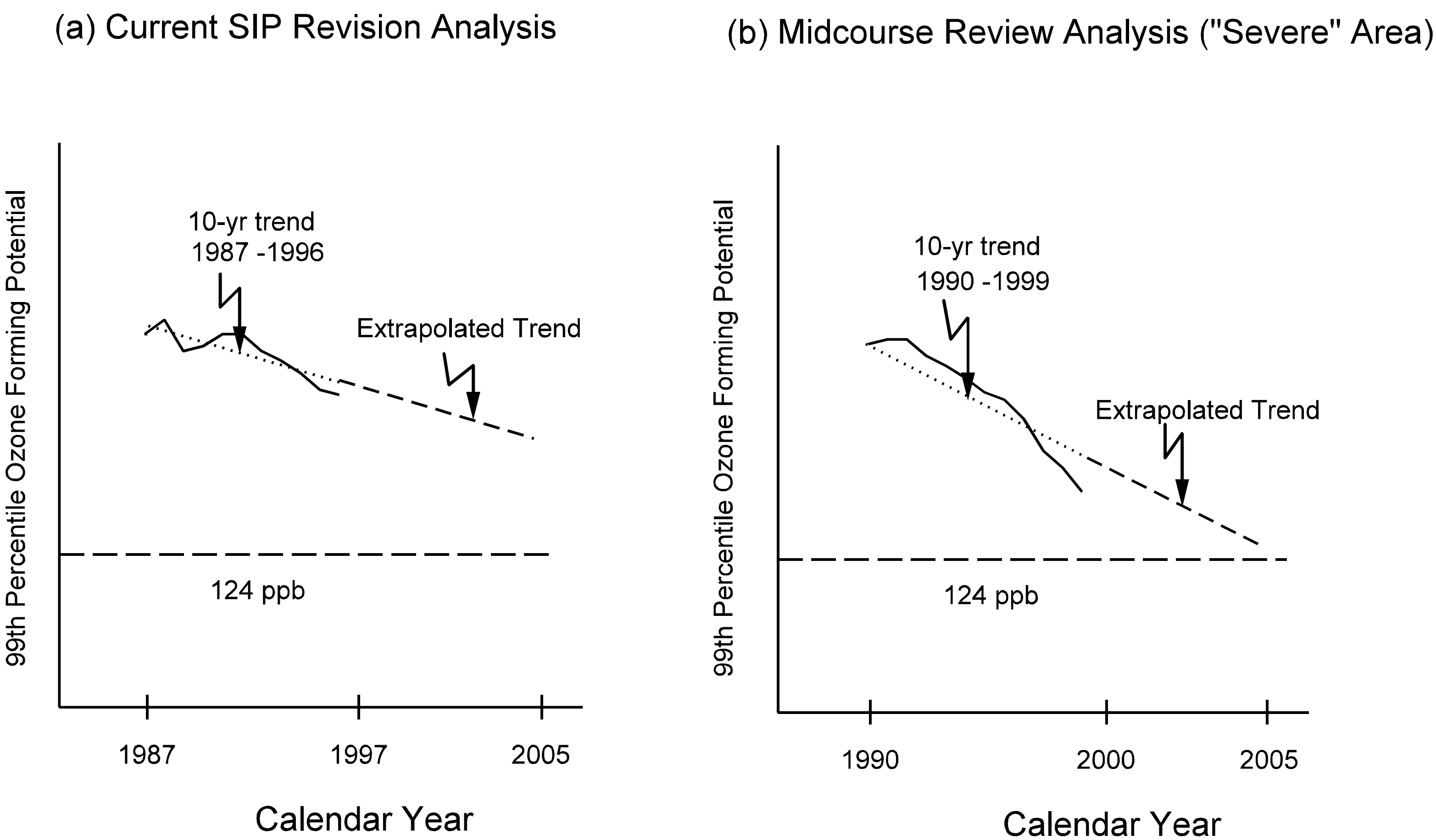
5.3.3 Trend Analysis

This analysis entails normalizing trend parameters (e.g, 99th percentile daily maxima) for meteorological differences observed year to year. The model described in Cox and Chu (1993) (or a similar type model which is more skillful in explaining adjusted variance in observed daily maximum ozone concentrations) can be used to normalize observed raw trends. We calculate normalized trends periodically for serious and above nonattainment areas with the Cox/Chu procedure. Recent normalized trends are presented in U.S.EPA (1994) and in U.S.EPA (1994a).

Figure 5.1 illustrates how the resulting information might be used in a weight of evidence determination. Figure 5.1(a) represents an analysis performed as part of a phase II demonstration (Nichols (1995)). The procedure calls for extrapolating the most recent 10-year normalized trend line (i.e., 1987-96) in 99th percentile daily maxima to 124 ppb. If the 124 ppb concentration is reached prior to the statutory attainment date and relative emission reductions at least as great as those in the preceding 10 years are anticipated, one might use this as supporting justification for allowing small deviations from the deterministic test's benchmark. The argument for permitting deviations from the benchmark would be strengthened if concurrent reductions in ambient precursor levels were observed and the regulations for further reductions were in place. The trend analysis could be repeated for a mid-course review (Figure 5.1(b)) and again at the statutory attainment date.

The weight given to evidence produced by the trend analysis is a function of several factors. The more air quality monitoring data available, the more confident we can be that potential exceedances of 124 ppb are being measured. Therefore, it follows that locations having large numbers of monitoring sites which measure ozone and precursors could give greater weight to trend analysis than could other locations with less extensive data bases. Other factors which favor giving a large weight to air quality trends include: (a) a high degree of skill (i.e., adjusted R^2) in the ability of the statistical model (used to normalize the data) to explain observed variance in daily maxima; (b) lack of bias in the statistical model's predictions of the highest observed daily maxima; (c) a short projection period to the statutory attainment date; (d) a suitable normalized trend parameter which approaches 124 ppb; (e) presence of a pronounced normalized trend line over the preceding 10-year period, and (f) regulations in place requiring further emission reductions. Presence of a statistically significant downward trend exceeding that expected from the photochemical grid modeling results, would support deviation from a benchmark.

FIGURE 5.1. EXAMPLE TREND ANALYSIS IN ATTAINMENT DEMONSTRATIONS



5.3.4 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 observational approaches currently under investigation within the research community: use of indicator species (Trainer, et al., 1993, Milford, et al., 1994, and Sillman, 1995), the relative incremental reactivity (RIR) approach (Cardelino and Chameides, 1995), the smog produced algorithm approach (Blanchard and Roth, 1994) and receptor models (Lewis, et al., 1993, and Henry, et al., 1994).

Currently, the potential for observational models to quantify precursor reductions needed to demonstrate attainment is not clear. In the near term, we believe the principal value of these methods is to identify control directions (VOC, NOx, both) and source categories making major contributions to ozone formation. Thus, their role in a weight of evidence determination is to provide means for corroborating whether a control strategy identified in a photochemical grid modeling analysis is addressing key contributors to observed high ozone. Thus, they may be particularly valuable if used in concert with an incremental cost/benefit analysis (see Section 5.3.5). Weight given to results from observational models depends on suitability and comprehensiveness of the available monitoring network, consistency of results with physical/chemical understanding of ozone formation and transport, and presence of several analyses which complement one another and appear to be consistent. If these techniques confirm a strategy, the "weight of evidence" they provide supports a more aggressive approach (i.e., trying to come closer to meeting the benchmark). If the results are contradictory, they may support a position that controlling certain emissions further in pursuit of the benchmark should be postponed.

5.3.5 Considering Incremental Costs and Benefits

When used in concert with other supporting information, we believe this is a legitimate consideration in assessing whether additional efforts are warranted to meet the deterministic test's benchmark. By "incremental" costs/benefits, we mean the difference in costs and the difference in benefits which accompany the current strategy being considered vs. a strategy which comes closer to meeting the benchmark. We recommend that incremental benefits be estimated by computing total dosage in the domain above 40 ppb with and without the incremental control measures in question. Our understanding is that there is no apparent lower threshold for effects from exposure to ozone. Our suggestion to use "40" ppb rather than "0" ppb stems from a commonly used value for continental background and from a desire to better highlight differences between strategies. Dosage is best weighted by the spatial distribution of population. Thus, units would look something like, "person-ppb-hours".

The incremental cost/benefit test we suggest is a qualitative rather than quantitative one. If small incremental benefits are accompanied by large incremental costs, this supports not immediately pursuing this particular strategy to come closer to passing the benchmark. We must emphasize that, due to the qualitative nature of these kind of calculations, we are not suggesting 1:1 tradeoffs between costs and benefits. Rather, we suggest that if the model predictions appear to be relatively unresponsive to additional controls, resulting in large incremental costs, it may be appropriate to conclude that model results are close enough to the benchmark, given other corroborative evidence.

5.3.6 Other Optional Analyses

Other types of analyses, in addition to those described, may be used to support a weight of evidence determination. The rationale underlying use of such an analysis, results obtained and how these results support or do not support a conclusion that attainment is likely need to be documented.

5.3.7 Transport

See Section 4.2.2.

5.3.8 Summary, Weight of Evidence Determinations in the Deterministic Approach

Table 5.1 summarizes factors influencing "weight of evidence" determinations and their use in evaluating model results which nearly meet the benchmark(s) in the attainment test. The middle column of the table includes factors which would increase the weight given to evidence produced by the indicated analysis. The right-hand column describes outcomes which would contribute to evidence supporting a deviation from the modeled attainment test's benchmark(s).

Table 5.1. Factors Affecting Weight of Evidence and Acceptance of Model Results Nearly Meeting the Attainment Test's Benchmark(s)

Type of Analysis	Factors Increasing Weight of Evidence	Factors Supporting Deviation from the Benchmark
Photochemical Grid Model	-good performance -extensive data base -short projection period -confidence in inventories & projections	-overpredictions -major improvement in predicted AQ using a variety of indicators -results come very close to meeting the benchmark -other, peer reviewed grid models predict comparable or better improvement in ozone
Trend Data	-extensive monitoring network -precursor & ozone trends avail. -statistical model normalizing trend explains much variance -little bias in statistically predicted highest ozone -short projection period -pronounced, stat. significant normalized trend -continued, comparable relative reductions in emissions provided for	-pronounced downward normalized trend exceeding that anticipated with grid model

Observational Models	-extensive monitoring network -QA'd, self-consistent results -plausible, physical explanations for findings	-indicates sources other than those in modeled strategies play significant roles
Selected Episodes	-all met.regimes corresponding w. high obs. O3 considered -met.ozone potential of episodes exceeded ~ 1/year	-observed O3 >> design value -ExEx of met.ozone potential << 1/yr
Incremental Costs/Benefits	-good documentation for cost estimates -lack of alternatives for reducing emissions -lack of model responsiveness for variety of strategies as benchmark is approached	-lack of model responsiveness accompanied by high incremental costs
Other (Optional) Analyses	-rationale for each is documented	

5.4 Summary of Deterministic Approach

The Deterministic Approach includes a test with a benchmark requiring that **predicted daily maximum ozone in all surface grid cells on all primary episode days be $\leq$ 124 ppb.** A focused model performance evaluation may be used to exclude a limited number of modeled exceedances if it is shown that these likely result from an artifact of the model.

The Deterministic Approach includes an optional weight of evidence determination which may be undertaken to demonstrate attainment in the face of model results which do not meet the test's benchmark. The burden of proof which this determination must meet increases as photochemical grid modeling results deviate further from the benchmark.

A weight of evidence determination consists of several identified corroborative data analyses. Results from these analyses are used qualitatively to decide whether attainment of the NAAQS is likely even though the test's benchmark is not quite met.

6.0 MULTI-STAGE NATURE OF THE ATTAINMENT DEMONSTRATION PROCESS

In Sections 1.2 and 2.0, we noted that a major purpose of our redefinition of the modeled attainment test for ozone is to take better account of uncertainty in assessing whether a control scenario will be sufficient to meet the NAAQS. Thus far, we have addressed this goal by incorporating corroborative data either directly into a test (i.e., the statistical test) or by making use of these data in weight of evidence determinations (the Deterministic Approach and the Statistical Approach).

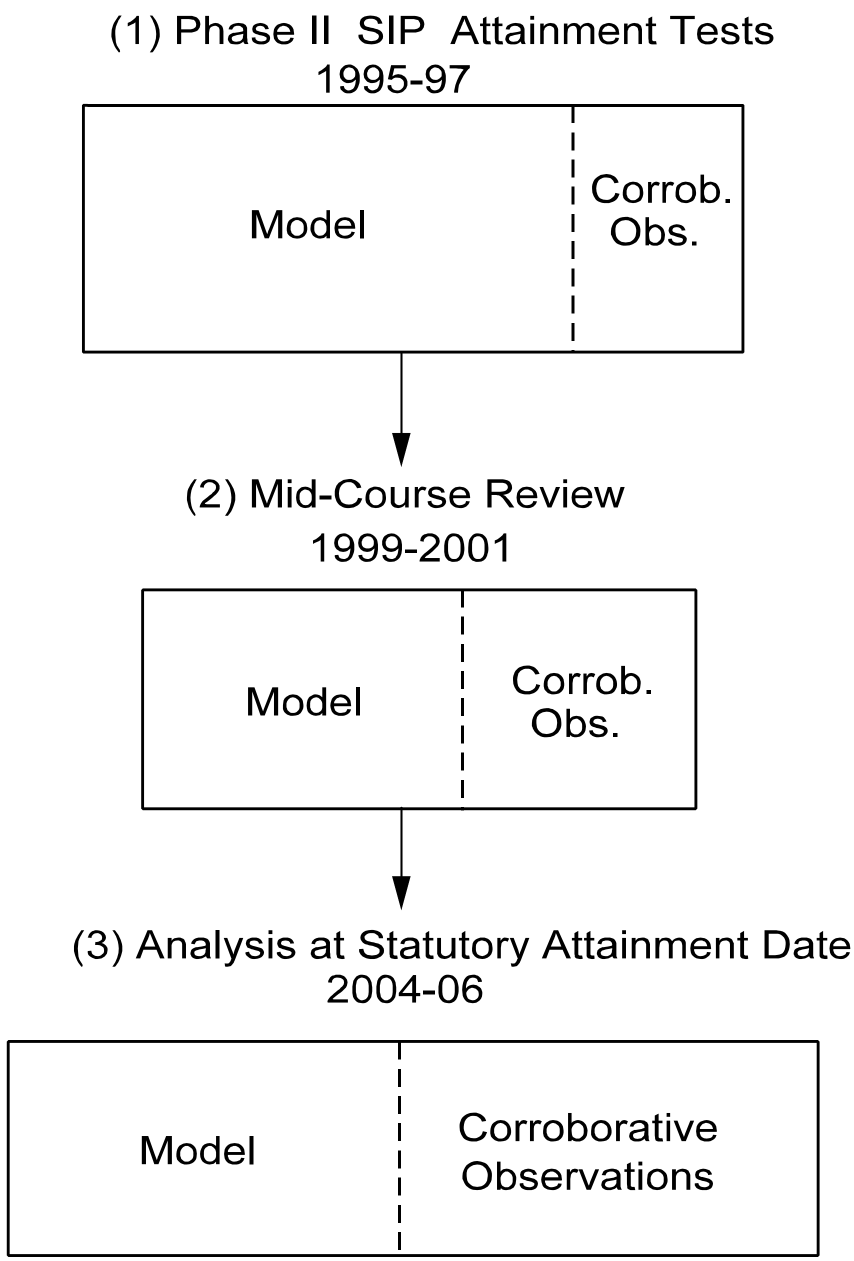
One of the most important causes of uncertainty however, is our inability to confidently predict what will happen 10-15 years into the future. Uncertainty in predicting economic growth, population changes and new technology causes major uncertainty in the estimates of projected emissions this far into the future. The assumptions made about growth can have pronounced effects on projected emissions and, therefore, on the levels of control needed to attain the NAAQS. This arises from the compounding effect over many years. To illustrate, the difference in emissions projected from 1990-2005 using a 3%/year growth rate as opposed to a 2%/year growth rate is 21%--a difference larger than the composite effect of several control measures. Inability to predict <u>where</u> changes will occur exacerbates this uncertainty.

Because of the uncertainty inherent in long term projections, we believe that a technically viable attainment demonstration needs to contain provisions for periodic reviews of the monitoring, modeling and inventory data to assess the extent to which refinements to the originally identified attainment strategy are needed. Such reviews could be coordinated in some manner with the Clean Air Act's required reasonable further progress assessments. At a minimum, we require that an acceptable attainment demonstration for a "severe" or "extreme" nonattainment area contain provisions for at least one “mid-course” review between completion of the phase II analyses and the statutory attainment date, as well as a review at or shortly before the required attainment date. We also strongly recommend that model applications in the phase II analyses make projections to the period to be considered in a mid-course review. This will permit more meaningful comparisons between observations at the mid-course review and the modeled projections. These comparisons, in turn, will support weight of evidence determinations performed as part of the mid-course reviews.

Thus, we require at least a 3-stage analysis: the current (1995-97) phase II analysis; a mid-course review (circa 1999-2001) and a third review at or shortly before the statutory attainment date for "severe" areas (circa 2004-2006). As described in Sections 4.0 and 5.0, the first stage of the analysis (i.e., the 1995-97 phase II analysis) consists of approaches which use photochemical grid modeling and a series of corroborative analyses with monitored data, statistical models and meteorological information. As the statutory date approaches, we anticipate greater reliance on observed data will be possible. The procedure we describe takes advantage of more reliable and extensive (e.g., PAMS) monitored air quality and meteorological data to corroborate predictions from past modeling and from improved modeling tools available in the future. Most importantly, projections are made over progressively shorter timeframes. Figure 6.1 presents a conceptual view of the required multi-stage attainment demonstration.

FIGURE 6.1

MULTI-STAGE NATURE OF SIP DATA ANALYSIS



7.0 SUMMARY

We have presented two approaches which are acceptable for demonstrating attainment of the NAAQS for ozone. These are identified as a "Statistical Approach" and a "Deterministic Approach". The approaches are designed to meet three objectives:

1. provide reasonable assurance that the control scenario which is selected will be sufficient to attain the NAAQS by statutory dates,

2. take account of uncertainty inherent in the data bases and in the available means for performing photochemical grid modeling, and

3. more closely replicate the monitored attainment test and, in so doing, more closely replicate the NAAQS which permits occasional exceedances at a number of locations.

The Statistical Approach includes a statistical test and a weight of evidence determination. Weight of evidence can be used to justify small deviations from the test's benchmarks. Because the test already incorporates some corroborative information and is less conservative than the test in the Deterministic Approach, burden of proof needed to deviate from benchmarks is greater than for the Deterministic Approach.

The statistical test contains three benchmarks. The first of these limits the number of days in which modeled exceedances are allowed at any given location within the model domain to three. This limit is reduced if three or fewer "severe" days are modeled. The second benchmark limits the magnitude of an allowed modeled exceedance. This limit is based on severity of modeled episode days, as well as on distributions of daily maximum ozone concentrations presently observed at monitoring sites where the NAAQS is attained. The third benchmark requires the number of occurrences in which hourly ozone predictions exceed 124 ppb to be reduced by at least 80%. This final benchmark is included as a safeguard in the event the photochemical grid model underpredicts observed ozone. It may be waived for any day on which no such underpredictions occur.

The statistical test depends critically on use of a procedure to rank severity of episode days. Rankings are based on a statistical model relating highest daily maxima to meteorological variables. If this statistical model cannot explain at least 65% of the variation in an area's highest daily maximum ozone concentration, the Statistical Approach may not be used to demonstrate attainment of the NAAQS.

We have refined existing guidance on episode selection to enable consideration of the ranked severity of candidate days. We have also recommended that greater priority be given to selecting days having observed ozone near the ozone design value. When multiple non-attainment areas are modeled, the design value for the most severe of these governs the choice of episodes. Weight given to the ranked values in episode selection also depends on the degree to which the statistical model underlying the rankings explains variance in the observed highest daily maximum ozone concentration.

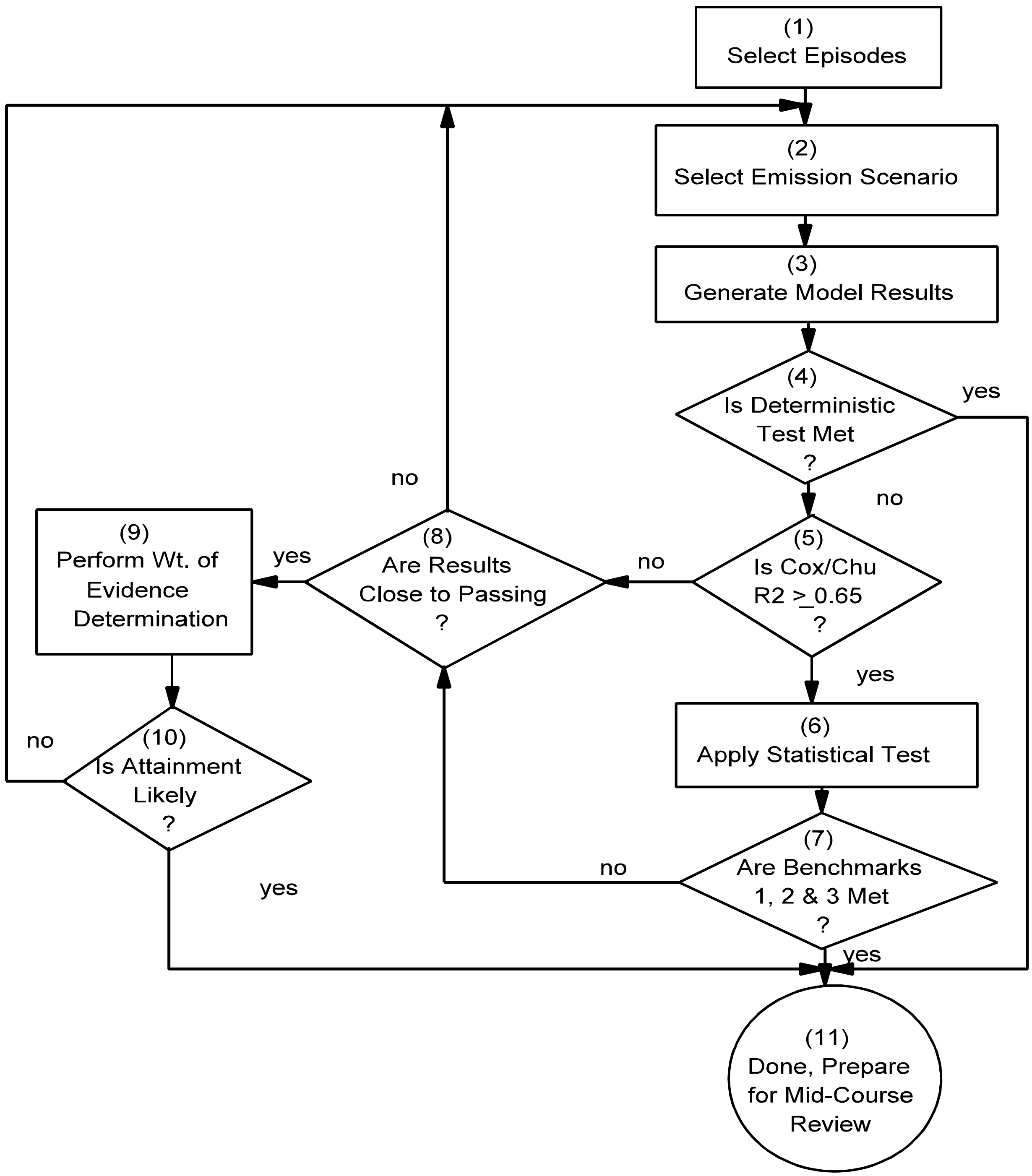
The second acceptable approach for demonstrating attainment is called the "Deterministic Approach". It consists of a deterministic test containing a single benchmark. The Deterministic Approach also includes a weight of evidence determination which may be used to see whether a deviation from the test's benchmark is warranted. The deterministic test's benchmark is met if there are no modeled exceedances of 124 ppb in any surface grid cell during any of the modeled primary episode days. A limited number of exceptions is allowed if a focused model performance evaluation shows that the modeled exceedance(s) is likely a result of an artifact introduced by the model rather than physics and chemistry of the atmosphere.

The Deterministic Approach allows greater latitude in using weight of evidence determinations to permit deviations from its test’s benchmark than is appropriate for deviations from benchmarks in the statistical test. This stems from the more conservative nature of the deterministic test and the greater likelihood that a strategy leading to results which nearly meet the benchmark will lead to attainment of the NAAQS. The weight of evidence determination is used to qualitatively support an argument that model results close to the benchmark demonstrate attainment.

Figure 7.1 summarizes the sequence of activities in the recommended approaches for demonstrating attainment. We believe that the two approaches we have recommended are consistent with the NAAQS and monitored attainment test in that they both permit occasional modeled exceedances. Consideration of episode severity in the choice of episodes is also consistent with the NAAQS.

Use of corroborative information either directly in a test or as part of a weight of evidence determination helps address uncertainty. However, a major (perhaps the major) source of uncertainty arises from projecting future events over long periods of time. Therefore, we have required that a technically viable attainment demonstration addressing the 2005 timeframe include provisions for at least one "mid-course" review of monitored, emission and modeling information. To make this review more insightful, we recommend that current model demonstrations include projections and predictions for the time period envisioned for this review.

FIGURE 7.1
FLOW CHART: USE OF ATTAINMENT APPROACHES



8.0 REFERENCES CITED

Blanchard, C.L. and P.M. Roth, (1994), Spatial Mapping of Preferred Strategies for Reducing Ambient Ozone Concentrations Nationwide, Prepared under EPA Contract 68D10154, Joseph Bufalini, EPA Project Officer.

Cardelino, C.A. and W.L. Chameides, (1995), "An Observation Based Model for Analyzing Ozone Precursor Relationships in the Urban Atmosphere", J.Air and Waste Management Association, 45, pp.161-181.

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.

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.

Jang, J., H.E. Jeffries, D. Byun and J.E. Pleim, (1994), "Sensitivity of Ozone to Model Grid Resolution: Part I. Application of High-Resolution Regional Acid Deposition Model", Submitted for Publication in Atmospheric Environment.

Jeffries, H.E. and S. Tonnesen, (1994), "A Comparison of Two Photochemical Reaction Mechanisms Using Mass Balance and Process Analysis", Atmospheric Environment, 28, pp.2991-3003.

Lewis, C.W., T.L. Conner, R.K. Stevens, J.F. Collins and R.C. Henry, (1993), "Receptor Modeling of Volatile Hydrocarbons Measured in the 1990 Atlanta Ozone Precursor Study", Paper 93-TP-58.04, 86th Annual AWMA Meeting, Denver, CO., June 1993.

Milford, J.B., D. Gao, S. Sillman, P. Blossey and A.G. Russell, (1994), "Total Reactive Nitrogen (Noy) as an Indicator of the Sensitivity of Ozone to Reductions in Hydrocarbon and Nox Emissions", J.Geophysical Research 99D, pp.3533-3542.

Nichols, M.D., (March 2, 1995) Memorandum to Regional

Administrators, USEPA Regions I-X, Subject: "Ozone Attainment Demonstrations".

Sillman, S., (1995), "The Use of Noy, H2O2 and HNO3 as Empirical Indicators for Ozone-NOx-ROG Sensitivity in Urban Locations", J.Geophysical Research, 100, pp.14175-14183.

Tonnesen, S. and H.E. Jeffries, (1994), "Inhibition of Odd Oxygen Production in the Carbon Bond Four and Generic Reaction Set Mechanisms", Atmospheric Environment, 28, pp.1339-1349.

Trainer, M., D.D. Parrish, M.P. Buhr, R.B. Norton, F.C. Fehsenfeld, K.G. Anlauf, J.W. Bottenheim, Y.Z. Tang, H.A. Wiebe, J.M. Roberts, R.L. Tanner, L. Newman, V.C. Bowersox, J.F. Meagher, K.J. Olszyna, M.O. Rogers, T. Wang, H. Berresheim, K.L. Demerjian and U.K. Roychowdhury, (1993), "Correlations of Ozone with Noy in Photochemically Aged Air", J.Geophysical Research 98, pp.2917-2925.

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

U.S. EPA, (1994), National Air Quality and Emissions Trends Report, 1993, EPA-450/R-94-026.

U.S. EPA, (1994a), Clean Air Act Ozone Design Value Study: Final Report, A Report to Congress, EPA-454/R-94-035.

APPENDIX A: EXAMPLE APPLICATION

DESCRIPTION

For the purpose of this exercise we plan to demonstrate modeled attainment for a nonattainment area being simulated by a UAM domain comprised of 20 rows and 20 columns of 5Km x 5Km grid cells. The model was run for three primary days. Figures A.1, A.2 and A.3 depict predicted daily maximum concentrations above 124 ppb for days 1, 2, and 3, respectively. The Cox/Chu rankings for the days are 55, 32, and 215, respectively. Only concentrations above 124 ppb are displayed. Grid cells are outlined with dashed lines and subregions are highlighted with bold lines. The highest daily maximum ozone predictions within the domain are 129, 130 and 122 ppb, respectively. The highest monitored air quality observed on those days was 145, 138 and 132 ppb. The design value for the area is 140 ppb and the maximum ozone air quality for the last three years was 138, 130, and 126 ppb, respectively. The model performance calculations for highest-prediction accuracy are -10%, 8%, and 3% for all three days, respectively. For days 1 and 2 the number of daylight grid cell hours above 124 ppb are 635 and 400 for the model performance runs. The attainment strategy reduces these numbers to 23 and 21.

APPLICATION OF DETERMINISTIC TEST

We first look at the daily maximum concentration for each grid cell on each primary day. Notice only day 3 has no concentrations above 124 ppb. The deterministic test's benchmark is not passed, because we have daily maximum concentrations above 124 ppb for days 1 and 2.

APPLICATION OF STATISTICAL TEST

To apply the statistical test we proceed as follows:

Benchmark 1 allows exceedances on days for which the expected exceedance frequency is less than two times per year (i.e., Cox/Chu ranking < 83 for a 41-year period of record). Results for Day 3 are acceptable, because even though the ranking is greater than 83 (i.e., 215) there are no predicted exceedances. The rankings for days 1 and 2 are less than 83. Therefore, benchmark 1 allows exceedances on one of these days in each of the indicated subregions (n=2, therefore n-1=1 for any subregion within the domain). Again, look closely at figures A.1 and A.2. Close inspection of the subregions reveals that no subregion had more than one day of exceedances. In other words, across both days, the exceedances do not overlap or occur within the same subregions. Therefore benchmark 1 is passed.

Benchmark 2 limits how high an exceedance concentration can be in order to be consistent with air quality monitored at attainment sites. According to Table 4.2 the acceptable upper limit for days 1 and 2 is 130 ppb (rankings for both days are above 21 and 41 years of data were used, so ExEx > 0.5 times per year). Exceedances on both days are at or below 130 ppb. Therefore benchmark 2 is passed.

Benchmark 3 requires us to look at the model performance statistics for underpredictions on days for which exceedances are allowed. The UAM guidance for regulatory application of UAM recommends the calculation of an unpaired highest-prediction accuracy (AU). The AU values for days 1, 2 and 3 are -10%, 8% and 3%, respectively. (Note: using the conventions described in the 1991 Guidance, a negative number represents an <u>over</u>prediction). Benchmark 3 requires that days with allowed exceedances and an AU value greater than 5% demonstrate an 80% improvement in the areal coverage above 124 ppb for daylight hours (i.e., 8:00 AM-8:00 PM, LST). Day 3 does not have to demonstrate 80% improvement because the AU is less than 5% and no exceedances were allowed. Day 1 is allowed exceedances but does not have to demonstrate 80% improvement because the AU is less than 5%. However, for day 2, exceedances are allowed and the AU model performance statistic is greater than 5%. Therefore, we add all daylight grid cell hours above 124 ppb for the model performance run on day 2 (400 grid cell hours). After growing emissions to the future attainment year and applying the attainment control strategy we again count the number of grid cell hours above 124 ppb (21 grid cell hours). This constitutes a 95% improvement in the air quality above 124 ppb (((400-21)/400)=.9475). Benchmark 3 is passed. For this exercise the statistical benchmarks are passed and the modeling satisfactorily demonstrates attainment of the ozone NAAQS.

USE OF WEIGHT OF EVIDENCE

For the above exercise the statistical benchmarks are passed. Suppose this had not been the case and a different outcome had prevailed. The following case is an example of how information obtained through the attainment test's benchmarks and other analysis may provide "weight of evidence" sufficient to demonstrate attainment. Through the introduction of additional information unique to this case, we plan to illustrate a case in which modeling plus additional analysis provides sufficient evidence that attainment is probable by the statutory date, even though one or more of the statistical test’s benchmarks are not passed.

Suppose the peak model prediction for day 1 was 135 ppb instead of 129 ppb. This is 5 ppb above the limit 130 ppb and benchmark 2 is failed. For this exercise assume benchmarks 1 and 3 are passed as previously presented. Looking more closely at day 1 we see our model performance indicates that the model overpredicted by 10%. There is a possibility that the future estimates are also overpredicted. 10% would imply as much as a 13 ppb (more than 5) fluctuation beyond 124 ppb. To further support our case we calculate the areal improvement. On day 1, 635 daylight grid cells are above 124 ppb. The future attainment strategy reduces this number to 23 grid cells. This represents a significant improvement in air quality (96%). We then refer back to the air quality observed on all three of our episode days, 145, 138 and 132, respectively. Not only is day 1's concentration the highest observed for all three episodes but it is greater than our air quality design value which is 140 ppb. Also, our current air quality for the last three years has shown a steady improvement (138, 130, 126 ppb, respectively). It is reasonable to believe that based on this information our control strategy will provide for attainment by the statutory attainment date.

Figure A.1: DAY 1; Cox/Chu Ranking=55

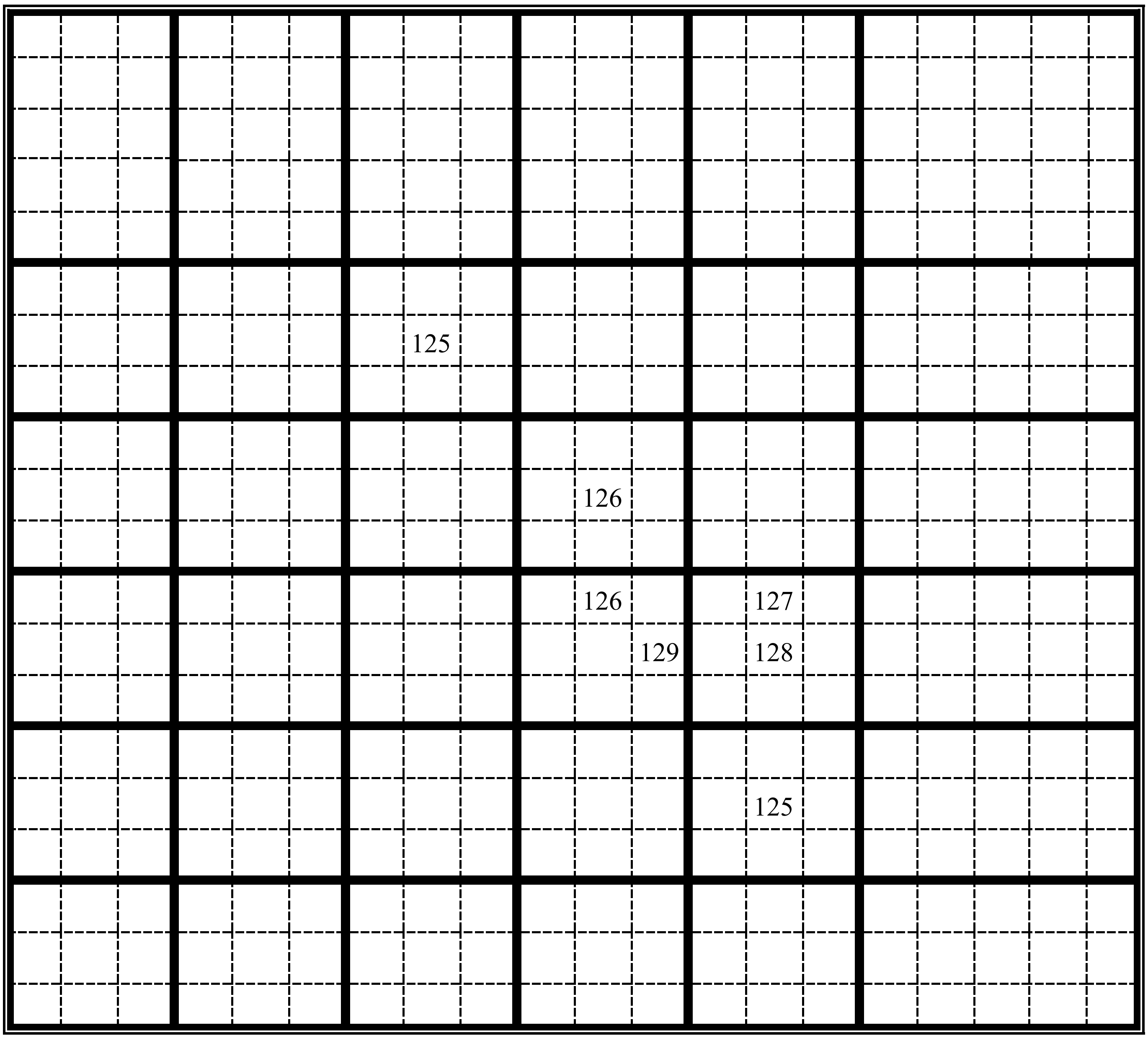


Figure A.2: DAY 2; Cox/Chu Ranking=32

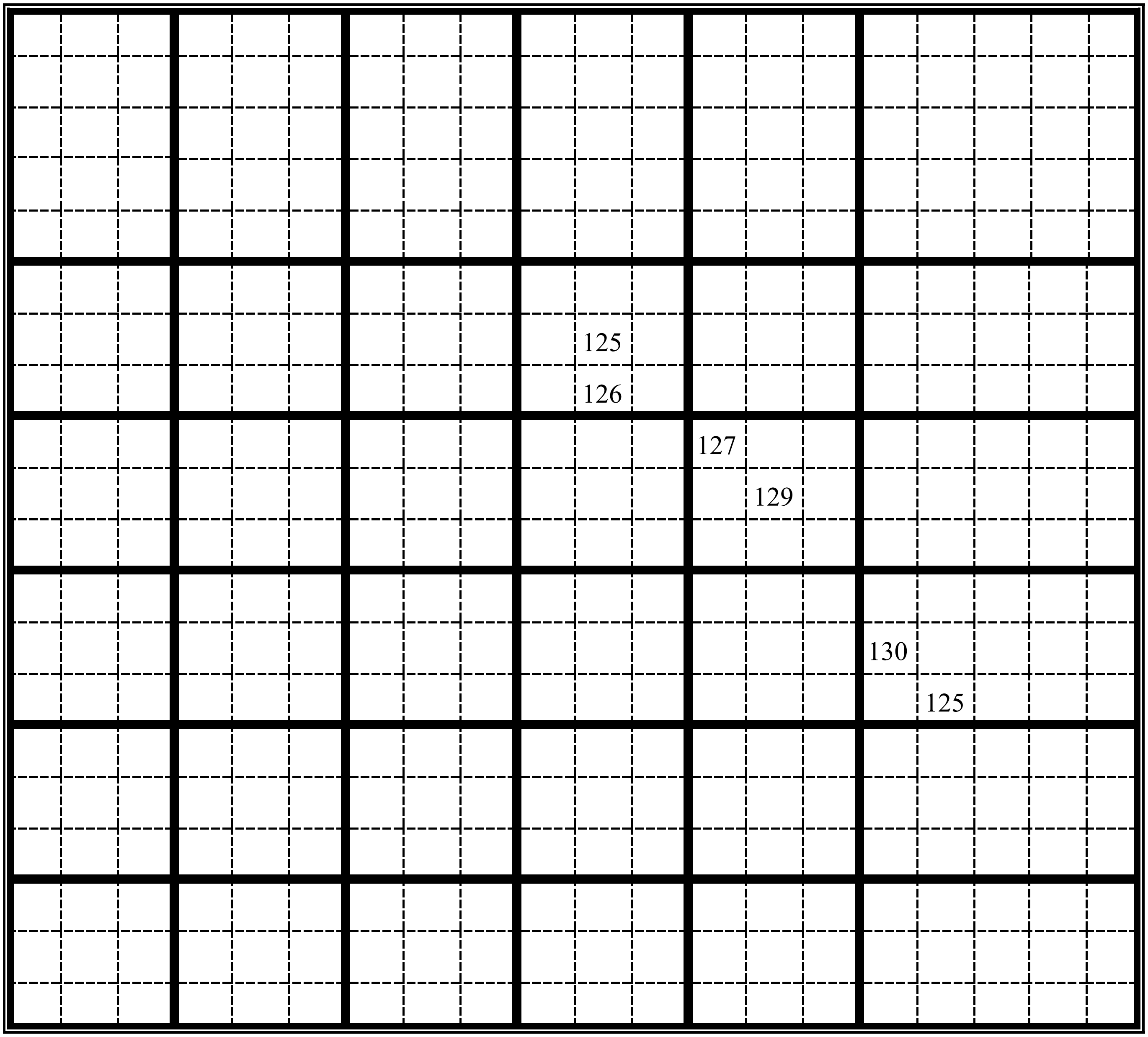
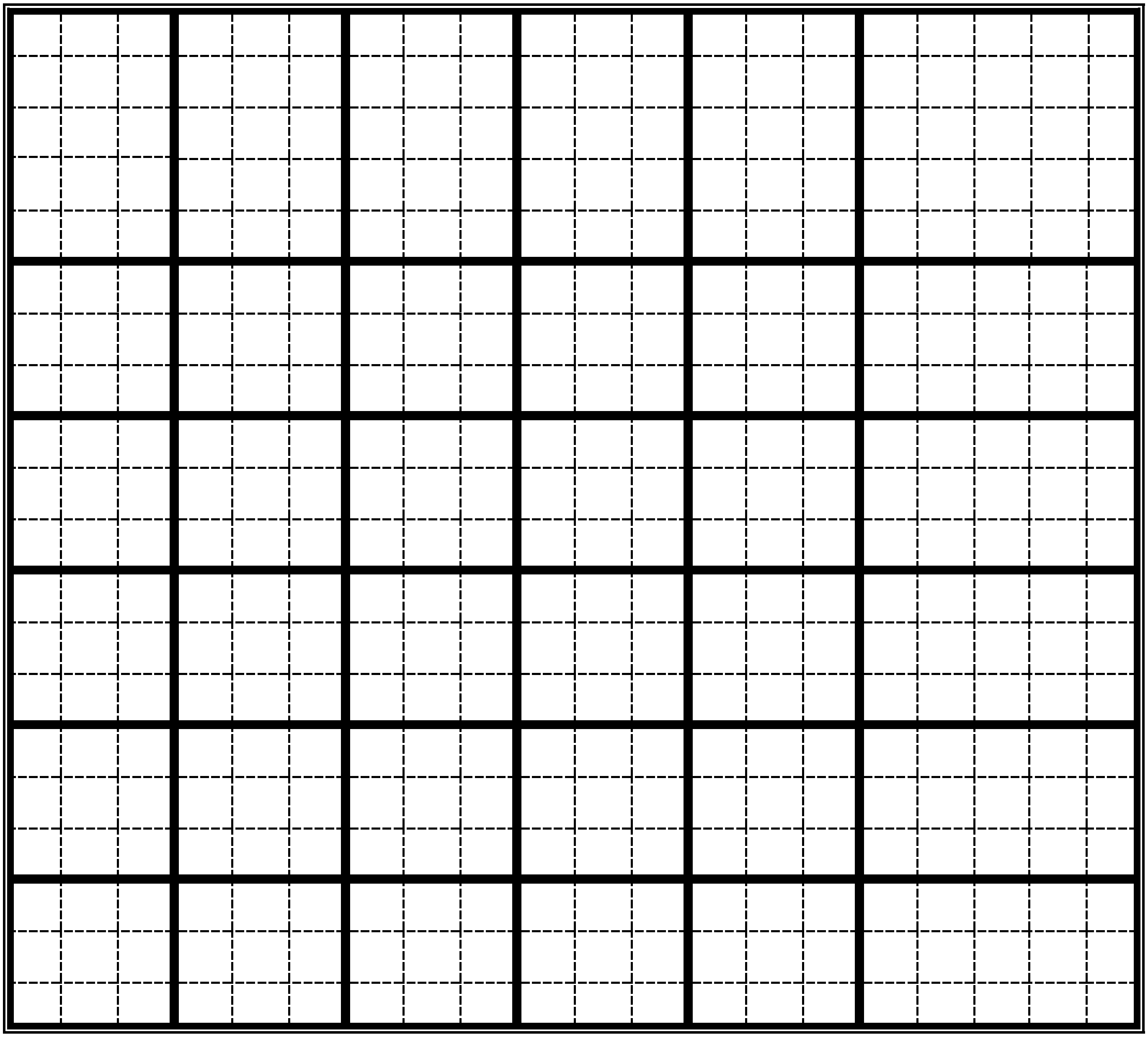


Figure A.3: DAY 3; Cox/Chu Ranking=215



APPENDIX B: GLOSSARY

Glossary

Benchmark

A measure or criterion against which a model result is evaluated. An example of a benchmark (used in the deterministic attainment test) is lack of any hourly ozone predictions greater than 124 ppb. A test consists of one or more benchmarks.

Deterministic Approach

A procedure for assessing whether a model result implies that a proposed strategy is likely to result in attainment of the NAAQS. The approach consists of two parts: a "deterministic test" and a "weight of evidence determination". Both of these latter terms are described below.

Deterministic Test

A test which incorporates a single criterion to determine whether a proposed control strategy demonstrates attainment: there must be no predicted ozone values > 124 ppb. In contrast to the "statistical test" (described below), it does not consider likelihood that meteorological conditions more severe than those modeled will occur.

Episode Days

An "episode" is a period of interest which is chosen for modeling to support State implementation plans to meet the air quality standard for ozone. Generally, an "episode" is a period of one or more days in which ozone concentrations exceeding the level specified in the ozone air quality standard (i.e., 0.12 ppm) have been monitored. Each day in such a period is referred to as an "episode day". The first day of a modeled episode is usually discounted, because model predictions may be too dependent on poorly known data. Model results obtained for the remaining days are compared to the national air quality standard using the attainment demonstration approaches described in this guidance. For this reason, these days are referred to as "primary episode days".

Exceedance

A term, coined by the EPA, meaning a model prediction (or observation) which is greater than the level specified in a

National Ambient Air Quality Standard (NAAQS). Thus, for ozone, an exceedance is any prediction greater than 0.12 ppm. Consistent with conventional rounding procedures and the practice long followed with monitored data, we consider concentrations of 115-124 ppb to be equivalent to 0.12 ppm.

Meteorological Ozone Forming Potential

A value for a daily maximum ozone concentration which is estimated using a regression model derived to describe correspondence between observed daily maximum ozone and several meteorological variables over a limited period (say 10 years). Because the regression model's independent variables contain only meteorological terms, it is possible to apply the regression equation over a many year period, prior to the time in which ozone measurements are available. All that is required is for the appropriate meteorological measurements to have been made. Thus, it is possible to rank "meteorological ozone forming potential" using a very long period of record (i.e., a large data base). This allows us to draw inferences about the number of times per year we would expect a modeled day's calculated "meteorological ozone forming potential" to be exceeded. We use this information to relate an ozone concentration modeled with a photochemical grid model to the "expected exceedance" form of the ozone NAAQS.

Meteorological Regime

A meteorological regime describes a set of meteorological conditions which are shared in common by a subset of days being considered for modeling. In our guidance, we distinguish among meteorological regimes according to broad, prevailing wind patterns (i.e., windfields). Orientation of sources of VOC and Nox with respect to one another and with respect to receptor sites (e.g., monitor sites) is determined by the wind pattern. We believe the relationship between control strategies and predicted ozone is affected by the way in which sources of precursors are oriented toward one another. To ensure that a proposed strategy is generally effective, we recommend that episodes selected from several meteorological regimes be modeled.

Mid-course Review

A "mid-course review" is a reassessment of modeling analyses and more recent monitored data to reaffirm that a prescribed control strategy is appropriate for attaining the ambient air quality standard for ozone by statutory dates. We require a mid-course review as a means for addressing uncertainty inherent in making projections in emissions and air quality many years into the future.

NAAQS

Abbreviation for "national ambient air quality standard". The NAAQS for ozone is that the daily maximum hourly concentration should not be expected to exceed 0.12 ppm more than 1.0 times per year at any monitoring site.

Observational Models

An "observational model" is one which relies on monitored air quality, and in some cases, meteorological data, to draw inferences about the types of control strategies which are likely to be effective in reducing ambient ozone. They typically examine relative amounts of monitored volatile organic compounds and nitrogen oxides as well as species of pollutants which may be indicative of emissions of a particular type of source of ozone precursors. They are called "observational" models, because their fundamental inputs are <u>observed</u> air quality data.

"Severe" Episode Day

We define a "severe" episode day as one whose meteorological ozone forming potential is not expected occur as frequently as twice per year. That is, the "expected exceedance frequency" (ExEx) is less than 2.0 per year. In the "statistical attainment test", described below, we relate a modeled episode day's expected exceedance frequency to the NAAQS for ozone, which allows 1.0 expected exceedance per year of a daily maximum ozone concentration of 0.12 ppm at every location.

Statistical Approach

A procedure for assessing whether a proposed control strategy is likely to result in attainment of the ozone NAAQS. The approach has two components: a "statistical test" (described below) and a "weight of evidence determination", also described below.

Statistical Test

A test which incorporates three criteria to determine whether a proposed strategy demonstrates attainment of the ozone NAAQS. The test includes (1)a limit on the number of days having modeled exceedances at each location; (2) a limit on the size of a modeled exceedance, and (3) in cases where a model underpredicts observed ozone, a requirement that a minimum level of improvement in predicted ozone be exceeded. The test is "statistical", because it considers the expected frequency with which the severity of a modeled episode day is exceeded.

Subregion

In this guidance, the term “subregion” refers to contiguous geographical areas approximately 15 km x 15 km. In the statistical test, we allow one or more exceedances of the ozone NAAQS to occur in each subregion on as many as 3 “severe episode days”.

UAM

An abbreviation for the “Urban Airshed Model”. The UAM is a 3-dimensional Eulerian photochemical grid model. As of June 1996, version IV (i.e., UAM-IV) of the UAM is recommended by the U.S. Environmental Protection Agency for use in attainment demonstrations needed to support State implementation plans to meet the ozone NAAQS.

Violation

A “violation” of the ozone NAAQS occurs when the expected frequency of a modeled or monitored exceedance is greater than 1.0 per year at any location. Thus, an exceedance at any location is not a violation unless there is reason to believe it will occur more frequently than once per year. The distinction between “violation” and “exceedance” is what permits a modeled attainment test to allow modeled exceedances.

Weight of Evidence Determination

A weight of evidence determination is the second component of the Deterministic Approach and the Statistical Approach. It entails use of results from a statistical or deterministic test, monitored data, additional modeled results and other data to make a judgment, consistent with all the available evidence, whether a proposed control strategy will attain the NAAQS by the statutory date. It is included as a part of an attainment demonstration in recognition of uncertainty inherent in the application of photochemical grid models for this purpose. Use of corroborative data is intended to reduce this uncertainty. This Guidance identifies different kinds of analyses that should be considered in a weight of evidence determination, as well as factors which affect how heavily each analysis should be weighed and outcomes which are consistent with allowing small deviations from the benchmarks in the deterministic or statistical attainment tests.

APPENDIX C: RELATED ISSUES

I. <u>General Questions About The Attainment Demonstration Process And The Relationship Of The Attainment Test To This Process</u>

1. Given that we typically only model 3 episodes while monitors measure ozone every day, why do we believe that the current modeled attainment test may be conservative?

The current modeled attainment test may be conservative for several reasons. First, our guidance focuses modeling on days observing the most severe ozone concentrations. Since these are the conditions which have already led to highest observed ozone, it is assumed that reducing ozone to the level of the NAAQS under these circumstances should also lead to levels below 120 ppb on other days. In short, the episode selection process does not choose a random sample of days to model. Rather, it is directed toward choosing those days for which it is anticipated attaining the NAAQS will be most difficult. The modeled attainment test should be viewed in this context.

The monitored attainment test is acknowledged as the definitive means for determining whether an area is attaining the ozone NAAQS. The modeled attainment test differs from the monitored attainment test in several important ways. First, the modeled test allows <u>no</u> exceedances of 120 ppb in any surface grid cell. In contrast to a limited number of monitoring sites, there are over a thousand of these cells in a typical application. Next, the monitored test allows up to three exceedances of 124 ppb (not 120 ppb) at each monitoring site during a three year period. Putting this another way, if there were a network of 15 monitors, each having complete sampling and observing 3 exceedances, an area could have 45 exceedances over 3 years and still meet the NAAQS. In contrast, the modeled test considers many more locations, yet permits no exceedances despite preselecting meteorological episodes observed to be most conducive to high monitored ozone.

2. How does transport (i.e., high boundary conditions) affect the attainment test?

Assumed boundary conditions do not affect the attainment test per se. Rather, the degree of transport and how it changes between the base period and the statutory attainment date affect the model results. As shown in Figure S.1 (executive summary), the model results are then compared to the benchmark(s) in the statistical or deterministic test for demonstrating attainment. The benchmarks themselves are unaffected by the boundary conditions used as input to the model. Although it is possible to confine application of the attainment test to only a portion of the modeling domain, this is a policy decision outside the scope of this guidance.

If model results come close to meeting the test's benchmark(s), some of the analyses included in the description of a weight of evidence determination (Section 5.3) may indirectly consider the role of transport in deciding whether the outcome of a test is close enough to a benchmark to provide sufficient confidence that attainment is demonstrated. For example, use of observational models and cost/benefit analysis in concert could determine that locally predicted ozone is not likely to be responsive to additional controls proposed on a local source or group of local sources. One reason for this may well be presence of a major transport component from sources outside the local jurisdiction. If there are ongoing efforts to develop a regional strategy (e.g., such as in the Ozone Transport Assessment Group (OTAG)), evidence of this nature could be used as a factor in deciding that model results which nearly meet applicable benchmarks are sufficient to demonstrate attainment.

It is important to recall that a technically acceptable attainment demonstration for an area with a statutory date of 2005 or beyond should have provisions for at least one mid-course review. If there does not appear to be a viable regional strategy for reducing transport and if there is little apparent progress toward meeting the benchmarks or NAAQS, a more stringent strategy could be invoked at that time.

3. Given the large uncertainty associated with projections over many years, would a demonstration which identifies measures to meet air quality goals in intermediate years (rather than at the statutory date), together with a commitment to conduct one or more mid-course reviews to identify subsequent necessary measures constitute an acceptable demonstration?

We understand the argument that uncertainty associated with model predictions increases the further one projects into the future. However, the answer to this question lies outside the scope of this guidance. The question is one that must be resolved by a policy decision which factors in legal constraints as well as technical considerations.

II. <u>Questions About The Statistical Test And Episode Ranking Schemes</u>

4. Using a long-term period like 41 years for ranking could lead to more frequent exceedances over limited periods due to long term periodicities in meteorological conditions. Wouldn't it be better to consider only the last 10-15 years, where meteorological conditions appear to be more conducive to high ozone in many parts of the country?

The answer to this question depends on how one should interpret the meaning of the NAAQS for ozone. The NAAQS is met if the expected number of times the daily maximum hourly ozone concentration exceeds 0.12 ppm is less than or equal to 1.0 per year. Since the NAAQS makes no mention of specific time frames for its application, the longer the period used to rank individual days, the more reliable the distribution of rankings should be.

The benchmarks in the statistical test are unaffected by the period chosen for ranking days. This follows because meeting the first two benchmarks is determined by the <u>expected exceedance frequency</u> of modeled episode days. However, the ranked value corresponding to an expected exceedance frequency varies depending on the number of years the ranking scheme is based upon. The impact of choosing a shorter period can be illustrated using Table 4.2. Suppose we use a ranking scheme based on 10-years data (rather than 41). The ranked values corresponding with ExEx rates of 0.5, 1.0 and 2.0 exceedances per year are "6", "11" and "21" respectively. This contrasts with ranks of "21", "42" and "82" appropriate for a 41-year sample.

The limit on predicted daily maximum ozone for days with ExEx between 0.5 and 2.0 times/year would continue to be 130 ppb. Looking at the column labeled "10" in Table 4.2, we see that the limit on the top ranked day is more restrictive than is the case for a 41-year period of record. Thus, there is a balance which occurs. Longer periods consider many more days. Those days which are highest ranked have higher permitted daily maxima. However, there is a greater likelihood that the distribution used to compute the rankings is a comprehensive one, and the probability that the top ranked day is exceeded is much smaller.

5. Variability in observed daily maxima depends on variability in emissions as well as in meteorology. Shouldn't a ranking procedure include emissions variability as a determining factor?

We agree that day to day variations in highest observed daily maximum ozone concentrations are likely to be affected by daily changes in emissions. The guidance allows flexibility in the choice of variables for use in statistical models to explain observed variation in daily maximum ozone so long as the criteria specified in Section 4.3.4 are met.

In choosing a statistical model to serve as a basis for ranking episode days, one should keep in mind the primary purpose for doing so. This is to provide a means of characterizing the shape of the distribution of highest daily maxima <u>after</u> a control strategy is implemented. This characterization allows us to judge whether an exceedance, modeled after controls are simulated is, nevertheless, consistent with meeting the NAAQS. Any control strategy is likely to change emissions in a major, systematic way. This could mean that the rankings assigned to days in the base case may no longer be valid.

A second problem with including day to day variation in emissions in a ranking scheme is that it is doubtful that these are known nearly so well as the daily fluctuations in key meteorological variables. If this information were known so precisely, should it not be included as input to the photochemical model? Inclusion of a poorly known independent variable in a regression equation would seem to open the door to speculative assumptions about how the variable differs from day to day.

In short, we are open to suggestions that are plausible and appear to work well in explaining variations in observed daily maximum ozone concentrations. However a proposed ranking procedure, which includes emissions as an independent variable in the underlying regression equation, would need some accompanying explanation concerning why our concerns about this are not well founded.

6. Why not simply use air quality observations directly to characterize severity of episode days?

As noted in the response to the preceding question, we are amenable to use of different models for characterizing variation in daily maxima. Clearly, this one is excellent in characterizing severity during the base case, but will it be in the future? The advantage use of meteorological variables has over air quality and emissions variables is that there is no reason to believe that they will change in any systematic way in the foreseeable future. Further, if emissions are reduced in future years, one would expect the distribution of future air quality concentrations to become increasingly dependent on meteorological fluctuations.

The problem of reconciling historical data with more current observations poses a second disadvantage with directly using air quality observations to rank days. For example, looking at air quality observations alone, might convince one that a 1983 episode was more conducive to high ozone than one in 1995. This might not actually be the case, since observations in 1995 may be reduced as the result of control efforts over the past 12 years.

A third, though minor, disadvantage to directly using air quality observations for ranking purposes in the Statistical Test is that the period of record would likely be limited to at most 15 years. Consequences of this are discussed in the response to question 4.

7. What does one do if there is more than one MSA with rankings for the meteorological ozone forming potential in a modeling domain and these rankings differ? Does a single ranking apply for an entire domain?

Choice of which set of rankings to use depends on the location of the highest and most pervasive exceedances modeled on each day as well as on the wind field. It is conceivable that rankings from different MSA's may be used for different episode days. In general, for each modeled day, one should choose the ranking from the MSA where the highest and most numerous exceedances are predicted. Consideration may also be given to the location of observed exceedances. If there is a tossup among two or more MSA's using the preceding criteria, but there is a major difference in the amount of emissions from the different MSA's, use the ranking for the MSA with the greatest emissions. If this last criterion is also inapplicable, choose the MSA with the highest ranking for the day in question. We do not recommend using different rankings for different parts of the typical size domain. We feel this would make the test unduly complicated. However, case by case exceptions could be considered, particularly for large regional scale domains on days where the windfield suggests little interaction among the MSA's.

It may also happen that the model’s worst or most numerous predicted exceedances occur in a location which we have not considered with our default ranking procedure. If ozone and meteorological observations are available nearby, an additional set of rankings may be developed for this area. The Cox/Chu approach need not be used for this purpose if the criteria identified in Section 4.3.4 are met by an alternative ranking approach.

8. What is the basis for allowing a predicted daily maximum of 130 ppb if a day's estimated severity is expected to be exceeded more than once per year?

The ranking procedure carries with it attendant uncertainties. Rankings are assigned using a statistical model which, typically, explains about 70% of the observed day to day variation in the highest daily maximum ozone. This means that some of the variation is unexplained (uncertain). In addition, as we note in Section 4.3.2, rankings exhibit some sensitivity to the regression model chosen to explain variations in observed highest daily maxima. The sensitivity increases as we move away from the most extreme episode days. We have used the preceding information to conclude that ExEx estimates between 0.5 and 2.0 are essentially equivalent to 1.0 ExEx/year in the statistical attainment test. Our guidance permits exceedances in each subregion on some (but not all) days with ExEx values in this range. An exceedance is a value greater than 124 ppb. We chose 130 ppb (the value corresponding to 0.5 ExEx/year), since the uncertainty in the analysis suggests all the exceedance rates in the 0.5-2.0 range are equivalent. Some might argue that this is not protective of the NAAQS. Our rejoinder is that it is easier to add controls as a result of a mid-course review (in which the uncertainty is diminished) than it is to compensate sources and the public for requiring unnecessary measures.

9. Why doesn't the default ranking scheme consider regional transport? How could a measure of regional transport be considered in locations where this is believed to be important? How could other phenomena such as flow reversal be considered?

The Cox/Chu equations used in the default procedure for ranking severity of days includes terms for morning and afternoon wind speed and direction. However, there is no term(s) reflecting widespread occurrence of meteorological conditions which are favorable for high ozone. Considering meteorological ozone forming potential calculated within a day or two in nearby MSA's might be a way to consider regional transport over a long period of record. It was not feasible for us to do this in deriving the default procedure, because it would have meant going through several iterations for each of the 32 areas we considered.

We have ranked data sets for a long period of record for each of the 32 areas mentioned above. Each ranking has an associated meteorological ozone forming potential. This information could be made available to a State which requests it so that it might be examined for MSA's located within several hundred miles of an MSA which is the focus of a modeling analysis. Meteorological ozone forming potential or ranked values calculated within a day or two at these nearby MSA's could be examined to see whether they are useful in explaining day to day variation in the observed highest daily maximum ozone concentration within the MSA which is the focus of the SIP attainment demonstration. As described in Section 4.3.4, we are open to the use of other procedures for ranking severity of days, so long as the criteria mentioned in that Section are met.

Other reviewers have hypothesized that flow reversal resulting in transport of a city's own plume back over the city is a condition which is likely to coincide with high observed ozone. If this is believed to be a major problem, not covered adequately by the morning and afternoon wind velocity terms in the default procedure, terms like the "change in afternoon vs. morning wind direction" might be explored to see whether the skill of the regression equation in explaining variation in daily maximum ozone is improved. Section 4.3.4 provides flexibility to tailor a ranking scheme for a particular location, so long as reasonable criteria are met.

10. Will rankings be based solely on 1953-1993 data? How can episodes occurring after 1993 be considered?

We can add new years to the data base periodically. This should enable consideration of episodes from later years (with a short time lag) using the default approach. As can be seen from Table 4.2, increasing the period of record beyond 41-years affects the ExEx values corresponding to ranked values in minor ways. Similarly, the limit in the magnitude of an allowed exceedance is pretty insensitive to minor increases in the period of record used to rank days. For previously selected episodes, we would allow a State to retain rankings based on the 41-year period of record if the modeling was underway and if the State chose to retain the ranking.

III. <u>Questions Relating to Uncertainty and Weight of Evidence</u>

Determinations

11. Doesn't a weight of evidence determination establish a new, defacto benchmark which relaxes the NAAQS?

The NAAQS is unaffected by the modeled attainment test. States are obligated to meet the NAAQS by the statutory dates. The weight of evidence determination is introduced so that other information, in addition to the photochemical grid model results, can be used to help make the best judgment we can about whether a proposed strategy is likely to be successful in meeting the NAAQS within required timeframes.

Unless benchmarks in the deterministic and statistical tests are nearly met, the weight of evidence provided from other analyses will need to be very compelling to overcome that resulting from the photochemical grid model. If the photochemical grid model has been applied in accordance with published guidance, we believe its predictions should be given considerable weight. Nevertheless, there is uncertainty associated with the predictions. This is why use of corroborative information is desirable.

Use of corroborative information is not intended to result in defacto, relaxed benchmarks. If results of corroborative analyses are also consistent with the conclusion that a strategy will be insufficient to meet the NAAQS by the statutory date, attainment would not be demonstrated.

12. Uncertainty is a two-edged sword. The two recommended modeled approaches for demonstrating attainment appear to be biased toward preventing unnecessary control measures. What protections exist against a modeling analysis underestimating needed control measures?

As noted in the response to question 1, we believe the current test is more conservative than what the NAAQS requires. This conservatism was introduced to ensure the NAAQS would be met, given uncertainty attendant in the modeled estimates. In the two approaches we are now recommending for modeled attainment demonstrations, we attempt to consider factors leading to uncertainty more explicitly in the approaches or we introduce requirements intended to reduce uncertainty. For example, the statistical test incorporates consideration of episode severity in its benchmarks. The deterministic test (which is really very similar to the existing test), may be supplemented with corroborative analyses in order to justify not meeting its benchmark. The intent of the corroborative analyses is to reduce the uncertainty about whether a strategy will be sufficient to attain the NAAQS. Reducing the uncertainty (i.e., the corroborative analyses support adequacy of the strategy) justifies relaxing the requirement to meet the benchmark. We believe the requirement for an attainment demonstration to include provision for one or more mid-course reviews offers the most important protection against prescribing insufficient controls. This brings to bear stronger corroborative evidence, as well as modeling with smaller uncertainty (due to shorter projection periods) in assessing whether an existing strategy is adequate to attain the NAAQS.

13. Are the types of analyses which can be considered in a weight of evidence determination limited to those described in Section 5.3?

No. Other analyses may also be considered. We recommend that they be identified and that the rationale for their use be described to the appropriate U.S.EPA Regional Office before resources are expended.

TECHNICAL REPORT DATA

(Please read Instructions on reverse before completing)

1. REPORT NO.	2.	3. RECIPIENT'S ACCESSION NO.
EPA-454/B-95-007		

4. TITLE AND SUBTITLE	5. REPORT DATE
Guidance on Use of Modeled Results to Demonstrate Attainment of the Ozone NAAQS	June 1996
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS U.S. Environmental Protection Agency Office of Air Quality Planning and Standards Research Triangle Park, NC 27711	10. PROGRAM ELEMENT NO. 11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED 14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT

This document provides guidance on use of results from photochemical grid models to test whether a proposed control strategy is likely to be sufficient to demonstrate attainment of the one expected exceedance form of the ozone NAAQS. The guidance described herein should be used in concert with Guideline for Regulatory Application of the Urban Airshed Model (EPA-450/4-91-013, July 1991) to apply grid models to determine attainment of the ozone NAAQS. This guidance describes two acceptable approaches for demonstrating attainment: a statistical approach and a deterministic approach. It also describes how corroborative evidence may be used with each approach in a weight of evidence analysis to determine whether attainment of the NAAQS is likely. Because of attendant uncertainties in the modeling process, the guidance requires use of subsequent mid-course reviews to support any needed refinements to an attainment strategy.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/ Group
Air Pollution Atmospheric Dispersion Models Statistical Models	photochemical grid models ozone attainment tests meteorological conditions	
18. DISTRIBUTION STATEMENT Release Unlimited	19. SECURITY CLASS *(Report)* Unclassified	21. NO. OF PAGES 77
	20. SECURITY CLASS *(Page)* Unclassified	22. PRICE

EPA Form 2220-1 (Rev. 4-77) PREVIOUS EDITION IS OBSOLETE