

AIR QUALITY MODELING PROTOCOL

Meteorological, Emissions and Air Quality Modeling for an Integrated Assessment Framework in Support of the Southern Appalachians Mountain Initiative

Prepared For:

The Southern Appalachians Mountain Initiative (SAMI)
Technical Oversight Committee

59 Woodfin Place
Asheville, NC 2880

Prepared By:

A.G. (Ted) Russell
Georgia Power Professor
School of Civil & Env. Engineering
Georgia Institute of Technology
200 Bobby Dodd Way
Atlanta, GA 30332-0512
work: (404) 894-3079
fax: (404) 894-9724
e-mail: trussell@pollution.ce.gatech.edu

Richard (Dick) McNider
Professor of Mathematics
Department of Mathematical Sciences
University of Alabama in Huntsville
Huntsville, AL 35899

work: (205) 922-5756
fax: (205) 922-5755
e-mail: mcnider@atmos.uah.edu

James G. Wilkinson
School of Civil & Env. Engineering
Georgia Institute of Technology
200 Bobby Dodd Way
Atlanta, GA 30332-0512
work: (404) 894-3089
fax: (404) 894-9724
e-mail: jwilkins@isis.ce.gatech.edu

Jennie Moody
Research Assistant Professor
Department of Environmental Science
University of Virginia
Clark Hall
Charlottesville, VA 22903
work: (804) 924-0569
fax: (804) 982-2300
e-mail: jlm8h@virginia.edu

M. Talat Odman
School of Civil & Env. Engineering
Georgia Institute of Technology
200 Bobby Dodd Way
Atlanta, GA 30332-0512
work: (404) 894-2783
fax: (404) 894-8266
e-mail: talat.odman@ce.gatech.edu

April 21, 1997

TABLE OF CONTENTS

1.0	INTRODUCTION	1
1.1	Background	1
1.2	SAMI Air Quality Modeling Study Objectives	2
1.3	Purpose of the Protocol	4
1.4	Overview of the Study	4
1.5	Document Structure	4
2.0	URBAN-TO-REGIONAL MULTISCALE (URM) MODEL	6
2.1	URM	6
2.2	RSM Modifications	8
2.3	Aerosol Modeling	8
2.4	DDM-3D Modifications	8
2.5	Parallel versus Sequential Processing	10
3.0	MODELING DOMAIN AND DATA AVAILABILITY	11
3.1	Modeling Domain	11
3.2	Data Availability	14
3.2.1	Emissions Data	14
3.2.2	Air Quality Data	14
3.2.3	Meteorological Data	15
3.3	Quality Assurance	15
4.0	URM INPUT DATA	16
4.1	Emissions	16
4.1.1	Area Source Emissions	16
4.1.2	Elevated Point Source Emissions	16
4.2	Air Quality	16
4.2.1	Surface Level Initial And Boundary Conditions	17
4.2.2	Upper Levels Boundary And Initial Conditions	17
4.3	Meteorological Data	17
4.3.1	Mixing Depths	17
4.3.2	Wind Field	18
4.3.3	Surface Temperature Field	18
4.3.4	Surface Absolute Humidity	18
4.3.5	Total Solar Radiation	18
4.3.6	Ultraviolet Radiation	18
4.4	Surface Deposition	19
4.4.1	Land Use	19
4.4.2	Surface Resistance	19
4.4.3	Surface Roughness	19

4.4.4	Average Surface Roughness	20	
4.5	Grid Definition	20	
4.6	Chemistry	20	
4.7	Command Files	23	
5.0	MODEL PERFORMANCE EVALUATION	24	
5.1	Principles	24	
5.2	URM Model Evaluation Process	26	
5.2.1	Step 1: Mechanistic Evaluation	26	
5.2.2	Step 2: Initial Screening	26	
5.2.3	Step 3: Operational/Scientific Evaluation	28	
5.3	Statistical and Graphical Tools to Be Used in the Evaluation	29	
5.4	Investigative Simulations To Be Considered In The Scientific Evaluation	30	
5.5	Provisional Rejection Criteria	33	
5.6	Domain-wide Versus Subdomain Performance Evaluation	35	
6.0	BASE CASE, BASELINE, RESPONSE SURFACES, AND EMO MODELING	37	
6.1	Base Case	37	
6.1.1	Episodes	37	
6.1.2	Expected Results	37	
6.2	Baseline	37	
6.2.1	Episodes	37	
6.2.2	Expected Results	37	
6.3	Response Surfaces	37	
6.3.1	Use of DDM-3D	37	
6.3.2	Expected Results	37	
6.4	Emissions Management Options (EMOs)	37	
6.4.1	Episodes	37	
6.4.2	EMO Development	37	
6.4.3	Expected Results	37	
6.5	Expected Publications	37	
7.0	REFERENCES	38	
APPENDIX A -- MODEL EVALUATION PROCEDURES		39	

1.0 INTRODUCTION

The air quality modeling protocol documents the process to perform air quality modeling analyses needed to support the Southern Appalachian Mountains Initiative (SAMI) in its effort to provide a regional strategy for assessing and improving air quality in the southern Appalachian Mountains.

Important Note. This document should be viewed as a work in progress. As a *first of its kind* study, that being an integrated assessment of ozone, acid deposition, and regional haze, new information will inevitably be revealed about how an integrated air quality modeling project such as this should be conducted. In particular, air quality model performance evaluation guidelines will likely be the subject for much debate during the course of the SAMI study. Please note, further, that although federal guidelines exist for the treatment of urban-scale ozone modeling results, no such guidelines exist for regional-scale ozone modeling studies, though the defacto standard has become the same urban-scale ozone modeling performance evaluation metrics. Further, the urban-scale modeling performance evaluation metrics and criteria continue to be scrutinized by the scientific community. No formalized air quality model performance evaluation criteria exist for acid deposition or regional haze.

In addition to the likely debate that will surround the model performance evaluation criteria, the proposed sensitivity and uncertainty analysis efforts, which are proposed as a standard component of this and all future studies, will also likely stir heated debate. Although the science and the methods that surround the proposed sensitivity and uncertainty analysis techniques have been extensively peer reviewed, these techniques are generally untested as formal methods in stakeholder air quality modeling, such as the SAMI air quality modeling study. Much new information will be revealed as a result of these proposed, standard analyses; hence, a not insignificant learning process will be inevitable as this information is converted to knowledge.

Also, the episodes that will be modeled have not yet been chosen, inhibiting our ability to determine what data will be available for model application and evaluation. For example, the PAMS network, NARSTO-Northeast and the Southern Oxidants Study can provide additional data for specific periods.

1.1 Background

Studies that have been conducted in national parks and national forest wilderness areas of the southern Appalachian Mountains have documented adverse effects to visibility, streams, soil, and vegetation. Poor air quality in the region has been implicated as the major source of the adverse effects. Beginning in 1990, the Federal Land Managers for Shenandoah National Park, Great Smoky Mountains National Park, and Jefferson National Forest/James River Face Wilderness Area made several adverse impact determinations in their review of proposed air permits for major new sources. Although it is known that the air pollution levels which currently

affect park and wilderness resources come from existing sources of pollution -- large and small, mobile and stationary, near and distant -- the relative contribution of each source type to the regional air pollution problem is not well quantified. Further, the current air pollution levels in these areas typically do not exceed federal standards; yet the current air pollution levels are threatening the natural ecosystems, resources, diversity, and beauty of the Southern Appalachian Mountains.

Over the course of the next decade, the 1990 Clean Air Act Amendments (CAAA) require major reductions in primary airborne pollutants, including sulfur oxides (SO_x), nitrogen oxides (NO_x), and volatile organic compounds (VOC), as well as secondary pollutants such as ozone. Although the reductions are expected to produce air quality improvements, there is uncertainty whether the results will be enough to protect and preserve the delicate ecosystems and natural resources of the Southern Appalachians, especially in Class I areas.

In March 1992, a conference was held in Gatlinburg, Tennessee to examine scientific evidence on the air pollution effects in the Southern Appalachian Mountains and to explore ideas to mitigate these effects. As a result of the discussions held at the Gatlinburg conference, the eight states surrounding the Southern Appalachian Mountains (Alabama, Georgia, Kentucky, North Carolina, South Carolina, Tennessee, Virginia, and West Virginia), the Environmental Protection Agency (EPA), the National Park Service (NPS), and the U.S. Forest Service met in June 1992, to launch the Southern Appalachian Mountains Initiative (SAMI). Through this cooperative effort, the SAMI participants seek to establish the necessary forums and processes which will be used to assess and mitigate the air pollution problems of the southern Appalachian Mountains region.

1.2 SAMI Air Quality Modeling Study Objectives

To this end, SAMI is building an integrated assessment framework (IAF) which will model the environmental and socioeconomic responses to changes in air emissions which result from various emissions management options (EMOs). SAMI has proposed an IAF that is divided into six linked modules:

- > Base year emissions inventory, emissions projections, and control costs;
- > Air chemistry and transport;
- > Effects -- acid deposition (aquatic and terrestrial);
- > Effects -- ozone (vegetation);
- > Effects -- visibility; and
- > Socioeconomic consequences.

This protocol addresses one component of the *air and chemistry transport* module -- the air quality modeling. Two other components of the *air and chemistry transport* module are the meteorological modeling and the emissions modeling which are addressed in other protocols.

The purpose of the SAMI air quality modeling study (i.e. air chemistry and transport component of the IAF) is to characterize the air pollution formation processes that affect air quality in the southern Appalachian Mountains. To adequately characterize the processes of interest, a number of historical air pollution episodes (i.e. base case episodes) which include ozone exceedances, elevated levels of acid deposition, and/or regional haze problems are to be studied from which the following can be assessed:

- > The ability of the air quality model, the Urban-to-Regional Multiscale (URM) model, to reproduce adequately observed air quality for a number of historical air pollution episodes;
- > The viability of using projected emissions estimates of the base case episodes (i.e. the baseline emissions estimates) for use in studying EMOs; and
- > The robustness of the air quality modeling results as they pertain to use in the SAMI effects studies.

Determination of the number and type of historical episodes is being conducted under the SAMI Episode Selection contract; however, it is known that fifty (50) days from the years 1988 through 1995 will be selected for modeling purposes. It is not known how these fifty days will be constructed in terms of multi day episodes (e.g. 5-10 day episodes; 10-5 day episodes; 2-12 day, 1-6 day, 1-10 day, and 2-5 day episodes, etc.). Of note, ramp-up days will be added to each episode to help minimize the effects of initial air quality conditions. Two ramp-up days were found to be suitable in a recent sensitivity study of ozone that was conducted by Yang et al. (1997). The final determination of ramp-up days will be based on the meteorology of the episode.

The five goals of the SAMI air quality modeling study are to: (1) apply the URM to a number of historical episodes in the SAMI region; (2) evaluate the viability of using baseline episodes (i.e. future year emissions estimates projected from the base case emissions estimates [of note, the future years are 2010 and 2040]) to model EMOs; (3) construct baseline sensitivity isopleths of selected air quality metrics which will be used to help develop EMOs; (4) conduct baseline, with selected EMOs applied, air quality model runs to determine the efficacy of the selected EMOs; and (5) provide selected air quality model results for use in the SAMI effects modeling. Other objectives of the study include:

- > Assess the modeling and analysis needs/constraints affecting the SAMI air quality modeling study;
- > Prepare and update an air quality modeling protocol that guides the application of the URM to the SAMI region;
- > Prepare air quality model inputs for the episode days of interest;
- > Conduct model performance evaluation analyses to establish the reliability of the URM and supporting data sets for judging the efficacy of EMOs in the SAMI region;
- > Assist SAMI staff and other stakeholders in evaluating EMOs; and

- > Transfer the full set of modeling files and data bases used in the study to the SAMI Technical Oversight Committee (TOC).

1.3 Purpose of the Protocol

This protocol documents the activities necessary to conduct the air quality modeling required for the SAMI study. These activities include: (1) selection of appropriate modeling tools and data bases; (2) construction of an integrated air quality modeling framework to study the pollution processes for ozone, acid deposition, and regional haze; (3) performance evaluation of the URM air quality model on selected base case episodes; (4) evaluation of baseline episodes for use in EMO testing; (5) construction of the baseline air quality metric sensitivity isopleths for use in EMO development; (6) evaluation of air quality modeling predictions as a result of EMOs; (7) preparation of URM results for use by the SAMI effects modelers.

1.4 Overview of the Study

The SAMI air quality modeling study consists of several elements:

- > Development of a final work plan (completed 13 April 1997)
- > Development of the air quality modeling protocol (this document) that describes the background, objectives, and procedures to be followed in the modeling analysis;
- > Preparation of the air quality inputs for each modeling episode;
- > Development of the base year, base case episodes, and baseline episodes boundary conditions;
- > Application and performance testing of the URM for the base year and base case episodes;
- > Application of the URM to each baseline episode to establish future-year (2010 and 2040) baseline conditions;
- > Development of *interim* across-the-board emissions reduction strategies to estimate the level of controls needed to meet specific air quality objectives;
- > Application of the URM for a variety of emissions sensitivity simulations for a number of episodes to explore the effects on air quality metrics due to various emissions reduction scenarios including tradeoffs between ground-level versus elevated sources, point versus mobile sources, local versus more distant sources, etc.;
- > Application of the URM using selected EMOs (as applied to the baseline episodes) to demonstrate the efficacy of the EMOs to attain various air quality goals; and
- > Documentation of the URM air quality modeling analysis results and submittal of the results to the SAMI TOC for their review and approval.

1.5 Document Structure

Section 2 describes the URM, the modifications necessary to URM for use in the SAMI study, and advantages/disadvantages of running URM in parallel and sequential mode. Section 3 discusses what data sets will be used to develop the initial and boundary conditions and the performance evaluation databases for each episode. In addition, quality control procedures for the air quality data sets will be discussed. Section 4 describes the URM input files and how the input files are constructed. Section 5 describes the URM model performance evaluation procedures and the acceptance/rejection criteria. Section 6 discusses the types of air quality modeling that will be conducted: base year, base case, air quality metric response surfaces, baseline, and baseline with EMOs. Furthermore, Section 6 describes the foundation for reporting the air quality modeling results and what the results may be. Section 7 presents the references.

2.0 URBAN-TO-REGIONAL MULTISCALE (URM) MODEL

The following section presents a brief overview of the capabilities of URM and briefly discusses the modifications to URM that are underway to support the needs of the SAMI study.

2.1 URM

The most challenging aspect of the project is the air quality modeling. Three different aspects of air quality are to be modeled: ozone, acid deposition and fine particulate matter. URM, and its monoscale predecessor the California/Carnegie Institute of Technology (CIT) model, have been applied to, and evaluated for, most of the air quality aspects desired, including ozone (e.g. Kumar et al. 1993, 1995; Russell, 1996), fine particulate matter (in its monoscale version, Russell et al. 1986; Pilinis et al. 1993) and dry acid deposition (in its monoscale version, Russell et al. 1986). (All the capabilities of the monoscale version, and more, e.g. the plume-in-grid model, are available in the multiscale version.) While it does include aqueous phase chemistry, which has been evaluated, (McNair, 1995) and used for wet deposition, it has not yet been evaluated for its ability to accurately reproduce wet acid deposition fields.

URM, and its monoscale version, the CIT model, is likely the most widely used and accepted photochemical air quality model amongst the academic, scientific air quality modeling community. They (one or both) are being used at Georgia Tech, Caltech, University of Michigan, University of California (Berkeley and Irvine), Carnegie Mellon, and Florida for ozone, fine particulate and deposition modeling, primarily at the mesoscale (large urban and regional areas). It is also being used by a number of governmental and international institutions, including Los Alamos National Laboratory, the Instituto Mexicano del Petroleo, Ecole Polytechnique Federale de Lausanne and Nissan. The model has been applied over a number of different domains, including Los Angeles, Mexico City, Athens, the Kanto Plain, the Northeastern United States, Eastern United States, Switzerland, and the U.S.-Mexico border (e.g. Harley et al. 1993; Kumar et al., 1993; Kumar and Russell, 1996ab; Dominguez, 1996). Application to these areas with very different emissions characteristics, and many of which are topographically very complex, have demonstrated its capabilities over a wide range of conditions. While much of its use has been towards research, it has also been used for regulatory guidance in California and Pennsylvania. The extent of its use for photochemical modeling is probably second only to UAM-IV, a model which is not appropriate for this study. In this study, the multiscale version (URM) would be used. Multiscale is similar, though numerically preferable to, using two-way nested grids. URM is described briefly in McNaughton and Mayes (1995), and will not be repeated here. More complete descriptions are available in the literature (e.g. McRae et al. 1982, 1983; Russell et al., 1985, 1986ab; Harley et al. 1993; Kumar et al. 1994; Kumar and Russell, 1996ab). Below, additional points not covered in the SAMI review are given.

URM has recently undergone more extensive evaluation for its ability to reproduce photochemical oxidant levels in the Northeast (Kumar and Russell, 1996b). Recent analysis

against observed data suggests that it captures the plume growth and dynamics well. URM/CIT has been developed and maintained by Dr. Russell's research group, and is a public domain model. URM has an aerosol module which has been evaluated successfully (e.g. Russell et al. 1987; Pilinis et al. 1989). Current capabilities include following the evolution of secondary sulfate, nitrate, ammonium and organics. An updated module is now being implemented, and will be available for this project.

At this time, URM follows four aerosol size bins: three bins below 2.5 microns and one between 2.5-10 microns. This way, combination of the first three bins results in $PM_{2.5}$ while combining all four size bins yields PM_{10} . Given the uncertainty in the particulate matter emissions and the limited time available for the modeling, the number of size bins will not be changed (increased). Results from the aerosol modeling will be used for visibility calculations. The decision on how to output the aerosol results (i.e., $PM_{2.5}$ or PM_{10} or both) will be made after discussion with the SAMI Effects Subcommittee. Results from the aerosol modeling are also important for assessing dry and wet deposition. Also, as indicated in the SAMI executive summary and terrestrial effects reviews, an important process that will be captured will be the impact of acidic clouds on mountains.

McNair (1995) used URM to successfully simulate the chemistry and physics of a stratus cloud impacting the San Gabriel Mountains during the 1987 Southern California Air Quality Study. Comparison between observations and predictions showed very good agreement when only five vertical layers were used in URM. As part of the EURAD modeling project, a reduced form wet acid deposition capability was added (Giovannoni et al. 1995). It is important to note that McNair (1995) application used 5-km grid size. Since we are planning to use 12-km grid size over the Southern Appalachian Mountains, the effect of such acidic clouds may not be resolved explicitly. A more comprehensive module for capturing wet acid deposition is currently being added to URM -- the Reactive Scavenging Module (RSM). RSM is considered the state-of-the-science for wet deposition processes. RSM represents in-cloud processes with twenty (20) vertical layers which should sufficiently resolve all events of interest; however, the validity of this assumption will be investigated. While McNaughton and Mayes (1995) note that URM uses a very fast and proven chemical solver, URM also has a very fast and accurate advection routine, that does not require re-calculation of the chemical dynamics at multiple resolutions (as is the case for some nesting methods). This is important for multiscale calculations, as is desirable for the SAMI project.

Seven vertical layers for the model domain, extending to a depth of about 6000 meters above ground level, are proposed as the primary vertical structure for the SAMI study. In past studies, between five and twenty layers have been used; however, little difference in the ground level air quality results have been observed as a result of the finer vertical detail, except at night (e.g. Byun and Dennis, 1995) for ozone. This problem was corrected when the lowest cell was reduced in size. Based on past experience, a surface layer that is about twenty meters in depth (or less) appears to capture the pollutant dynamics near the ground; however, an appropriate vertical structure will be determined during the preliminary modeling and in consultation with the

TOC. The choice of seven layers is a compromise between added detail and computational costs. It is important to note that while URM will use seven layers, the cloud module (RSM) will have a different internal structure with 20 layers. This will assure that cloud processes are accurately resolved for reliable acid deposition modeling. Nevertheless, we also propose to conduct at least one episode simulation with additional vertical layers (15) and extending up to about 15 kilometers to quantify the difference in model results. Extension to 15 kilometers will also help in the investigation of the role of stratospheric intrusion. Similarly, the choice of horizontal grid structure is a compromise. Based on past studies, finer meshes work well to pick up the pollutant dynamics in source regions, and less finely resolved grids are necessary in regions with low emissions density (e.g. Kumar et al., 1994). An initial part of the project will be to test different grid structures and their interaction with the RAMS fields. The final grid structure to be chosen will retain the fidelity of the results over the Southern Appalachians found using the base grid, though may use less nesting in some areas to reduce computational requirements. The choice of grid will be done in consultation with the SAMI Technical Oversight Committee.

2.2 RSM Modifications

URM has not been used for wet deposition in past studies. As part of this project, the Reactive Scavenging Module (RSM) developed by Berkowitz et al. is being added to URM for this purpose. RSM is one of the most advanced representations of the processes leading to acid deposition that has been used for three-dimensional modeling. RSM includes descriptions of gas-phase pollutant scavenging by droplets, atmospheric chemistry, acid deposition integration and convective mixing under the clouds. Originally, this routine was intended as the update for RADDM, but was not implemented due to time limitations.

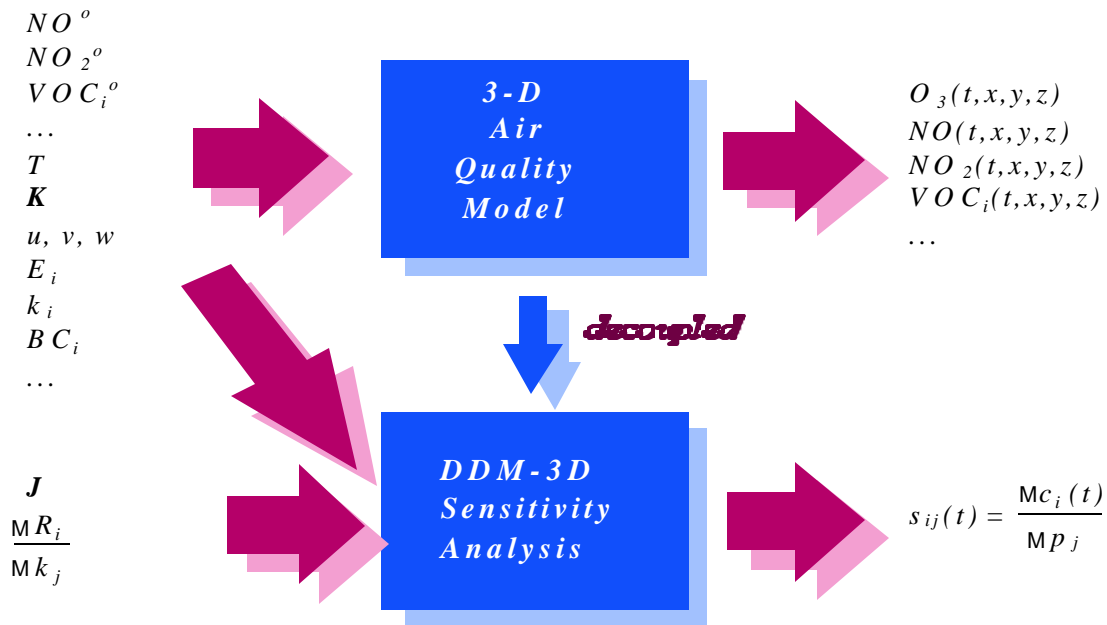
2.3 Aerosol Modeling

Aerosol dynamics will be followed in URM using an equilibrium approach for the inorganic salts. As proposed originally, the aerosols will be divided into two classes: PM_{fine} and PM_{coarse} , with a cut-off of 2.5 μm . The cut-off is chosen to concur with the cut-off used by the IMPROVE monitoring data, which is then used for effects modeling. While the technique is available to give greater size resolution, the way the data will ultimately be used does not use that increased detail, and there is a significant computational overhead. (The visibility models are based on using $PM_{2.5}$ speciated data.) Primary compounds will be followed directly from emissions. The feasibility of the incorporation of the recent results on how semi-volatile, secondary organics are formed (Odum et al., 1997) will be assessed. Those results suggest that the condensation of secondary organics is primarily dependent on the existence of other condensed phase organics present, as opposed to the level of inorganic particulate matter. We plan to investigate the use of more size bins, as well, though this is primarily a check on the sensitivity/uncertainty of using a calculation using only two size bins.

2.4 DDM-3D Modifications

An added feature of URM/CIT that is of particular interest to this study is the recent development of a method to directly, and with very little computational expense, do model sensitivity analysis of multidimensional air quality models. The method, DDM-3D (Yang et al., 1997), is numerically more exact and less limited than the traditional brute force approach of conducting multiple simulations and finding model sensitivities by difference. It is also much more computationally efficient. In essence, it is possible to derive how model results depend on any input (both spatially and temporally) or model parameter at the same time the model is being exercised. These results can, possibly, reduce the computational intensity of this project, or can be used to show how the sensitivity of ozone and particulate matter to emissions vary as a function of the emissions level. We are currently using DDM-3D to develop Area-of-Influence (AOI) methods, as being examined by the EPA FACA committee on Ozone, Particulate Matter and Regional Haze.

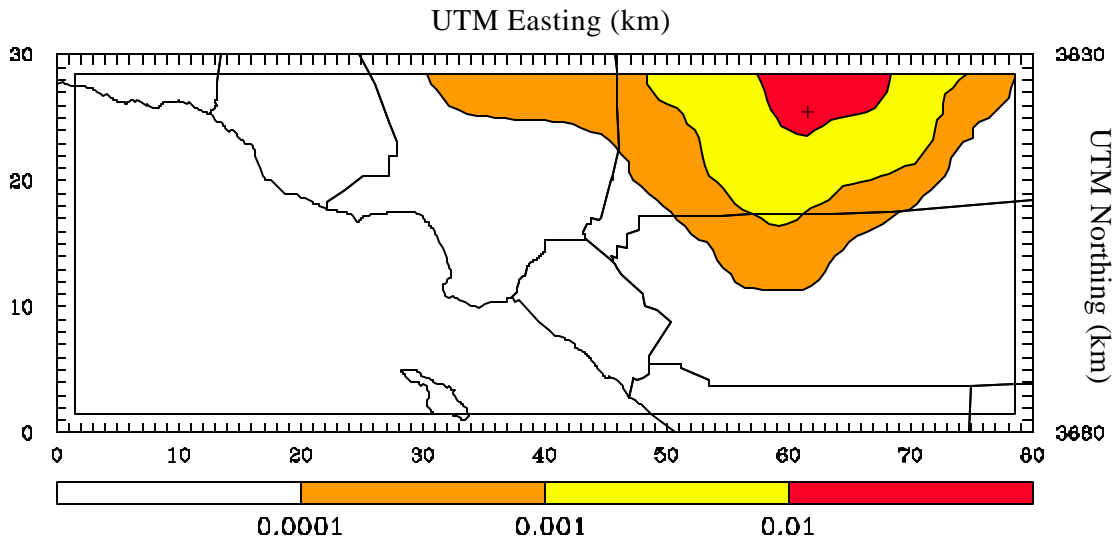
Decoupled Direct Method is a formal sensitivity analysis, which was developed for photochemical kinetic problems without transport process in the model (Dunker, 1984). Recently, this method has been further expanded to a three-dimensional air quality model, i.e. CIT photochemical model applied to southern California. Using a decoupled numerical technique, comprehensive sensitivities of a model can be efficiently solved without numerical instability problems. DDM-3D is schematically shown in Figure 2-1.



The magnitudes of model parameters and inputs can range dramatically in space and time, and, likewise, sensitivity coefficients to different parameters may differ by several orders of magnitude. Thus, it is useful to define and use semi-normalized sensitivity coefficients. Given a model parameter, p_j , its variation is defined here as $p_j(\mathbf{x},t) = e_j P_j(\mathbf{x},t)$, where $P_j(\mathbf{x},t)$ is the unperturbed field, which can vary in time and space, and e_j is a scaling variable with a nominal value of one. Here, what is termed the semi-normalized sensitivity coefficient, s_j , is calculated by the partial derivative of a species concentration, c_i , to the scaling variable e_j of parameter j , e_j .

In this study DDM-3D will be used to calculate the sensitivities of ozone, PM (which will be used for visibility), and acid deposition to precursor emissions. The important precursors are SO₂ (PM and acid deposition), VOCs (ozone and PM, and to a lesser extent, acid deposition), NOx (ozone, acid deposition and PM), ammonia (PM, acid deposition), primary PM_{2.5} (PM), and CO (contributes to ozone formation much like a VOC). The geographic location and type of precursor emission sources must be considered. Distance from receptor areas of interest is important, as well as direction. Further, differentiating between source types such as elevated versus low level may prove of interest. To illustrate how the process works, suppose we are interested in how ozone levels at Shenandoah National Park (SNP) are affected by elevated NOx emissions from Ohio Valley (OV). Ozone concentration at SNP would be c_i in the discussion above and all the elevated NOx emissions from the specified region would be grouped as parameter P_j . With DDM-3D, URM would not only calculate the ozone concentration at SNP, c_i , but its sensitivity, s_{ij} , to elevated NOx emissions from OV. In other words, we would know by how much the ozone concentrations would change when the base value P_j of those emissions are perturbed. This is invaluable information that could only be obtained through multiple model runs if the so-called "brute-force" method were used instead of DDM-3D. We are able to calculate approximately 20 such sensitivities by doubling the computation time for concentrations alone. To obtain the same information through the brute-force method would require thousands of model runs, each time with a different value of precursor emissions of interest.

An example of the results of a DDM-3D application are presented in Figure 2-2. In Figure 2-2, the sensitivity of the local ozone field to a 1% increase in the domain-wide ozone initial conditions is presented for the third day of the 27-29 August 1987 ozone episode in the Southern



California Air Basin (Los Angeles, CA). For example, the location of the peak ozone is identified by the '+' symbol (i.e. grid cell 62, 26). The ozone sensitivity at this point is 0.015 ppb per 1% increase in the domain-wide ozone initial conditions. Therefore, a 20% increase in the ozone initial conditions is expected to result in a 0.3 ppb increase (20×0.015) in the ozone at grid cell (62,26).

The use of DDM-3D will potentially eliminate the need to conduct one or more of the sensitivity analyses that are described in Section 5.4 of this document. In particular, the sensitivity analyses associated with modified boundary conditions, wind speeds, and emissions will likely need not be conducted since the use of DDM-3D will produce more robust results in a much more computationally efficient manner.

Figure 2-2. Example results (ozone sensitivity to ozone initial conditions) of a DDM-3D application over the Southern California Air Basin (Los Angeles, CA) -- third day of a three day episode (27-29 August 1987). Scale represents a change in *local* ozone (ppb) per 1% *domain-wide* increase in ozone initial conditions.

2.5 Parallel versus Sequential Processing

Another feature of URM is that it has been parallelized for use on multiple workstations (Kumar et al. 1997). Specifically, the URM transport and chemistry modules have been constructed such that their work can be distributed to multiple machines or run sequentially on only one computer. This capability is very important to the success of the SAMI project. The parallel version of the URM will be used for the base case and baseline episode runs. The sequential version of the URM will be used for the sensitivity runs which will be used to construct isopleths. The reason for using the sequential version for the sensitivity runs is that, in essence, we would be modeling a number of runs in parallel, as opposed to parallelizing within a single

run. This is computationally faster since there is some overhead with running any model in a parallel mode.

3.0 MODELING DOMAIN AND DATA AVAILABILITY

This section identifies the SAMI modeling domain and grid specification for use in the URM. This domain is exactly the same as that used in the emissions modeling, but it is different than that used in the meteorological modeling for reasons discussed in the meteorological modeling protocol. Also, an overview of the data needs for the URM are discussed.

3.1 Modeling Domain

The proposed URM and emissions modeling domain is shown in Figure 3-1. Table 3-1 (subject to change) lists the air quality monitoring stations in the SAMI domain. Please note that Table 3-1 was constructed solely on the basis of the contents of the AIRS data base for 1995. The AIRS data base maintains air quality data for SO₂, VOC, NO, NO₂, NO_x, and O₃. Other air quality data of interest for the SAMI project is currently being collected. Further, some stations that are identified in Table 3-1 may not have been operating in years prior to 1995. The URM air quality modeling grid is a multiscale grid in the standard Mercator projection:

Projection:	standard Mercator
Units of measure:	meters
Origin of latitude:	26.0 degrees
Longitude of true scale:	-99.0 degrees
number of computational nodes:	4176 per layer
southwest corner (x, y):	0.0, 2680000.0
southeast corner (x,y):	3264000.0, 2680000.0
northeast corner (x,y):	3264000.0, 5368000.0
northwest corner (x,y):	0.0, 5368000.0
southwest corner (lon,lat):	-99.0, 25.997
southeast corner (lon,lat):	-66.399, 25.997
northeast corner (lon,lat):	-66.399, 47.355
northwest corner (lon, lat):	-99.0, 47.355

The domain presented in Figure 3-1 is subject to change based on the outcome of subsequent grid resolution tests and consultation with the SAMI TOC.

The SAMI stakeholders have expressed concern about the lack of finer resolution over the Ohio River Valley in Kentucky, Ohio, and West Virginia. We expect to model the large sources in this area via the URM plume-in-grid capability. Of note, the Ohio River Valley may be one of a few source areas that will be modeled via a coupled coarse grid and plume-in-grid treatment. The URM plume-in-grid capability allows us to model important sources in an area whose sources will otherwise be modeled using a coarse resolution grid. If a finer resolution grid were extended over the Ohio River Valley, there would likely be no need to use the plume-

in-grid capability since the finer resolution grid would capture the important details of the plume evolution (e.g. as shown by Kumar et al. 1993 over the northeastern United States). Further, extending a finer resolution grid over the Ohio River Valley is likely to result in significantly longer URM run times. Therefore, plume-in-grid treatment of important sources in the Ohio River Valley coupled with the use of a coarse resolution grid over the same area allows us to keep URM execution times down but at the same time allow us to capture important plume dynamics from suspected sources that impact the SAMI domain. If the plume-in-grid treatment proves to be ineffective in capturing the Ohio Valley plumes, we will try using finer resolution grids as well as the plume-in-grid model coupled with a finer resolution grid.

Table 3-1. Air Quality Monitoring Sites In The SAMI Region.

(State)			Parameters Monitored									
County Name	Site ID	Site Name	NO	NO ₂	NO _x	O ₃	NMHC	Elements	NO3	SO4	Deposi	

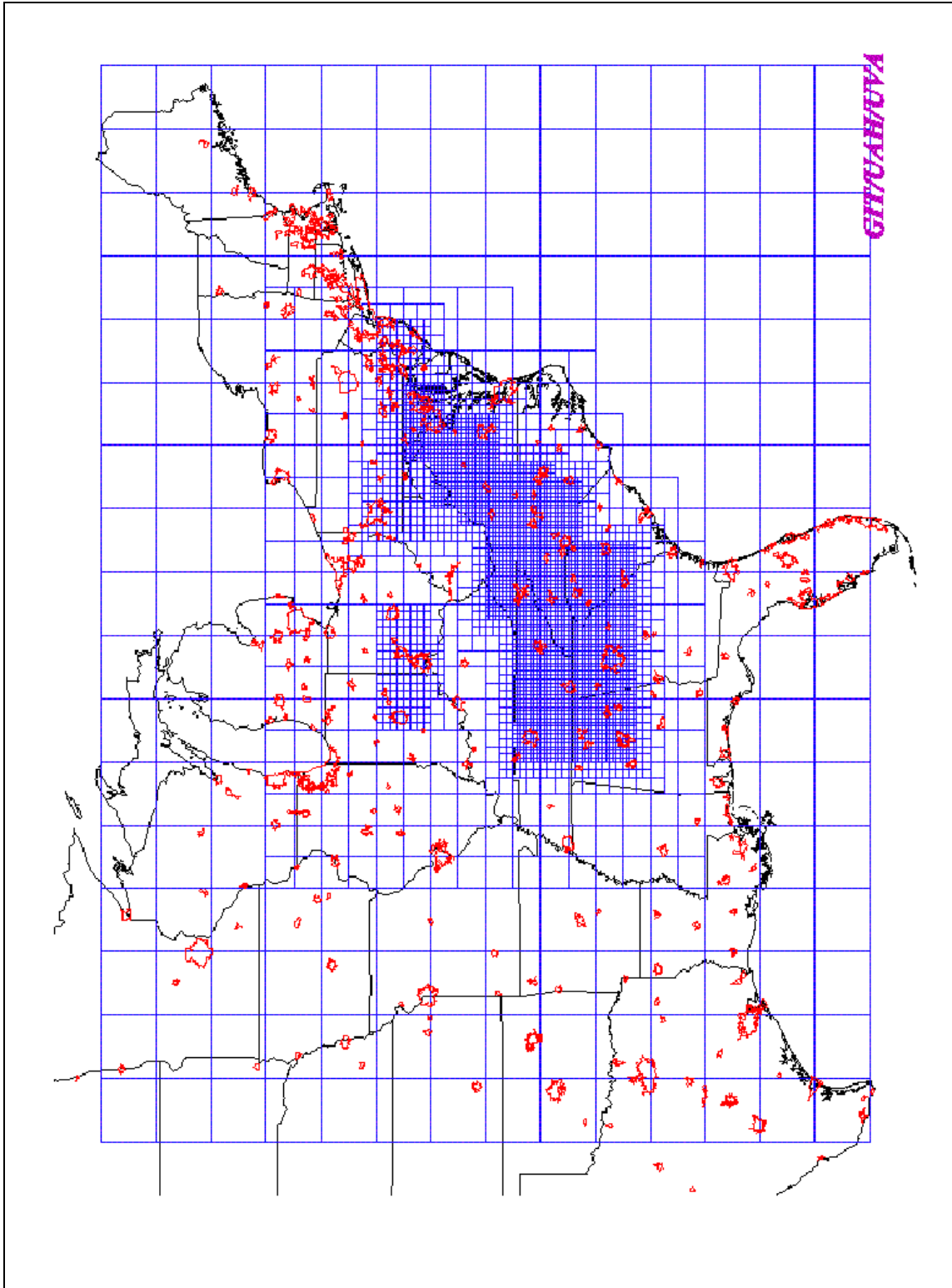


Figure 3-1. Proposed URM Modeling Grid.

Seven (7) grid layers are proposed in the vertical direction. Grid resolution tests are expected to measure the sensitivity of pollutant concentrations to vertical structures with fifteen

(15) layers. Further, the fifteen layer grid resolution tests are also meant to examine the possible impact of stratospheric ozone intrusion due to tropopause folding. Table 3-2 shows the preliminary estimates at the thicknesses of these seven and fifteen layer structures. As with the horizontal grid structure, the vertical grid structure is subject to change based on the results of the meteorological modeling and grid resolution tests. The increased vertical resolution, over standard air quality modeling practice, is aimed at mitigating the artificial dilution of pollutants, especially elevated point source plumes that are not treated through plume-in-grid (PIG) modeling, in the model when few, relatively coarse resolution vertical layers are used (Tesche et al., 1992).

The use of finer resolution near the surface of the domain (as compared to the more coarse resolution aloft) allows us to capture the steeper concentration gradients that typically exist in the near-surface troposphere. Further, the finer resolution in the near-surface troposphere allows us to capture the evolution of the mixing depths during the day. Again, however, the final decision on the horizontal and vertical grid structure will not be made until the preliminary air quality modeling and the meteorological modeling results are diagnosed.

Table 3-2. Initial Estimate For The Seven And Fifteen Layer Vertical Depths.

Number of Layers	Total Depth (meters)	Structure (Cell thicknesses in meters)
7	6000 m	20, 80, 200, 400 800, 1600, 2900
15	14,350 m	20, 80, 200, 200, 200, 200, 300, 300, 300, 400, 500, 900, 2000, 2000, 2000

3.2 Data Availability

3.2.1 Emissions Data

Emissions data for use in URM are derived from EMS-95. The SAMI emissions modeling protocol provides details on the source and processing of the emissions data.

3.2.2 Air Quality Data

Table 3-1 lists the air quality monitoring sites in the SAMI region. Aerometric data from these NAMS and SLAMS stations are in the process of being obtained from AIRS. The air quality data are used in the development of URM model inputs and in evaluating the model's performance. Additional air quality data is expected to be obtained from PAMS, NARSTONE, and IMPROVE archives. The data availability is significantly influenced by the episode choices.

AIRS. AIRS data will be used to set boundary conditions and to initialize the model for the photochemically-related pollutants. AIRS data will also be used for model performance evaluation.

PAMS. PAMS data, when and where available, will be used to set boundary conditions and to initialize the model for the photochemically-related pollutants. PAMS data will also be used for model performance evaluation. An added advantage of the PAMS data over the AIRS data will be the speciated VOC data though the PAMS data is limited in spatial extent and to episodes after 1995.

NARSTO-NE. The NARSTO-NE intensive data will be used in much the same way that the PAMS data is anticipated to be used. This study has been conducted during 1995 and 1996 summers.

IMPROVE. The IMPROVE network was developed and implemented to collect particulate matter data for use in visibility assessment in Class I areas. Speciated data has been taken at a limited number of sites in the SAMI area, as well as the rest of the eastern United States being modeled. (There are more in the western United States, though that information will not be used here.) IMPROVE sites in the eastern United States were added in 1991.

3.2.3 Meteorological Data

Meteorological data for use in URM are derived from RAMS and MM5. The SAMI meteorological modeling protocol provides details on the source and processing of the meteorological data.

3.3 Quality Assurance

The emissions modeling protocol and the meteorological modeling protocol discuss the QA procedures associated with each of their respective assignments. However, errors are inevitably discovered during the air quality modeling runs. Most external problems are generally associated with the emissions. When such emissions problems arise and are appropriately identified, the problems in the input data files are corrected and selected modules of EMS-95 are rerun. In the more uncommon case when meteorological problems are identified, RAMS (or MM5) is rerun. Other more subtle errors are typically discovered during the model evaluation phase. During the model evaluation, the emissions estimates and air quality modeling results are run through MAPS© which produces graphical displays of the emissions estimates and air quality fields over the modeling domain and temporal graphs of the same. These graphs are examined for anomalous values (e.g. localized emissions over a lake, or high emissions in an rural setting), and when discovered, the problems in the emission input files are determined, corrected, and selected modules of EMS-95 rerun. The meteorological, air quality, and land-use inputs are plotted and examined to ensure: (a) accurate representation of the observed data in the URM-ready fields, and (b) temporal and spatial consistency and reasonableness.

4.0 URM INPUT DATA

The proposed URM modeling domain (Figure 3-1) is a multiscale grid with horizontal cell sizes in the east-west and north-south directions that range from 192 kilometers (coarse grid structure) to 12 kilometers (fine grid). Intermediate horizontal cell sizes are 96 kilometers, 48 kilometers, or 24 kilometers. The proposed vertical structure is described in Section 2. All data inputs are prescribed at the nodes (i.e. the intersections of the grid cells). The URM data inputs and the source of the data are described in the remainder of this section. All data formats for the files that are described in this section can be found in McRae et al. (1992).

4.1 Emissions

A primary determinant of pollutant concentration levels within an airshed is the emission of contaminant materials into the atmosphere. These emissions, which can be produced from a variety of different activities, enter the airshed model either through the boundary conditions or as direct source terms in the atmospheric diffusion equation. This subsection briefly describes the area source emissions, which enter the URM as a surface level boundary condition, and the elevated point source emissions which enter the URM directly as source term unless treated explicitly by the plume-in-grid module. All emissions are specified in units of $ppm \cdot sec^{-1}$ and are specified per the SAPRC90 chemical mechanism. Furthermore, the emissions fields are also spatially and temporally resolved.

4.1.1 Area Source Emissions

Area source emissions estimates will be comprised of standard area sources, biogenics, mobile sources, and low-level point sources. The area source emissions estimates will be prepared by EMS-95 for direct input to URM. Please refer to the emissions modeling protocol for details on the development of the area source emissions estimates.

4.1.2 Elevated Point Source Emissions

Elevated point source emissions estimates will be comprised of the point sources whose effective plume heights exceed twenty (20) meters (i.e. the surface grid cell) and the point sources which will be treated explicitly in the URM plume-in-grid module. The elevated point source emissions estimates will be prepared by EMS-95 for direct input to URM. Please refer to the emissions modeling protocol for details on the development of the elevated point source emissions estimates.

4.2 Air Quality

Air quality data is needed by the URM to specify the initial and boundary conditions for the atmospheric diffusion equation. The URM needs three air quality files to run: the surface level initial conditions; the upper levels boundary and initial conditions; and the day zero upper

levels initial conditions which is used only to initialize the first ramp-up day. In summary, the modeling domain is demarcated by urban, rural, and major water bodies in order to use different total hydrocarbon splitting factors; ground level initial conditions are interpolated to a vertical profile where necessary; and boundary conditions are specified at the north, south, east, and west faces of the modeling domain and a zero-flux boundary condition is applied at the top of the modeling domain. Finally, all air quality data is spatially and temporally resolved.

4.2.1 Surface Level Initial And Boundary Conditions

Most initial condition data are derived from ground level measurements that have been interpolated to the modeling domain. A distance-weighted interpolation or a kriging procedure is used to prepare the fully specified initial condition field. These data are used to derive the initial conditions in the vertical. All the observed air quality data in the SAMI domain, including the SLAMS and NAMS (i.e. AIRS air quality data base), PAMS, NARSTO-NE, and IMPROVE as well as any special study data, are used to specify the initial conditions.

4.2.2 Upper Levels Boundary And Initial Conditions

The vertical distribution of the initial and boundary conditions in the modeling domain are established by one of two methods. The first method sets the initial and boundary conditions in each vertical cell by explicitly specifying concentrations. The second method uses an auxiliary observed air quality field which contains concentrations measured in the late afternoon on the previous day to set initial and boundary conditions aloft. The observed data can be used with different scaling factors for each vertical cell. Boundary conditions at the top of the modeling domain are assumed to follow a zero-diffusive flux condition. This approach minimizes the impact of upper level boundary condition setting on model results. As part of this project, we plan to develop an upper level boundary condition setting procedure that can account for the influx of stratospheric ozone based on potential vorticity.

4.3 Meteorological Data

The meteorological fields that are used by URM have a major impact on pollutant dispersion and transport. Six meteorologically related files are used in URM: the mixing depths; the wind field; the temperature field; the absolute humidity field; the total solar radiation scaling field; and the ultraviolet radiation scaling field. All the meteorological fields are spatially and temporally resolved.

4.3.1 Mixing Depths

The mixing depths define the vertical elevation where atmospheric stability changes. Note that because the URM uses terrain following coordinates, the mixing depths are measured

from ground level and not mean sea level. This depth can be internal to a computational cell and are temporally variable. The vertical extent of the modeling domain is sufficiently high to incorporate the mixing height variations. The mixing depths are diagnosed directly from the K_{zz} field which is computed by the meteorological model. In URM, the mixing depth is used for various parameterizations; one example is the plume-rise module. Preprocessing (i.e. format change, units conversion, and elevation transformation) of the meteorological output is necessary to convert the data to a form suitable for use by URM.

4.3.2 Wind Field

The wind field describes the advective and vertical flow. Advective flow is the primary means of horizontal transport of pollutants. The horizontal and vertical winds are fully specified by the meteorological model. Minor preprocessing (i.e. format change) of the meteorological output is necessary to convert the data to a form suitable for use by URM.

4.3.3 Surface Temperature Field

The temperatures are needed to determine the thermochemical rate constants in the chemistry module. The temperatures are reported as dry (not wet) bulb values. The temperatures are fully specified by the meteorological model. Minor preprocessing (i.e. format change) of the meteorological output is necessary to convert the data to a form suitable for use by URM.

4.3.4 Surface Absolute Humidity

The absolute humidity field provides a direct measure of the water vapor concentration in the air. The absolute humidity field is fully specified by the meteorological model. Minor preprocessing (i.e. format change) of the meteorological output is necessary to convert the data to a form suitable for use by URM.

4.3.5 Total Solar Radiation

The total solar radiation scaling field in conjunction with the wind field and surface roughness characteristics is used to compute the atmospheric stability. The clear sky total solar radiation values at the surface are computed based on standard astronomical formulation (McRae et al., 1981). The scaling field is computed based on measures of the total solar radiation at the surface and the fraction of sky cover (i.e. cloud cover) which is derived from meteorological modeling results. Preprocessing (i.e. format change, units conversion, and elevation transformation) of the meteorological output is necessary to convert the data to a form suitable for use by URM.

4.3.6 Ultraviolet Radiation

Default ultraviolet radiation values are based on the time zone and latitude of the airshed. If the elevation of the node is greater than 1000 meters, the ultraviolet radiation values need to be scaled to reflect the increase in actinic irradiance with height above sea level. The scaling factors are determined by solving the radiative transfer equations.

4.4 Surface Deposition

An important process that affects the predicted concentrations is the interaction of the pollutants with the earth's surface. Deposition takes two forms: wet and dry. The dominant process that influences dry surface deposition is turbulent transport. However, once the pollutants get near the surface, surface uptake is controlled by two factors: molecular diffusion and chemical resistance. Wet deposition is wholly controlled by precipitation.

4.4.1 Land Use

The land use file describes the major use characteristics of each cell. The URM currently recognizes seven major land use types: urban, agricultural, rangeland, forest, water, wetlands, and barren. The seven major categories are split into thirty-two subcategories which roughly correspond to the USGS standard land use categories. Each subcategory has an assigned surface roughness value which is used to determine the surface roughness height data that is described below. The last major update to the USGS standard land use data file was in 1976; however, the USGS has for been continually updating this data since 1976. Further, the USGS has indicated over the past few years that a entirely new land use data base was being built which should represent post-1990 land use configurations. The USGS standard land use data will be examined to determine its suitability for use in the SAMI project. Also, the same land use data base will be used in the meteorological, emissions, and air quality modeling -- a subtle quality control aspect that has been overlooked in probably all air quality studies.

4.4.2 Surface Resistance

Dry deposition chemical surface resistance is described for each land use category and pollutant for five total solar radiation ranges. This value represents the potential to uptake pollutants through dry deposition. Dry deposition chemical surface resistance is explicitly defined for ozone and SO₂. For species other than ozone and SO₂, the dry deposition chemical surface resistance is either specified as a factor of another species, typically SO₂, or as absolute values if the species do not exhibit a seasonal dependence or the seasonal dependence is not known. As part of the SAMI project, the standard surface resistance values will be reviewed to determine their suitability for use in this study.

4.4.3 Surface Roughness

Surface roughness heights are defined for each land use category. As part of the SAMI project, the standard surface roughness values will be reviewed to determine their suitability for use in this study.

4.4.4 Average Surface Roughness

The average cell surface roughness is a measure of the overall land use in the grid cell. As with surface resistance, this value represents the potential for that cell to uptake pollutants through dry deposition. This value is computed as a summation of the land use fraction weighted values of the natural logarithms of the land use surface roughnesses.

4.5 Grid Definition

The proposed modeling domain is described in Section 3.1. As shown in Figure 3.1, URM is capable of computing on a grid where the “patches” of high resolution do not necessarily have rectangular boundaries. It is therefore possible to increase the resolution only in areas of greatest interest (sources and receptors) without the burden of a high-density grid for areas in between. This results in significant computational savings, especially for the SAMI domain where the primary area of interest (i.e., the Southern Appalachian Mountains) and the coastline are oblique to the domain. However, most of the input to the model (e.g., meteorological and emissions data) are provided on grids with rectangular boundaries. For example, the RAMS model will provide meteorological data on rectangular “nests” (see Figure 2 in Meteorological Modeling Protocol). It is URM’s responsibility to pick and choose the data for the nodes where air quality computations will be conducted. Two grid definition files are necessary in order to do this: one defining the grid for the input data and another that defines the computational grid. We typically define air quality model inputs over a larger domain and with higher density than is necessary since we generally do not have *a priori* knowledge as to the resolution and extent of the grid for “best” computational results. After a couple of air quality model runs have been made, we are in a better position to identify areas that have little or no impact on the study area of interest. Once such areas are identified, the computational grid can be readjusted and subsequent air quality model runs can be made with no or little computations over those areas. The decision to exclude areas from the computational exercise will be made in consultation with the SAMI TOC.

4.6 Chemistry

The chemical mechanism that is used in URM is the SAPRC90 mechanism (Carter, 1990). Although a newer version of the SAPRC chemical mechanism exists, SAPRC93, there is concern that the SAPRC93 mechanism may exhibit bias under certain circumstances (Carter, 1993). We propose to use the SAPRC chemical mechanism over the CB-IV chemical mechanism because the SAPRC mechanism is considered the state-of-the-science. The SAPRC chemical mechanism has better evolved with time more so than many of the tropospheric chemical mechanisms including CB-IV. The SAPRC mechanism better represents

the dynamics of the emissions inventory since the mechanism is based partially upon the mix of the primary emitted species -- something the CB-IV chemical mechanism does not do.

The chemistry expansion effort is underway to add the reactions necessary to support the heterogeneous chemistry that is necessary for the SAMI project. Table 4-1 lists the species that will be used in the final version of the URM chemistry. Please note that the species that are directly emitted are identified with a check mark in the *Primary Emissions?* column. Also, the THC and RHC splits that are used in the chemistry will be reviewed to determine their suitability for use in the SAMI project.

Of note, the isoprene chemistry that is part of SAPRC90 is considered the best science currently available. In fact, the revised isoprene chemistry of the CB-IV chemical mechanism is based in part on the work that went into the development of the SAPRC90 isoprene chemistry.

Table 4-1. SAPRC90 Chemical Species.

Species	Description	Primary Emission?
NO	NITRIC OXIDE	
NO2	NITROGEN DIOXIDE	
O3	OZONE	
HONO	NITROUS ACID	
HNO3	NITRIC ACID	
HNO4	PERNITRIC ACID	
N2O5	NITROGEN PENTOXIDE	
NO3	NITRATE RADICAL	
HO2	HYDROPEROXY RADICAL	
CO	CARBON MONOXIDE	
HCHO	FORMALDEHYDE	
MEK	METHYL ETHYL KETONE	
MGLY	METHYLGLYOXYL	
PAN	PEROXY ACETYL NITRATE	
RO2	TOTAL RO2 RADICALS	
RCO3	CH3CO3 RADICAL	
ETHE	ETHENE	
CRES	CRESOLS AND OTHER ALKYL PHENOLS	
NPHE	NITROPHENOLS	
HO2H	HYDROGEN PEROXIDE	
C	CARBON ATOMS	
LN	LOST NITROGEN ATOMS	
OOH	LUMPED HYDROPEROXY SPECIES	
RRP	RO2-RO2-PRODUCT	
RHP	RO2-HO2-PRODUCT	
OLRI	OLE-RI, O ATOM REACTIONS WITH OLEFINS	
O3SB	O3OL-SB, REPRESENTS CONVERSION OF SO2 TO SO3 TMB - 1, 2,4-TRIMETHYLBENZENE	
MEOH	METHANOL	

ETOH	ETHANOL	
MTBE	METHYL-T-BUTYLEETHER	
TMC	2,2,4 TRIMETHYLPENTANE	
GLY	GLUOXAL	
RNO3	ALKYL NITRATES, RONO2	
GPAN	GLYOXYL DEVELOPED PAN	
PHEN	PHENOL	
TOLU	TOLUENE	
BALD	BENZALDEHYDE AND OTHER AROMATIC ALDEHYDES	
PBZN	PEROXY BENZOYL NITRATE	
AFG1	AROMATIC RING FRAGMENTS 1	
AFG2	AROMATIC RING FRAGMENTS 2	
CCHO	ACETALDEHYDE	
RCHO	PROPIONALDEHYDE AND ALL HIGHER ALDEHYDES	
ACET	ACETONE	
PPN	PEROXY PROPYL NITRATE	
ETBE	ETHYL-T-BUTYLEETHER	
MECY	ME-CYCC5, METHYLCYCLOPENTANE	
MCYC	3MCYCPTC (3 METHYLCYCLOPENTENE)	
ETHA	ETHANE	
BUTA	BUTANE	
MEC5	2-ME-C5, 2-METHYLPENTANE	
BENZ	BENZENE	
CBEN	C2BENZ (ETHYLBENZENE)	
OXYL	OXYLENE	
MPXY	METHYLHYDROPEROXIDEXYLENE (MP-XYLENE)	
PRPE	PROPENE	
BUTD	13BUTD (1,3 BUTADIENE)	
M1BT	2-METHYL-1-BUTENE	
ISOP	ISOPRENE	
M2BT	2-METHYL-2-BUTENE	
AAR1	GENERAL ALKANE AND AROMATICS	
AAR2	GENERAL ALKANE AND AROMATICS	
AAR3	GENERAL ALKANE AND AROMATICS	
OLE1	GENERAL ALKENES	
OLE2	GENERAL ALKENES	
NH3	AMMONIA	
NIT	AEROSOL NITRATE	
SO2	SULFUR DIOXIDE	
SO3	SULFUR TRIOXIDE, RAPIDLY FORMS H2SO4	
APNE	A-PINENE	
UNKN	UNKNOWN	
PRPA	PROPANE	
OSD	O*1D2, O SINGLET D	
O	O ATOM	
HO	HYDROXYL RADICAL	
CCO	CCO-O2 RADICAL	
C2CO	C2CO-O2 RADICAL	

BCO2	BZ-CO-O2 RADICAL	
RO2N	ALKYL NITRATE RO2 RADICAL	
RO2X	RO2-XN RADICAL	
RO2P	RO2-NP, PHENOL RO2 RADICAL	
RO2R	GENERAL RO2 #1 RADICAL	
R2O2	GENERAL RO2 #2 RADICAL	
COCO	HCOCO-O2 RADICAL	
HCO3	HOCOO RADICAL	
BZO	PHENOXY RADICAL	
BZNO	BZ(NO2)-O	
O2	OXYGEN	
H2O	WATER VAPOR	
CH4	METHANE	
CO2	CARBON DIOXIDE	
H2	HYDROGEN	
K	POTASSIUM	
NA	SODIUM	
CA	CALCIUM	
MG	MAGNESIUM	

4.7 Command Files

There are two URM command files that are prepared for each modeling day. The first command file directs the execution of URM, and the second command file tells URM where all the input files are located and where all the output files are to be stored.

5.0 MODEL PERFORMANCE EVALUATION

5.1 Principles

It is important to establish guidelines for determining whether the air quality modeling system (i.e., the emissions, meteorological and dispersion models and their supporting data sets) has performed well enough for the SAMI base case simulations to serve as the basis for baseline development and EMO development and testing. In this following section, a framework for determining whether subsequent use of the URM is appropriate based on the model performance evaluation results and conclusions.

The Model Should Be Viewed As A System. When a *model* is rejected or accepted, this is meant in a broad sense. For purposes of model evaluation, the *model* includes not only the air quality model (URM), but also the other components of the system -- EMS-95 (the emissions model), the RAMS (and MM5) meteorological model, the supporting aerometric and emissions data base, and any other related analytical and numerical procedures used to produce modeling results. In fact, studies to date suggest that model performance is more sensitive to the inputs, than the model formulation and structure. Further, other studies suggest that the inherent spatial variability leads to significant deviations between point measurements used in evaluation and the spatially averaged concentrations that are being simulated. Part of the model testing process is to identify flawed model components. Of note, it is possible for the meteorological and emissions modeling results to perform well, but the entire episode can be rejected from further consideration if the air quality modeling results fail to meet performance evaluation specifications.

Model Acceptance Is A Continuing Process of Non-Rejection. The use of explicit or implied model "acceptance" criteria is avoided. Modeling systems should be accepted gradually as a consequence of successive non-rejections. Over time, confidence in a model builds as it is exercised in a number of different applications (hopefully involving stressful performance testing) without encountering major or fatal flaws that cause the model to be rejected.

Previous Experience Is Used As A Guide. Previous air quality modeling experience serves as a principal guide to set criteria for model rejection. Interpretation of the SAMI air quality modeling results against the backdrop of previous regional-scale modeling experience will call attention to the existence of potential performance problems and will suggest whether the model should be tested further or rejected. However, it should be noted that although a number of region-scale photochemical modeling studies have been conducted for which such an interpretation can be made for ozone and its precursors, there is a notable dearth of studies on a regional-scale for acid deposition and regional haze.

Thus far, the only criteria proposed for model rejection are those proposed by Tesche (1992) for the SARMAP and LMOS studies, by Reynolds and Wagner (1993) for the

Southeast Michigan Ozone Study (SEMOS), and by Tesche et al., (1995) in the Wasatch Front Ozone Modeling Study (WFOS). Please note that the aforementioned studies are all at the urban-scale. Tesche et al. (1990) rejected the idea of setting "model performance acceptance criteria". Instead, a general framework consisting of "rejection criteria" was established that specifically calls for a case-by-case analysis of the evaluation results. This approach has been adopted at least in part by the California Air Resources Board (ARB) in their statewide photochemical modeling guidelines (ARB, 1992) and in EPA's national guidelines on UAM-IV modeling (EPA, 1991). The general criteria are as follows (Tesche et al., 1990):

"The model should produce peak (unpaired) estimation accuracy, overall bias, and gross error statistics from the investigative simulations, the other numerical measures and diagnostic tests suggest unusual or aberrant behavior. If the simulation results fall within these ranges, additional diagnostic analyses... .. should be carried out to determine whether the simulation should be rejected. If a simulation falls outside of any one of these general ranges, it becomes incumbent upon the modeler to explain why the performance is poorer than that commonly achieved in past applications. It will also be necessary to explain whether the causes of poorer performance will adversely affect the use of the model in control strategy evaluations. Otherwise, the particular base case in question should be judged inadequate."

However, those criteria were based on historical modeling, and they did not fully account for the spatial inhomogeneities. Further, compensatory errors may have been used to improve performance. For example, it is now widely believed that the emissions from many source categories were biased low in historical modeling studies, and this may have been compensated for by increasing boundary conditions. Recent studies have shown that using the "official" inventory which did not account for the underestimates in emissions led to drastically low simulated ozone, and poor performance, but after inventory adjustment, performance was acceptable. These discoveries came after the original performance criteria were adopted, suggesting that such criteria must be reconsidered, if any rigid criteria should be used.

Model Rejection Criteria Must Remain Flexible. While the process for model testing is specified here, the proposed SAMI air quality modeling rejection criteria must remain flexible to accommodate this new application. The rejection criteria recommended in subsection 7.6 are based on past experience. It is unknown what specific problems are likely to arise in the SAMI air quality modeling study, but provisional rejection criteria which was developed on the basis of past experience have been constructed. The criteria suggested may need to be modified as experience is gained during the course of this study.

5.2 URM Model Evaluation Process

Performance testing of the URM will follow a step-wise process patterned after the ideal process outlined by Reynolds et al., (1994) and Roth et al., (1995). This evaluation

process is tailored to the schedule and data base constraints imposed by the current study. The URM evaluation will consist of three steps.

5.2.1 Step 1: Mechanistic Evaluation

The first step entails a *mechanistic evaluation* to explore the behavior of individual process modules within the URM. This step is necessary since the URM will undergo numerous modifications to incorporate an advanced wet deposition capability, heterogeneous chemistry, and enhanced aerosol fate processes. The purpose of the mechanistic evaluation is to identify possible flaws and/or systematic biases that may not be apparent when examining the model as a whole.

The various modules which comprise the URM have been extensively tested prior to their incorporation into the overall URM framework. As the direct result of the URM, and in particular its parent model the California/Carnegie Institute of Technology (CIT) air quality model, being publicly available, each major module has been investigated in one or more university or research organization projects. Accordingly, considerable confidence has been gained that major conceptual or coding flaws are not present in the model.

Again, however, since the URM will undergo additional modifications to support this project, additional mechanistic tests may be required of the new modules to be added to URM. At this time, no testing of the wet deposition, the heterogeneous chemistry, or aerosol processes module will be performed since the modeling components upon which these modules are based have been examined elsewhere (e.g. Berkowitz et al., 1989; Wexler and Seinfeld, 1991).

5.2.2 Step 2: Initial Screening

An initial screening of the URM modeling results for each episode is to be conducted. The initial screening consists of an examination of a number of statistical measures and graphical aids which are used to identify obviously flawed model simulations. Once identified, steps are taken to implement improvements to the model input files in a logical, defensible manner so that performance might be improved. If the initial screening suggests that no obvious flaws or compensating errors exist in the simulation(s), the performance evaluation moves to Step 3.

Visual Inspections. A number of graphical procedures will be used in the initial screening of the URM simulation results. For example, specific graphical displays to be produced and analyzed for each URM base case simulation include:

- > Spatial mean pollutants (e.g. ozone, PM_{2.5}, SO_x- wet deposition) time series plots;
- > Ground-level pollutant isopleths;
- > Concentration scatterplots;
- > Bias and error stratified by concentration; and
- > Bias and error stratified by time.

Rejection of Obviously Flawed Simulations. Initial screening of the URM results should allow the modeling team and the other stakeholders to identify obviously flawed simulations. Experience in air quality modeling is the best basis upon which to identify obviously flawed simulation results.

Model Performance Improvement. Efforts to improve URM model performance (i.e., to reduce the discrepancies between model estimates and observations) will be based on sound scientific principles. *Curve-fitting* or *tuning* activity will be avoided. The following principles will govern the model performance improvement process (to the fullest extent possible given the project schedule):

- > Any significant changes to the model or its inputs must be fully documented, both in hard copy and magnetic media where appropriate;
- > Any significant changes to the model or its inputs must be supported by scientific evidence, analysis of new data collected for the purpose, or by reanalysis of the existing data where errors or misjudgments may have occurred; and
- > All proposed changes to the model or its inputs will be reviewed by the Science Advisory and Policy Committee and should be reviewed by the stakeholders.

The following activities will be performed should it be necessary to improve URM model performance:

- > Typify the major problems with the base case simulation (e.g., peak ozone underestimation, timing problems, precursor underestimation, spatial offsets);
- > Identify model algorithms, process descriptions, or model parameterizations whose inadequate or incomplete characterization could potentially produce the apparent discrepancies;
- > Examine whether and to what extent modifications to preprocessor inputs, model algorithms, or model parameters can be made without moving outside the model's intended range of applicability, input uncertainty, scientific credibility, or reasonableness;
- > Assess whether the proposed changes may generate or exacerbate compensating errors (if they exist) in the simulation;
- > Implement any model or input changes and carry out the indicated runs;
- > Analyze the model results;
- > Determine if adequate performance has been achieved. If so, document the results of the diagnostic analyses and model improvement cycle(s) and submit the memorandum to the stakeholders for review. If not, recommend what to do next, such as:
 - perform another cycle of diagnosis and improvement;

- recommend the selection of another module that provides greater promise for achieving adequate performance;
- recommend that efforts be redirected to assess performance for another episode that was not selected for initial evaluation; (i.e., the initial episode[s] may have problems that cannot be fully resolved with the existing data base and the stakeholders may wish to determine whether adequate performance can possibly be achieved for another episode); or
- recommend that the objectives of the modeling program and associated model performance expectations be suitably revised.

Given the large uncertainties in both the emissions and meteorological data bases in the SAMI region, it is conceivable that one or more of the modeling episodes may not achieve satisfactory performance. If this occurs, alternative modeling episodes may have to be selected for use. While definitive criteria for rejecting an episode for inadequate model performance are difficult to prescribe in advance, suggested criteria are presented in section 5.5. These criteria are consistent with those use in other SIP studies around the country.

5.2.3 Step 3: Operational/Scientific Evaluation

If the initial air quality modeling results examined under Step 2 are free from obvious flaws, a more formal, and extensive evaluation of the simulation results will be conducted. Step 3 of the URM model performance evaluation consists of three tasks. First, not only will the graphical displays utilized in Step 2 be generated for precursor species (e.g. NO, NO₂, SO₂, PM_{2.5}, and PM₁₀), but the suite of analyses will be expanded to include those listed in Table 5-1. An examination of the VOC and PM species will be limited by the lack of speciated VOC and PM data from the SAMI domain. The graphical displays for the precursor species will be examined for obvious flaws. Should these be detected, the model performance improvement process may be implemented to identify and correct (if possible) the noted problems. The second and third steps involve the computation and assessment of several statistical and numerical metrics and the performance of specific model sensitivity experiments.

Table 5-1. Statistical Measures and Graphical Displays to be Used in the Evaluation of the URM Model.

Statistical Measure	Graphical Display
Maximum observed concentration	Modeled and observed spatial mean concentrations as a function of time
Maximum modeled concentration and deposition rates	Measures of peak estimation accuracy (A_{TS} , A_T , A_S , A_U , A)
Maximum modeled concentration and deposition rate at a monitoring station	Normalized bias as a function of time
- Ratio of maximum modeled to observed concentrations	Normalized gross error as a function of time

- Ratio of modeled to observed deposition rates	
Accuracy of peak estimation (paired in time and space)	Normalized bias as a function of concentration level and deposition rate
Accuracy of peak estimation (unpaired in time and space)	Normalized gross error as a function of concentration level and deposition rate
Average accuracy over all stations	Scatterplot of hourly concentration and deposition rate pairs
Normalized bias in hourly concentrations and deposition rates	Scatterplot of daily maximum concentration and deposition rate pairs
Mean bias in hourly concentrations and deposition rates	Quartile plots of hourly species concentrations
Normalized gross error in hourly concentrations and deposition rates	Daily maximum ground-level concentration and deposition isopleths
Mean gross error in hourly concentrations and deposition rates	
Variance in hourly concentrations and deposition rates	
Percent difference in mean observed and modeled aloft concentrations from the NARSTO and other special studies (e.g., ozone, NO, NO _x) if data are available.	

Once the base case for the first episode is established, the same procedures used to develop it will be applied to subsequent episodes. In other words, the model will be run in a *hands off* mode for the other episode(s). If inadequate performance results for subsequent episodes and if diagnostic performance improvement activities are employed to develop better base cases, then the procedures used for these later episodes must then be applied to the first episode. Note that this is likely to be the case since, the ozone, acid deposition, and regional haze episodes will probably be independent episodes. That is, each episode is likely to exhibit one, possibly two, air quality events.

5.3 Statistical and Graphical Tools to Be Used in the Evaluation

The operational evaluation process includes the calculation and analysis of several routine statistical measures and the plotting of specific graphical displays to characterize the basic performance attributes of the model. The scientific evaluation focuses on the model's performance in simulating various precursor and product species as well as the performance of a set of sensitivity and uncertainty simulations. Among the statistical measures to be examined in the operational evaluation are: different measures for characterizing the model's accuracy in estimating the hourly average concentrations and deposition rates; mean normalized bias to indicate the degree to which calculated one-hour concentrations and deposition rates are over- or underestimated; the variance, describing the dispersion of the residual distribution about the mean; and the mean normalized gross error, which quantifies the average absolute signed deviation of the concentration residuals. These measures will be calculated using the MAPS© software, described in Appendix A.

5.4 Investigative Simulations To Be Considered In The Scientific Evaluation

Investigative simulations are an essential component of the URM scientific evaluation and will be conducted as an integral part of this study. Investigative simulations serve much the same purposes as sensitivity simulations which are described later in this document:

- > Reveal internal inconsistencies in the model;
- > Provide basis for compensatory error analysis;
- > Reveal the parameters (or inputs) that dominate (or do not dominate) the model's operation;
- > Reveal error propagation through the model; and
- > Provide guidance for model refinement and data collection programs.

Up to six (6) investigative simulations may be performed for each episode as part of the URM performance evaluation. The exact nature of these simulations will be discussed with the stakeholders after the operational evaluation results have been reviewed. Potentially useful investigative simulations include:

- > Zero emissions;
- > Clean background;
- > Modified boundary conditions;
- > Zero dry deposition;
- > Modified turbulent diffusion;
- > Modified wind speeds;
- > Modified motor vehicle emissions estimates; and
- > Modified biogenic emissions estimates.

Please note that depending on the outcome of the incorporation of DDM-3D, a sensitivity method developed by Yang et al. (1997) specifically for use in three dimensional air quality models, into URM, separate air quality model runs for each of the following *investigations* may not be necessary since URM with DDM-3D will produce the same results as a consequence of the standard base case, base year, baseline, and baseline with EMOs air quality model runs.

Zero Emissions. The main purpose of the zero emission model run is to ensure that the base case simulation results are influenced appropriately by the emissions inputs. Reducing (or zeroing out) all the emissions (i.e point, area, mobile, and biogenics) should lead to much lower reactive species concentrations after the initial simulation day. The zero emissions simulation is performed by exercising the base case run with all emission values reduced to zero. All other model input files remain unchanged from the base case.

The results of the zero emissions simulation, and the other investigative simulations, are presented in three ways. First, a Deficit-Enhancement (D-E) plot is developed for each simulation. The D-E plot is developed by subtracting the metric of interest (e.g. hourly-average concentrations or deposition rates, daily maximum concentrations or deposition rates) from

those of the base case and plotting the residuals as isopleths. Time series plots for the run may also be developed. These are constructed with the base case estimates represented by a solid line and the zero emissions sensitivity run results depicted by a dashed line. Finally, a tabular listing may be developed of all monitoring stations and the maximum estimated concentrations or deposition rates (unpaired in time) in the base case and the zero emissions sensitivity run.

The zero emissions investigative run should produce significantly-reduced concentrations and pollutant deposition rates, close to background or to the levels reflective of the inflow boundary conditions. If not, there is reason to question the adequacy of the simulation. Lack of sensitivity to emissions may indicate inappropriately high initial conditions, improper boundary conditions, or some flaw in the model itself. Quite apart from these concerns, insensitivity to emissions raises serious questions about the usefulness of the simulated episode for EMO development and assessment in the SAMI region.

Clean Background Initial Conditions. The clean background initial conditions simulation reveals how much of the second (or third) day model estimates result from the initial field used to start the simulation. This simulation is performed by setting all initial concentrations to those typical of the clean background air. The reason for using clean background conditions, as opposed to zeroing the initial conditions is that the latter procedure may cause numerical instabilities.

Deficit-enhancement and time series plots may be used to display the results of this simulation. If the initial field is completely "washed out" of the model domain by the second or third day, the D-E plots will indicate essentially no differences between the sensitivity and base case runs on the following day(s). For stagnation episodes, some residual effects of initial conditions may be seen even on the third day of the simulation.

Background Level Boundary Conditions. The background-level boundary condition simulation examines the influence of boundary values on second (or third) day concentrations, particularly in regions where the base case estimates are highest. This simulation helps identify situations where the base case results are "driven" by the boundary conditions that are set above background levels. The background level boundary conditions simulation is performed by setting all inflow and outflow boundary values to low background values. This includes the top surface of the modeling region. Deficit-enhancement and time series plots will be used to display the simulation results. If the episode is influenced by transport, the D-E and time series plots should reveal significant impact of the boundary conditions in the interior of the computational domain.

Zero Dry Deposition. The zero dry deposition simulation addresses the influence of dry deposition on primary and secondary species concentrations. The zero deposition investigative run will be exercised by setting dry deposition velocities for all species to zero in URM and re-running the base case simulation. This entails a simple change to the URM input file. Dry deposition tests have not been regularly reported in previous URM evaluation studies

so little historical information exists to serve as a guide in interpreting the results of this investigation. Nevertheless, some general guidelines can be suggested. For primary species such as NO_x and VOCs, the downwind concentration fields should increase relative to the base case in a manner consistent with the deposition velocities for each primary species. For reactive species such as ozone, increases and decreases may occur in the D-E contour fields depending upon the interaction between ozone and NO fields.

Increased Turbulent Diffusion. Turbulent diffusion has a direct and often significant influence on pollutant concentrations. In this case, the vertical region of high turbulent diffusion will be increased, representing an increase in the mixed layer by that amount. The objective of the mixing depth sensitivity simulation is to reveal the degree to which pollutant concentrations are influenced by the depth of the mixed layer. One run may be performed in which the hourly mixing depths values are uniformly increased by 50% above the base case values. As such, this simulation should provide a bound on the change in pollutant estimates resulting from uncertainties in this input.

Increased mixing depths typically reduce pollutant concentrations although the reduction is less than a one-to-one change. One might choose, instead, to reduce the hourly mixing depths by 50%. The resultant increase in ozone concentrations under this scenario will typically be comparable in magnitude but of opposite sign as those for the mixing depth increase case. The results of this simulation will be presented graphically with D-E plots and with conventional time series plots.

Reduced Wind Speeds. Another investigative model simulation worth considering entails a 50% reduction in the magnitude of the wind speeds input to the URM. More complex and potentially very insightful windfield sensitivity runs might be conceived (e.g., random perturbation of node speeds and directions to reflect known systematic or compensatory errors in the meteorological model) but an initial characterization of the model's sensitivity to ventilation is an appropriate starting point. This simulation may be performed by decreasing all wind components by 50% and re-running the air quality model. Based on previous windfield sensitivity results, the magnitude of the hourly pollutant concentrations, including the peak value, should increase relative to the base case although the percentage increase should be less than proportional with wind speed reduction. Results of the windfield investigative simulation will be displayed through D-E plots and by time series plots.

Increased Motor Vehicle Emissions. A number of studies in California (Ingalls, 1989; Pierson et al., 1989; Fujita et al., 1991; Wagner and Wheeler, 1992) have indicated that on-road motor vehicle VOC emissions may be underestimated. This bias can be expected to lead to an underestimation in model estimated ozone concentrations. A simulation will be considered that involves increasing the motor vehicle VOC emissions by an across-the-board factor. Model estimates for ozone, nitrogen dioxides and SAPRC hydrocarbon species will be compared for this sensitivity simulation with those from the base case simulation using D-E plots and time series plots.

Decreased Biogenic Emissions. Although the recent body of work by Guenther et al. (1993), Geron et al. (1991), Williams et al. (1991) suggests that biogenic emissions are underestimated, the new biogenic emissions estimates formulation proposed by the same researchers is believed to result in significant overestimates of the biogenics. In a recent study, Wilkinson (1997) showed that by using the biogenic emissions estimates formulation of Guenther et al. (1993) and Williams et al. (1991), biogenic emissions in the South Coast Air Basin (Los Angeles, CA) may be overpredicted by an order of magnitude. However, in the same study, Wilkinson (1997) shows that the biogenic emissions are highly uncertain, and that the true biogenic emissions are likely captured with the new formulation.

In the SAMI study, biogenic emissions will play a significant role in ozone and aerosol formation. Therefore, an investigative run that examines the impact of biogenic emissions on pollutant concentrations should be conducted. These runs will be made on the basis of the biogenic emissions uncertainty analysis that will be conducted as a component of the emissions modeling. Results of the biogenic emissions investigative simulations will be displayed through D-E plots and by time series plots.

The above numerical, graphical, and investigative procedures constitute the set of operational tests that are suggested for each base case episode. If model performance is found to be poorer than expected, additional diagnostic and mechanistic tests and sensitivity tests (e.g., analysis of biogenic emissions sensitivity) should be considered in order to improve the simulations. The budget and schedule implications of these additional runs will be discussed with the stakeholders before any additional work is performed.

5.5 Provisional Rejection Criteria

With the principles offered in section 5.1 as a guide, the following criteria are proposed against which the URM hourly averaged ozone results should be judged. However, rejection criteria are not proposed for acid deposition, regional haze, and precursor species since the data for such criteria are limited and the review of this data is not complete.

Peak Unpaired Estimation Accuracy. The unpaired peak estimation accuracy, A_u , for each day following model initialization should: (a) be less than or equal to 20%, (b) be derived from a model estimate within 48 kilometers of the peak observation location, and (c) be based on a peak model estimate occurring within ± 2 hours of the maximum observation.

Normalized Bias. The mean normalized bias (MNBE) for each day following model initialization should be less than or equal to 20%, based on an ozone cutoff level of 40 ppb.

Normalized Gross Error. The mean normalized gross error for each day following model initialization should be less than or equal to 35%, based on an ozone cutoff level of 40 ppb.

Diurnal Bias. For each day following model initialization, the mean normalized bias, when plotted as a function of time of day, should not reveal major discrepancies (i.e., bias \geq 30%) within any time interval two hours or longer.

Diurnal Gross Error. For each day following model initialization, the mean normalized gross error, when plotted as a function of time of day, should not reveal major discrepancies (i.e., error \geq 50%) within any time interval two hours or longer.

Bias as a Function of Concentration Level. For each day following model initialization, the cumulative mean normalized bias, when plotted as a function of time of day, should not reveal major discrepancies (i.e., bias \geq 30%) over any concentration interval 20 ppb or larger.

Gross Error as a Function of Concentration Level. For each day following model initialization, the cumulative mean normalized gross error, when plotted as a function of time of day, should not reveal major discrepancies (i.e., error \geq 50%) over any concentration interval 40 ppb or larger.

Time Series. For each day following model initialization, the hourly ozone time series plots consisting of (a) the four-cell weighted average and (b) the model uncertainty envelope, (based on the highest and lowest estimates 1 cell from the monitor) should not exhibit unusual or inexplicable trends. Examples of such trends would include (a) a "bimodal" pattern at several monitors in the observed or estimated time series coupled with the absence of such a pattern in the other time series, (b) strongly "spiked" or broadly "flat" patterns at several monitors in the one time series but not the other, or (c) significant differences in the respective rates of rise or fall of the diurnal ozone profiles when these are not attributable to sub-grid scale (e.g., titration) effects.

Ground-Level Concentration Distributions. For each day following model initialization, hourly-averaged ground level ozone isopleths (containing the hourly ozone measurement at each monitor) shall not exhibit clearly discrepant or aberrant patterns. An example of such a pattern would be the existence of estimated ozone concentrations on the grid that exceeded by 25% or more the highest estimated concentration at a monitor (in the absence of a plausible, physically-based explanation.) If the maximum modeled concentration on the URM grid is significantly different from the peak monitored concentration and the two maxima are widely separated in space, this suggests the existence of potentially important errors in the modeled transport, emissions, or chemical processes. Confounding errors in all three processes are also likely in such a case and should be the cause for concern and additional model diagnosis.

Correlation of Estimate-Observation Pairs. For each day following model initialization, scatterplots of hourly-averaged ozone estimate-observation pairs should not exhibit a significant number of "outlier" points. In addition, scatterplots of the daily maximum ozone

estimate-observation pairs should not exhibit any outlier data points. The definition of outlier points and what constitutes a significant number of them must be determined on a case-by-case basis. However, an estimate-observation pair shall not be considered an outlier if there is plausible evidence that the discrepancy is the result of sub-grid scale effects or physical/chemical phenomena not addressed in the model's formulation.

5.6 Domain-wide Versus Subdomain Performance Evaluation

The model performance evaluation will not be conducted solely on a domain-wide basis nor will episodes be rejected solely on the basis of domain-wide model performance results. Instead, a number of subdomains will be established for which model performance will be evaluated. The episode rejection determination will be based on the results of all domains for which model performance is conducted. By “carving” the domain into numerous subdomains and investigating model performance in each subdomain, it may be possible to isolate stray performance. Further, this may help to identify gross problems in model inputs and possibly in model formulation.

A preliminary assessment of analysis subdomains is as follows:

- > north southern Appalachian Mountains;
- > south southern Appalachian Mountains;
- > northwest quarter area (i.e area immediately to the northwest and surrounding the southern Appalachian Mountains);
- > northeast quarter area (i.e area immediately to the northeast and surrounding the southern Appalachian Mountains);
- > southwest quarter area (i.e area immediately to the southwest and surrounding the southern Appalachian Mountains);
- > southeast quarter area (i.e area immediately to the southeast and surrounding the southern Appalachian Mountains);
- > all other northwest areas;
- > all other northeast areas;
- > all other southwest areas; and
- > all other southeast areas.

6.0 BASE CASE, BASELINE, RESPONSE SURFACES, AND EMO MODELING

6.1 Base Case

6.1.1 Episodes

6.1.2 Expected Results

6.2 Baseline

6.2.1 Episodes

6.2.2 Expected Results

6.3 Response Surfaces

6.3.1 Use of DDM-3D

6.3.2 Expected Results

6.4 Emissions Management Options (EMOs)

6.4.1 Episodes

6.4.2 EMO Development

Standard Methods.

Use of DDM-3D.

6.4.3 Expected Results

6.5 Expected Publications

7.0 REFERENCES

APPENDIX A -- MODEL EVALUATION PROCEDURES

MODEL EVALUATION PROCEDURES

Before the URM is applied to emission management option (EMO) design for the SAMI region, the URM model performance is to be tested via an operational/scientific evaluation. This provides some assurance to the SAMI stakeholders that the model is producing the right answer for the right reasons. The operational evaluation process includes the calculation and analysis of several routine statistical measures and the plotting of specific graphical displays to characterize the basic performance attributes of the model. The scientific evaluation focuses on the model's performance in simulating various precursor and product species as well as the performance of a set of investigative simulations. Among the statistical measures to be examined in the operational evaluation are: different measures for characterizing the model's accuracy in estimating the maximum one-hour average concentration; mean normalized bias to indicate the degree to which calculated one-hour concentrations are over- or underestimated; the variance, describing the dispersion of the residual distribution about the mean; and the mean normalized gross error, which quantifies the average absolute signed deviation of the concentration residuals. These measures are calculated using the MAPS© software.

The Model Performance Evaluation, Analysis, and Plotting Software (MAPS©) system package was developed for urban- and regional-scale meteorological, emissions, and photochemical model evaluations. The MAPS© system embodies all of the recommended statistical and graphical model testing methods for photochemical and meteorological models suggested by Tesche et al., (1990a) and subsequently adopted by the California Air Resources Board (ARB) and EPA in their photochemical modeling guidance documents (ARB, 1992; EPA, 1991). In addition, MAPS© contains statistical and graphical tools for analyzing emissions model estimates. MAPS© consists of a set of special-purpose FORTRAN codes, the National Center for Supercomputer Applications (NCSA) Hierarchical Data Format (HDF) data management libraries (ported to SUN and IBM RS/6000 platforms) and National Center for Atmospheric Research (NCAR) Graphics, Version 3.01.

Formulation of the MAPS© statistical measures and graphical procedures (generally applicable in both air quality and meteorological model evaluation) are presented in this section. Examples of the use of several of these procedures in meteorological and regional air quality modeling studies are presented in Anthes et al., (1987), Steyn and McKendry (1988), Dennis et al., (1990), Schere and Wayland (1991), Tesche et al., (1990a), and McNally (1994). In some of the definitions below, the variable \hat{c} represents a model-estimated or derived quantity, e.g., wind speed, wind direction, PBL height, ambient temperature. The subscripts e and o correspond to model-estimated and observed quantities, respectively. The subscript i refers to the ith hour of the day.

A.1 Mean and Global Statistics

Several statistical measures are calculated to provide an overall summary of the model estimates and observations and to support calculation of other statistical measures.

Mean Estimation (M_e). The mean model estimate is given by:

$$M_e = \frac{1}{N} \sum_{i=1}^N \phi_{ei}$$

where N is the product of the number of simulation hours and the number of ground-level monitoring locations providing hourly-averaged observational data. ϕ_{ei} represents the model estimate at hour i.

Mean Observation (M_o). The mean observation is given by:

$$M_o = \frac{1}{N} \sum_{i=1}^N \phi_{oi}$$

Here, ϕ_{oi} represents the observations at hour i.

Average Wind Direction. Because wind direction has a crossover point between 0 degrees and 360 degrees, standard linear statistical methods cannot be used to calculate the mean or standard deviation. Evaluations by the EPA (Turner, 1986) suggest that the method proposed by Yamartino (1984) performs well in estimating the wind direction standard deviation. Specifically, this quantity is calculated by:

$$\sigma_{\alpha} = \arcsin(\beta) \left[1 + 0.1547 \beta^3 \right]$$

where:

$$\beta = \left\{ \left[\frac{1}{N} \sum_{i=1}^N \left(\overline{\sin \alpha} \right)^2 + \frac{1}{N} \left(\overline{\cos \alpha} \right)^2 \right] \right\}^{1/2}$$

$$\overline{\sin \alpha} = \frac{1}{N} \sum_{i=1}^N \sin \alpha_i$$

$$\overline{\cos \alpha} = \frac{1}{N} \sum_{i=1}^N \cos \alpha_i$$

Here, α_i is the measured hourly or instantaneous wind direction value.

Standard Deviation of Estimation (S_{de}). The standard deviation of the model estimates is given by:

$$S_{de} = \left\{ \frac{1}{N} \sum_{i=1}^N \left(\phi_{ei} - M_e \right)^2 \right\}^{1/2}$$

Standard Deviation of Observations (S_{do}). The standard deviation of the observations is given by:

$$S_{do} = \left\{ \frac{1}{N} \sum_{i=1}^N \left(\phi_{oi} - M_o \right)^2 \right\}^{1/2}$$

Least Square Slope and Intercept Regression Statistics. A linear least-squares regression is performed to calculate the intercept (a) and slope (b) parameters in the following equation:

$$\hat{\phi}_{ei} = a + b \phi_{oi}$$

This regression is performed for each set of hourly (or instantaneous) data to facilitate calculation of several error and skill statistics.

Maximum Ratio (R_{max}). The maximum ratio is defined as the quotient of the maximum one-hour averaged model estimated concentration and the maximum hourly-averaged measurement, i.e.,

$$R_{max} = \frac{c_e(x, t)}{c_o(\hat{x}, \hat{t})}$$

where c_e is the estimated one-hour averaged pollutant concentration, c_o is the observed hourly averaged concentration, \hat{x} refers to the peak monitoring station location, \hat{t} is the time of the peak observation. The caret, $\hat{}$, denotes the time or location of the maximum observed concentration. There is no requirement that the maximum estimated and observed concentrations be paired in either time or space but for this measure we require that the maximum modeled concentration be taken from a monitoring station.

A.2 Difference Statistics

Residual (d_j). For quantities that are continuous in space and time (i.e., wind speed, temperature, pressure, pbl height, species concentrations) difference (or residual) statistics are very useful. Difference statistics are based on the definition of a residual quantity. A concentration residual, for example, is defined as:

$$d_i = \{c_e(x_i, t) - c_o(x_i, t)\}$$

where d_j is the i -th residual based on the difference between model-estimated (c_e) and observed (c_o) concentration at location x and time i .

Standard Deviation of Residual Distribution (SD_r). The standard deviation of the residual distribution is given by:

$$SD_r = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (d_i - MBE)^2}$$

where the concentration residual is defined as:

$$d_i = \{c_e(x_i, t) - c_o(x_i, t)\}$$

and MBE is the first moment, i.e., the mean bias error, defined shortly. This statistic describes the "dispersion" or spread of the residual distribution about the estimate of the mean. The standard deviation is calculated using all estimation-observation pairs above the cutoff level. The second moment of the residual distribution is the variance, the square of the standard deviation. Since the standard deviation has units of concentration, it is used here as the metric for dispersion. The standard deviation and variance measure the average "spread" of the residuals, independent of any systematic bias in the estimates. No direct information is provided concerning subregional errors or about large discrepancies occurring within portions of the diurnal cycle although in principle these, too, could be estimated.

Accuracy of Peak Model Estimates (A). Five related methods are used to evaluate the accuracy of the model's estimate of the maximum value of a spatially-distributed variable. This may be, for example, temperature, wind speed, pressure, or concentration. In the

definitions below we use the peak one-hour average concentrations for discussion purposes; however, these measures may be applied to several of the meteorological variables as well.

Several accuracy measures are used because there are different, informative, and plausible ways of comparing the peak measurement on a given day with model estimates. These five accuracy measures provide complimentary tests of the model's performance. When applied to ozone simulations, they are particularly useful from a regulatory perspective since they deal with peak ozone (or precursor) concentration levels.

Paired Peak Estimation Accuracy. The paired peak estimation accuracy, A_{ts} , is given by:

$$A_{ts} = \frac{\sum \{ |c_o(x, t) - c_e(x, t)| \}}{\sum \{ c_o(x, t) \}} \times 100\%$$

A_{ts} quantifies the discrepancy between the magnitude of the peak one-hour average concentration measurement at a monitoring station, $c_o(x, t)$, and the estimated concentration at the same location, $c_e(x, t)$, and at the same time, t . Model estimates and observations are thus "paired in time and space." The paired peak estimation accuracy is a stringent model evaluation measure. It quantifies the model's ability to reproduce, at the same time and location, the highest observed concentration during each day of the episode. The model-estimated concentration used in all comparisons with observations is derived from bi-linear interpolation of the four ground level grid cells nearest the monitoring station.

A_{ts} is very sensitive to spatial and temporal misalignments between the estimated and observed concentration fields. These space and time offsets may arise from spatial displacements in the transport fields resulting from biases in wind speed and direction, problems with the "timing" of photochemical oxidation and removal processes, or subgrid-scale phenomena (e.g., ozone titration by local NO_x emission sources) that are not intended to be resolvable by grid-based photochemical models.

Temporally-Paired Peak Estimation Accuracy. The temporally-paired peak estimation accuracy, A_t , is given by:

$$A_t = \frac{\sum \{ |c_o(x, t) - c_e(x, t)| \}}{\sum \{ c_o(x, t) \}} \times 100\%$$

A_t quantifies the discrepancy between the highest concentration measurement at a monitoring station and the highest model estimate at the same station or any other grid cell within a distance of, say, 25 km. This measure examines the model's ability to reproduce the highest observed concentration in the same subregion at the correct hour.

Spatially-Paired Peak Estimation Accuracy. The spatially-paired peak estimation accuracy, A_s , is given by:

$$A_s = \frac{\sum \{ |c_o(x, t) - c_e(x, t)| \}}{\sum \{ c_o(x, t) \}} \times 100\%$$

A_s quantifies the discrepancy between the magnitude of the peak one-hour average concentration measurement at a monitoring station and the highest estimated concentration at the same monitor, within 3 hours (before or after) the peak hour.

Unpaired Peak Estimation Accuracy. The unpaired peak estimation accuracy, A_u , is given by:

$$A_u = \frac{\sum \{ |c_e(x, t) - c_o(x, t)| \}}{\sum \{ c_o(x, t) \}} \times 100\%$$

A_u quantifies the difference between the magnitude of the peak one-hour average measured concentration and the highest estimated value in the modeling domain, whether this occurs at a monitoring station or not. The unpaired peak estimation accuracy tests the model's ability to reproduce the highest observed concentration anywhere in the region. This is the least stringent of the above four peak estimation measures introduced thus far. It is a weak comparison relative to the previous ones but is useful in coarse screening for model failures. This measure quickly identifies situations where the model produces maximum ozone concentrations in the air basin that significantly exceed the highest observed values within the network.

Average Station Peak Estimation Accuracy. The average station peak estimation accuracy, A_{si} , is given by:

$$A_{si} = \frac{1}{N} \sum_{i=1}^N |A_{si}|$$

where:

$$A_{si} = \frac{\sum \{ |c_e(x_i, t) - c_o(x_i, t)| \}}{\sum \{ c_o(x_i, t) \}} \times 100\%$$

Here, x_i is the i th monitoring station location. A_{si} is calculated by first determining the spatially-paired peak estimation accuracy, A_{si} , at each monitoring station. Thus, the average station peak estimation accuracy is simply the mean of the absolute value of the A_{si} scores, where the temporal offset between estimated and observed maxima at any monitoring station does not exceed three hours.

Mean Bias Error (MBE). The mean bias error is given by:

$$MBE = \frac{1}{N} \sum_{i=1}^N \{ (c_e(x_i, t) - c_o(x_i, t)) \}$$

where N equals the number of hourly estimate-observation pairs drawn from all valid monitoring station data on the simulation day of interest.

Mean Normalized Bias Error (MNBE). The mean normalized bias error, often just called the bias, is given by:

$$MNBE = \frac{1}{N} \sum_{i=1}^N \left\{ \frac{(c_e(x_i, t) - c_o(x_i, t))}{c_o(x_i, t)} \right\}$$

Mathematically, the bias is derived from the average signed deviation of the concentration residuals and is calculated using all pairs of estimates and observations above the cutoff level.

Cutoff levels of 4-6 pphm for ozone and 2 pphm for NO_2 are often used in modeling studies to reduce the influence that low measured or modeled concentrations (often occurring at night or on the upwind boundaries) have on the normalized bias statistics. In regions of exceptionally high ozone, e.g., the South Coast Air Basin, cutoff levels as high as 10 pphm are commonly used.

Mean Absolute Gross Error (MAGE). The mean gross error is calculated in two ways, similar to the bias. The mean absolute gross error is given by:

$$MAGE = \frac{1}{N} \sum_{i=1}^N \{ |c_e(x_i, t) - c_o(x_i, t)| \}$$

Mean Absolute Normalized Gross Error (MANGE). The mean absolute normalized gross error is:

$$MANGE = \frac{1}{N} \sum_{i=1}^N \left\{ \frac{|c_e(x_i, t) - c_o(x_i, t)|}{c_o(x_i, t)} \right\} \times 100\%$$

The gross error quantifies the mean absolute deviation of the concentration residuals. It indicates the average unsigned discrepancy between hourly estimates and observations and is calculated for all pairs above the cutoff level. Gross error is a robust measure of overall model performance and provides a useful basis for comparison among model simulations across different air basins or ozone episodes. Unless calculated for specific locations or time intervals, gross error estimates provide no direct information about sub-regional errors or about large discrepancies occurring within portions of the diurnal cycle.

Root Mean Square Error (RMSE). The root mean square error is given by:

$$RMSE = \left[\frac{1}{N} \sum_{i=1}^N \{ (\phi_{ei} - \phi_{oi})^2 \} \right]^{1/2}$$

The RMSE, as with the gross error, is a good overall measure of model performance.

However, since large errors are weighted heavily, large errors in a small subregion may produce large a RMSE even though the errors may be small elsewhere.

Systematic Root Mean Square Error (RMSE_s). A measure of the model's linear (or systematic) bias may be estimated from the systematic root mean square error given by:

$$RMSE_s = \left[\frac{1}{N} \sum_{i=1}^N \{ (\hat{\phi}_{ei} - \phi_{oi})^2 \} \right]^{1/2}$$

Unsystematic Root Mean Square Error (RMSE_u). A measure of the model's unsystematic bias is given by the unsystematic root mean square error, that is:

$$RMSE_u = \left[\frac{1}{N} \sum_{i=1}^N \{ (\phi_{ei} - \hat{\phi}_{oi})^2 \} \right]^{1/2}$$

The unsystematic difference is a measure of how much of the discrepancy between estimates and observations is due to random processes or influences outside the legitimate range of the model.

A "good" model will provide low values of the root mean square error, RMSE, explaining most of the variation in the observations. The systematic error, RMSE_s should approach zero and the unsystematic error RMSE_u should approach RMSE since:

$$RMSE^2 = RMSE_s^2 + RMSE_u^2$$

It is important that RMSE, RMSE_s, and RMSE_u are all analyzed. For example, if only RMSE is estimated (and it appears acceptable) it could consist largely of the systematic component.

This bias might be removed, thereby reducing the bias transferred to the photochemical calculation. On the other hand, if the RMSE consists largely of the unsystematic component ($RMSE_u$), this indicates further error reduction may require model refinement and/or data acquisition. It also provides error bars that may be used with the inputs in subsequent sensitivity analyses.

A.3 Skill Measures

Index of Agreement (I). Following Willmont (1981), the index of agreement is given by:

$$I = 1 - \frac{\sum_{i=1}^N (\text{RMSE}_i)^2}{\sum_{i=1}^N (|\text{PHI}_{ei} - \text{M}_o| + |\text{PHI}_{oi} - \text{M}_o|)^2}$$

This metric condenses all the differences between model estimates and observations into one statistical quantity. It is the ratio of the cumulative difference between the model estimates and the corresponding observations to the sum of two differences: between the estimates and observed mean and the observations and the observed mean. Viewed from another perspective, the index of agreement is a measure of how well the model estimates departure from the observed mean matches, case by case, the observations' departure from the observed mean. Thus, the correspondence between estimated and observed values across the domain at a given time may be quantified in a single metric and displayed as a time series. The index of agreement has a theoretical range of 0 to 1, the latter score suggesting perfect agreement.

RMS Skill Error (Skill_e). The root mean square error skill ratio is defined as:

$$\text{Skill}_E = \frac{\text{RMSE}_u}{\text{SD}_o}$$

Variance Skill Ratio (Skill_{var}). The variance ratio skill is given by:

$$\text{Skill}_{\text{Var}} = \frac{\text{SD}_e}{\text{SD}_o}$$

A.4 Graphical Tools

Many features of air quality and meteorological model simulations are best analyzed through graphical means. In addition to revealing important qualitative relationships, graphical displays also supply quantitative information. The main graphical displays that will be used to analyze the URM model performance results are as follows:

- > The relationships among the five accuracy measures;
- > The temporal correlation between estimates and observations;
- > The spatial distribution of estimated concentration fields;
- > The correlation among hourly pairs of estimates, observations and residuals;
- > The variation in bias and error estimates as functions of time and space; and
- > The degree of mismatch between volume-averaged model estimates and point measurements.
- > The distributional relationships between rank-ordered observations and rank-ordered model estimates.

Brief discussions of these plotting methods are as follows.

Accuracy Plots. Two accuracy plots are used. One depicts relationships between the peak five accuracy measures while the other plot summarizes the peak estimation accuracy at all monitoring stations, revealing the presence of subregional estimation bias if it occurs. The first

plot is a histogram that displays the calculated values of A_{ts} , A_t , A_s , A_u , and \cdot . The second plot is also a histogram showing the peak observed and estimated concentrations (unpaired in time) at each monitoring station above the cutoff concentration. Also contained on the plot is a shaded region corresponding to the normalized gross error.

Time Series Plots. Probably the most useful graphical procedure for depicting air quality model results is the time series plot. Developed for each monitoring station for which observed concentrations are available, this plot presents the hourly estimates and observations throughout the simulation period. The time series plot consists of the hourly averaged observations (boxes) and the hourly averaged estimates, the latter being fitted by a smooth continuous line. The model estimates are derived from bi-linear interpolation of the nearest four grid cells to the monitor. At each hour, the absolute value of the concentration residual will be calculated and plotted as a dashed line on the same plot.

With the time series plot one may determine the model's ability to reproduce the peak estimation, the presence or absence of significant bias and errors within the diurnal cycle, and whether the "timing" of the estimated concentration maximum agrees with the observation. By including the residual plot on the same graph, estimation biases are more apparent.

Spatial Time Series Plots. Conventional time series plots do not reveal situations where the model estimates concentrations comparable in magnitude to the observations a short distance away from the monitoring station. A second time series display, called a "spatial time series plot", is used for this purpose. These plots provide information about the degree to which model discrepancies result from the procedure for selecting the estimated values. There is no a priori reason to select the four-cell bi-linear average estimate over the estimate in the specific grid cell containing the monitor (i.e., the "cell value"), or perhaps the grid cell estimate within any of the four adjacent cells that is closest in magnitude to the observed value (i.e. the "best" estimate). Spatial time series plots are constructed for each monitoring station by plotting the hourly observations together with an envelope defined by the highest and lowest grid cell estimate within one cell of the monitoring station. MAPS© can easily examine multiple grid cell distances as well.

The spatial time series plots provide diagnostic information about the "steepness" of the concentration gradients in the simulated fields. A small envelope indicates relatively flat concentration gradients. Conversely, steep gradients may produce a fairly large envelope. Ideally, the measurement points will fall within the envelope. Spatial time series plots are one method of revealing the correspondence or "commensurability" between volume-averaged model estimates and point measurements.

Ground Level Isopleths. Ground-level ozone isopleths are developed for each hour of the episode to display the spatial distribution of estimated concentration fields. The isopleth plots are developed by computer-contouring the hourly, gridded ozone estimates. The information content of these plots are enhanced by including the following:

- > A base map identifying significant geophysical and political boundaries;
- > Locations of air monitoring stations;
- > The observed concentrations at each monitoring station by a bold numeral;
- > The location of the peak estimate (signified by an asterisk); and
- > The magnitude of the peak grid cell estimate.

Ground-level isopleths are also constructed based on the daily maximum concentration estimate in each grid cell. These "maximum" ozone isopleths supply direct information about the magnitude and location of pollutant concentrations and help to identify situations where sub-regional biases may be attributed to spatial misalignment of the estimated and observed concentration fields.

Scatterplots of Estimates and Observations. Scatterplots are a useful means of visually assessing the extent of bias and error in hourly ozone estimate-observation pairs. Hourly scatterplots are developed by plotting all hourly-averaged estimate-observation pairs for which the observed concentration exceeds the cutoff value. Similarly, daily maximum scatterplots are developed from the pairs of maximum hourly estimated and observed values at each monitoring station. The estimated maximum is the highest value simulated within three hours of the observed maximum. In these plots, the solid diagonal line with 1:1 slope will be used to identify the perfect correlation line and the dashed lines enclose the region wherein estimates and observations agree to within a factor of two. The lines of agreement can be made more stringent if desired.

The scatterplot is used to give a quick visual indication of the extent of over-or underestimation in the hourly estimates and whether there appear to be strong nonlinearities in model estimates and observations over the concentration range studied. Bias is indicated by the preponderance of data points falling above or below the perfect correlation line. The dispersion (spread) of points provides a visual indication of the general error pattern in the simulation. Scatterplots help identify outlier estimate-observation pairs, i.e., a seemingly discrepant estimate-observation pair that may result from erroneous data, a fundamental flaw in the model, or some other cause that requires investigation. These plots provide little diagnostic information about sub-regional performance problems, temporal or spatial misalignments, or other inadequacies in the simulation. In addition, scatterplots mask the temporal correlation between various estimate-observation pairs.

Scatterplot of Residuals and Observations. Residual scatterplots are developed to describe the distribution of hourly average model discrepancies (positive and negative) as a function of concentration level. This graphical display is constructed from the data elements that make up the bias and error calculations. Hourly concentration residuals for all monitoring stations are plotted as a function of observed concentration for all pairs above the cutoff value. A daily maximum residual plot is also constructed based on data pairs involving the maximum

observed concentration at a monitor station and the maximum estimated value at the same station within three hours of the peak.

Residual scatterplots are used to characterize estimation discrepancy throughout the observed concentration range. The plot does not reveal the existence or causes of sub-regional or timing performance problems. Absence of bias is suggested by no systematic tendency for the data points to fall above or below the ordinate; however, as noted previously, important subregional biases may still exist in the presence of a zero overall bias estimate.

Bias Stratified by Concentration. Bias-concentration plots are derived from the residual distribution to depict the degree of systematic bias in hourly-averaged model estimates (paired in time and space) as a function of observed concentration level. This plot (and the companion error-concentration plot) aids in model diagnosis. The observed concentration range is divided into several equal-sized concentration bins and the normalized bias within each bin is calculated and plotted as a function of concentration level. A smooth line is then fitted through the bin-averaged values. The bias-concentration plot is used to reveal the existence of under- or over-estimation throughout the concentration range.

Gross Error Stratified by Concentration. Gross error-concentration plots are derived from the residual distribution to depict the error in model estimation (paired in time and space) as a function of observed concentration level. The observed concentration range is divided into several equal-sized concentration bins. Then, the average value of the normalized gross error within each bin is calculated and the bin averages are plotted as a function of the observed concentration level. MAPS© will display the mean normalized gross error on the plot for easy reference.

The gross error-concentration plot is used to reveal the variation in model error at various intervals throughout the concentration range. The plot must be interpreted carefully, however, remembering that the concentration residual is normalized by the observed value.

Bias Stratified by Time. Bias-time plots are developed to help identify specific time periods within the photochemical simulation when systematic patterns of under- or overestimation occur. The bias-time plot is constructed in a manner similar to the bias-concentration plot, except that the simulation period is discretized into a number of time intervals, usually 1-2 hours in duration. Systematic bias in model estimates during specific periods within the diurnal cycle may have several causes: biases in vertical mixing or wind transport; "timing" problems with the chemistry; non-representative temporal distributions assumed in the emissions inventory, and so on. While the bias-time plots may not clearly pinpoint the causes of bias, they may be helpful in defining the time intervals when the bias is most apparent. This helps focus subsequent diagnostic investigations.

Gross Error Stratified by Time. Gross error-time plots are developed to help identify specific time periods when gross errors in the model estimates may be a problem. This

plot is constructed in a similar manner as the error-concentration plot, except that the simulation period is discretized into a number of time intervals, usually 1-2 hours in duration. When interpreting the gross error-time and bias-time plots, one must remember that the concentration levels of all pollutants vary throughout the diurnal cycle.

- I. **Quantile-Quantile Plots.** Quantile-quantile plots are cumulative frequency distributions that provide a graphical characterization of the distribution of observed and modeled values over their entire ranges. Quantitative information that can be obtained from these distributions include estimates of the mean, median, and standard deviation. The plots also provide a visual characterization of how the estimates and observations are spread out with respect to the central value. They also readily display unpaired bias.

- URM Description
 - A. URM
 - B. RSM Modifications
 - C. Aerosol Updates
 - D. DDM-3D Modifications
 - E. Parallel versus Sequential Processing
- II. Data Availability And Quality Assurance
 - A. PAMS
 - B. IMPROVE
 - C. NARSTO-NE
 - D. AIRS
 - E. Quality Assurance
- III. URM Input Preparation Procedures
 - A. Modeling Domain
 - B. Emissions Estimates
 - 1. US
 - 2. Canada (specify importance of including these emissions)
 - C. Derivation Of Meteorological Fields From RAMS Results
 - 1. Mixing Depths
 - 2. Winds
 - 3. Surface Temperatures
 - 4. Absolute Humidity
 - 5. Solar/UV Radiation
 - D. Fields Common To URM and RAMS
 - 1. Surface Roughness
 - 2. Topography
 - 3. Land Use
- IV. Model Performance Evaluation
 - A. Categories of Evaluation
 - 1. Mechanistic
 - 2. Operational
 - 3. Scientific
 - B. Statistical Measures
 - 1. MAPS©
 - 2. Mean And Global Statistics
 - 3. Skill Measures
 - 4. Graphical Measures
- V. Base Case, Response Surfaces, Baseline, and EMOs
 - A. Base Case
 - B. Response Surfaces
 - C. Baseline
 - D. EMOs
- VI. Special Studies

- A. Use Of DDM-3D For Sensitivity Analyses
- B. Use Of DDM-3D For Uncertainty Analyses
- VII. Expected Results
 - A. Base Case
 - B. Response Surfaces
 - C. Baseline
 - D. EMOs
 - E. Peer Reviewed Publications Expected
- References