CHAPTER 4

THE AIR-QUALITY MODELING SYSTEM: *A Critical Component Of The Scientific Toolbox For Policy-Maker*

4. The Air-Quality Modeling System

POLICY RELEVANCE: Air-quality models are the principal tools available for evaluating alternative emission-control strategies for reducing ambient tropospheric O_3 concentrations. Consequently, effective policy-making requires models that are accurate in their simulations of the atmosphere and whose limitations are well characterized. Inaccuracies impose uncertainties of unknown character on the modeling results that decision makers use in crafting control strategies.

4.1 THE IMPORTANCE AND NATURE OF AIR-QUALITY MODELING SYSTEMS

Grid-based air-quality models provide a critical tool for tropospheric-ozone analysis and management. In addition to other merits, their ability to synthesize and assemble multiple, interactive elements of the air pollutant system (or mathematical representations thereof), and thus allow them to be analyzed collectively, is a particularly attractive and useful feature.

Because of this unique capability, air-quality models have found extensive application as research tools for scientific analysis. In addition, they have come to occupy a central role in the development of O_3 mitigation strategies by the policy community. Because of their importance, air-quality models, and *air-quality modeling systems*, are given special emphasis in this assessment.

The term "Air-Quality Modeling System" (AQMS) denotes the collection of numerical models used to simulate the formation, accumulation, transport, and removal of primary and secondary air pollutants over urban and regional spatial scales, including those ancillary modules that control input, output and graphical presentation of results (see Textbox 4.1). As noted in Chapter 2, AQMS applications vary among the three nations of North America, playing a supportive role Canada and Mexico, while being a required part of the development of State Implementation Plans (SIPs) in the United States. Despite the noted importance of these modeling systems, questions continue to arise concerning the reliability of model output and the potential offered by modeling to meet future planning needs. This chapter focuses on key issues pertaining to the strengths and limitations of current-generation AQMS and their ability to meet the needs of the policy community.[a]

[a] This chapter is not intended to provide a comprehensive examination of the AQMS and its applications. Reviews on this subject have been provided by Russell and Dennis (**CR19**) and Roth et al. (**CR20**).

4.1 The Components of the AQMS

Air-quality models use a mathematical representation of physical and chemical processes to determine source-receptor relationships for atmospheric constituents, in particular those derived from pollutant emissions. Because these models solve the governing equations in space and time using numerical techniques and a computer, they are referred to as numerical models. In addition, air-quality models that use a three-dimensional grid of cells to provide reference points for approximating the spatial averages needed to render a solution practical are often referred to as being grid-based or Eulerian. This grid may be viewed as a fixed reference frame wherein materials move about the three-dimensional domain by horizontal and vertical exchange, typically between adjacent grid cells, and chemical reactions occur at rates governed by average reactant concentrations in individual grid cells. Finally, because the air-quality model actually consists of a system of coupled modules (or models), it is referred to as an Air-Quality Modeling System or AQMS[a]. As illustrated in Figure 4.1, the AQMS typically includes:

- A Geographical Information System (GIS) (and other preprocessor codes) for manipulating and inputting data to the various other models of the system.
- An emission model for estimating gridded biogenic and anthropogenic emissions as a function of time, based on emission factors and activity (see Section 3.3).
- A meteorological model that simulates atmospheric physical processes and estimates gridded fields of variables such as winds, clouds, temperature, and humidity for the time span of the modeled period. As discussed in Section 3.5, this model can involve the continuous input of observed meteorological data using 4-Dimensional Data Assimilation (4DDA).
- An air-quality model, composed of a chemical mechanism, surface deposition and transport model and a numerical solver, which predicts the concentrations of atmospheric constituents on the basis of inputs from the emission and meteorological models.
- A post-processor and graphical analysis package for displaying and analyzing the model results.

In addition, an AQMS often includes a user interface that allows the user to conveniently and efficiently interact and manipulate model inputs, codes, and analysis routines.

[a]The term air-quality model can be ambiguous, sometimes referring to the photochemical air-quality model alone and sometimes to the overall modeling system. In this report, we use the term AQMS when specifically referring to the total system, and reserve the simple term model for situations where the basic model, without its ancillary components, is discussed.

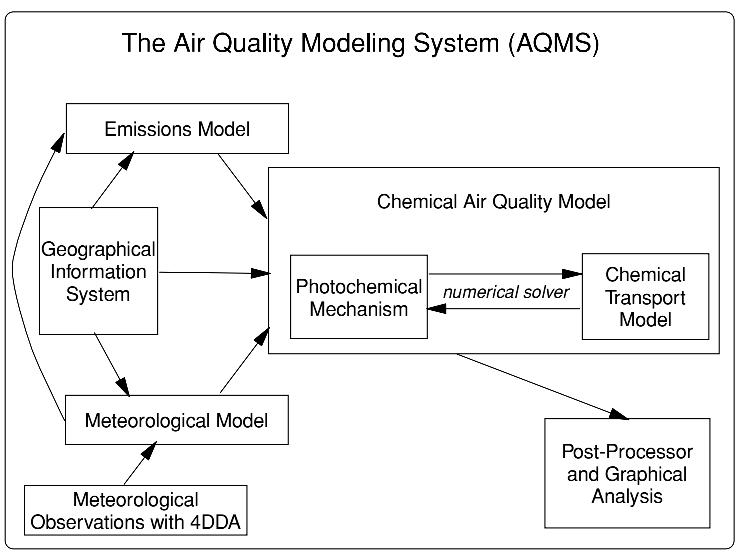

Figure 4.1 The Components Of A Typical AQMS.

4.2 KEY ISSUES PERTAINING TO AQMS APPLICATIONS

4.2.1 Issue: What are the principal needs that the AQMS is intended to address in policy-making?

Finding: The air-quality model is the only prognostic tool available to the policy-making community; i.e., a tool capable of quantitatively estimating future air-quality outcomes for conditions and/or emissions different from those that have existed in the past. Because of this unique capability, air-quality models have come to play a central role in the process of determining how pollutant emissions should be managed to achieve air-quality goals.

Discussion: Air-quality models are designed to calculate concentrations of O_3 and other pollutants and their variations in space and time in response to particular emission inputs and for specified meteorological scenarios. As explained in Textbox 4.2, the use of air-quality models in research applications and policy analyses can be dissimilar. Within the policy-making context they are used to evaluate the relative efficacy of alternative O_3 abatement strategies. More specifically, air-quality models are typically used to estimate:

- Patterns of O_3 concentration that result from specified emission patterns, be they historical, current, or future (e.g, to estimate exposure to specific populations).
- Changes in O_3 concentration patterns for specified changes in emissions of a particular compound or class of compound or a specific emission sector (e.g., to optimize emission control strategies).
- Relative roles of distant versus local emissions in producing enhanced O_3 concentrations (e.g., to assess the relative efficacy of local versus regional emission-control strategies).

Because air-quality models are imperfect representations of the real atmosphere, their output must have some level of error. For this reason, some

4.2 Dissimilarities in the Applications of the AQMS by the Scientist and Policy Makers

While the AQMS is a key tool of the policy maker, it is a product of and an important research tool of the scientist. However, the approach taken when using an air-quality model in a research mode can be quite different from that of the policy-making mode. For the scientist, air-quality models serve two broad functions. First, they serve as a means for synthesizing concepts, formulations, and information into a comprehensive numerical archive. Second, air-quality models provide a mechanism for elucidating the underlying physical and chemical processes responsible for the formation, accumulation, transport, and removal of air pollutants. To accomplish this, scientists often attempt to expose model weaknesses and inaccuracies by applying the model to conditions and situations that stress and challenge model parameterizations and algorithms and then comparing the resulting model outputs to appropriate real-world data. By diagnosing the root cause of the model's errors, the scientist hopes to develop a better understanding of the atmosphere and, by extension, of how to formulate more accurate and realistic model parameterizations and algorithms. Thus, in the context of scientific research, the inability of a given model application to faithfully reproduce observed data is not necessarily seen as a negative. However, the same cannot be said of the use of air-quality models in the policy-making context. In this case, the models are being used in a prognostic mode, and a prerequisite for such an application is typically a demonstration that the model is capable of reproducing observations from an historical episode. Thus, applications of air-quality models in policy-making tend to emphasize the model's ability to faithfully reproduce observations through appropriate adjustment and/or tuning of model inputs and parameterizations over some acceptable range of uncertainty.

measure of the uncertainty in air-quality model results should, in principle, be part of the results themselves. As discussed in more detail below, inability to quantify model uncertainty is one of the major shortcomings associated with the application of the current generation AQMS.

4.2.2 Issue: What advances in AQMS have taken place during the past decade?

Finding: Advances in AQMS have occurred in two broad categories:

1) The technical and scientific soundness of the parameterizations and algorithms used to simulate the relevant processes
2) The numerical/computational interface between the various AQMS components and between the model and the user.

Discussion: A number of advances have been made during the past decade in the scientific foundation of models, in the data supporting modeling applications, and in computational and data-handling capabilities. Principal among these advances are:

- Major improvements in the meteorological component of the modeling systems, with a move to advanced numerical meteorological models and an ability for assimilating observed meteorological data (see Section 3.5).
- Adoption of improved numerical solution techniques for the transport components of the governing equations.
- Introduction of "nesting" in models, permitting solution of equations at coarse to fine spatial resolutions as needed (Chang et al., 1997; Morris et al., 1991; Odman and Russell, 1991; Byun et al., 1995).
- Further development of "plume-in-grid" formulations, mitigating the error of instantaneously mixing the contents of a plume into a subset of grid elements (e.g., Myer et al., 1996; Kumar and Russell 1998).
- Updates to chemical mechanisms (see Section 3.2), including the introduction of recently estimated chemical-reaction rate parameters and more explicit representation of some reactions (e.g., Carter, 1996; Paulson and Seinfeld, 1992).
- Improvements in the treatment of biogenic emissions, notably isoprene, and the incorporation of the species arising from these emissions in the chemical mechanisms that are included in the AQMS (see Section 3.4).
- Initial development of algorithms for simulating aerosol dynamics (e.g., Lu et al., 1997; Bowman et al., 1997; Lurmann et al., 1997), preparing for the eventual modeling assessments of multi-pollutant issues (see Section 3.8).[b]
- Model simulations at larger, more encompassing regional scales, thereby reducing the impact of boundary conditions and permitting examination of longer-range transport issues.
- Improved approaches to sensitivity analysis whereby, in a single simulation, the sensitivity of model outputs to errors in a multitude of model inputs can be determined (Yang et al., 1998; Hwang and Byun, 1995).
- Introduction of methods for estimating the error distributions of model outputs as functions of errors in model inputs and formulation, and for estimating the relative culpability of the independent errors on the dependent (output) error (Reynolds et al., 1996; Hanna et al., 1998).

Other noteworthy advances primarily facilitate model application and diagnosis but do not improve the scientific foundation. These include:

- Developing comprehensive AQMS that combine data access and management, emission, meteorological, and air-quality models, pre- and post-processors, statistical and graphical analysis packages, and emission management decision support systems into one modularized platform, manipulated through a sophisticated graphical user interface.

[b] However, these treatments are fairly rudimentary and will benefit from continued development.

- Creating efficient "user-friendly" emission models for estimating emission rates and their spatial and temporal distribution from basic emission inventory and geographical data (see Section 3.3).
- Increasing computational speed and efficiency, in some instances through the use of parallel processing (see, for example, Dabdub and Seinfeld, 1994; Elbern, 1997).
- Developing and applying diagnostic tools that expose the inner workings of the models (Wang, 1997; Jeffries and Tonnesen, 1994; Jang et al., 1995) (see Textbox 4.3).

4.2.3 Issue: What are remaining limitations in AQMS?

Finding: AQMS, despite two decades of improvement, still contain significant limitations in their formulation and application.

Discussion: While both atmospheric and anthropogenic emission processes have important stochastic components, AQMS are basically deterministic approximations, representing to the extent possible average or mean realizations of the conditions of interest. To the extent that more extreme realizations are of interest and are significantly different in their consequences than mean realizations, the models will be unable to represent them. This is an inherent limitation in formulation.

Whatever grid size is selected for use, processes occurring at sub-grid scale cannot be represented dynamically. They can only be parameterized or treated as spatially averaged quantities. As a practical matter, dispersive processes resulting from turbulence at the fine scale, chemical reaction processes, and most emission processes will be represented by approximations: they always have important sub-grid scale components. How significant a limitation this poses is still not well understood.

As discussed, significant uncertainties often accompany both anthropogenic and biogenic emissions estimates. Estimation of biogenic emissions will remain challenging; reducing uncertainties significantly will require a large research investment over a long time.

Adequate knowledge of chemical reaction rates, intermediate formation, and product splits is available only for a minority (perhaps 20-25%) of chemical reactants. During the last decade, reduced support for research and confrontation with challenging experimental problems have resulted in a reduced rate of progress in elucidating needed rate and product information.

Only very limited capability of simulating cloud processes and their influences – on vertical mixing, attenuation of insolation, and reaction pathways and rates – resides in models today. Here, too, considerable investment and a long period of commitment to research are required to make substantive progress in expanding knowledge.

Considerable uncertainty exists in the understanding of interactions between gas and particle phases and between these phases and surfaces. Reasonably representative parameterizations are available for the deposition of some species-surface combinations, notably O_3. For others, such as NO_2, knowledge is very limited and rate estimates are highly uncertain.

The ability to estimate uncertainties itself is seriously limited and highly uncertain. Thus, the reliability of models is not well known or understood.

4.2.4 Issue: How are air-quality models evaluated, and what are the problems associated with these evaluations?

Finding: Model evaluations seek to identify and quantify errors and associated uncertainties in model output. Two of the more common approaches involve:

- Operational evaluation, which compares simulated and observed concentrations
- Uncertainty analysis, which assesses model uncertainty through an internal propagation of error analysis.

4.3 A Recent Finding Using Integrated Process Rate Analysis In An AQMS Application

Integrated Process Rate Analysis or IPRA is one of a number of new-generation on-line diagnostic tools for air-quality models that allows the contributions of individual processes at work within the model to be isolated from other competing processes and quantified (Wang, 1997). A modeling error of unexpectedly large magnitude was recently identified by applying IPRA to an AQMS simulation of a 1990 O_3 episode in central California using the model MM5 to provide meteorological inputs. The error results from an adjustment that forces mass conservation after each advection time step. It is particularly noticeable when vertical advection is large in grid cells where the terrain has substantial elevation changes. Under these conditions, this "mass adjustment" can be the primary source of trace materials, such as O_3, in the grid cell. The full ramifications have yet to be explored, but the fact that a non-physical or non-chemical process is responsible for most of the trace constituents in some grid cells is clearly worrisome. The problem appears to result largely from not using exactly the same coordinate system and/or gridding system in the air-quality and meteorological models. For example, air-quality models often have coarser vertical grid structure than the meteorological models providing their input wind vectors. Also, whereas the meteorological model may estimate wind vectors placed at the corners of the grid cells, the air-quality models may internally use wind vectors at the edges of the grid cells. This means that wind vectors must be interpolated from their coordinates in the meteorological model to those in the air-quality model; it is during this step that mass is not conserved, triggering the mass adjustment. Another aspect of the problem is the fact that non-hydrostatic meteorological models do not locally conserve mass (see Textbox 3.5).

The problem identified by IPRA is probably not unique to the MM5-AQMS pair of models but may occur in any pairing of prognostic meteorological and air-quality models when winds are interpolated from one to the other. Causes, consequences, and solutions to the problem are actively being investigated. However, initial evidence suggests that the problem would be diminished if the models were more tightly coupled, wherein they both shared exactly the same coordinate system, placed the wind vectors at identical points in space, and supplied the wind vectors from the meteorological model to the air-quality model at every time step (see Section 3.5).

Both approaches have advantages, but also limitations. Model errors obtained from operational evaluations can overestimate the actual errors, while uncertainty analyses can underestimate these errors.

Discussion: Air-quality models calculate pollutant concentrations by simulating complex nonlinear processes occurring in the atmosphere. In principle, two types of model error can be generated from this exercise. The first type is associated with a model's ability to simulate a past event. The other is associated with the predictive ability to simulate the atmosphere under conditions other than those observed in the past, such as during some future event having significantly different O_3 precursor emissions.

The second type of error is highly relevant for many policy-making applications (e.g., attainment demonstrations), and evaluating a model for this forecasting ability represents a particularly difficult and stringent performance test. Methods and data for this type of evaluation have been elusive, and consequently most model evaluations have concentrated on assessing model errors of the first type; i.e., the ability to simulate past events. This type of evaluation is thought to be necessary, but not totally sufficient, to gain confidence for model application in a predictive mode.

While model evaluation is often considered as a procedure for testing a single model application, it is often more of an iterative process of testing, diagnosing the causes of problems or errors, correcting errors or justifiably adjusting inputs, and re-testing. If the process is applied carefully and objectively, it can yield useful insights into the causes of inadequate model performance, data needs, and potentially attractive corrective actions. If, on the other hand, it is carried out for the sole purpose of optimizing model performance, it can potentially produce compensating errors that hide major model deficiencies and affect the model's response to changing environmental conditions and precursor emissions.

Models are evaluated by:

- Comparing simulated concentrations with observed concentrations during specific air-pollution episodes (i.e., an operational evaluation).
- Comparing simulated concentrations from two different air-quality models for the same episode, domain, and input data (i.e., a comparative evaluation).
- Examining the computer code for errors, estimating the error in the numerical solution techniques, and assessing the extent to which process representations reflect the latest scientific understanding, (i.e., a code audit and scientific evaluation).
- Evaluating the soundness of individual process descriptions and the reasonableness of the results that they produce (i.e., a diagnostic evaluation).
- Calculating the sensitivity of model results to changes in inputs and process descriptions (i.e., a sensitivity analysis).
- Carrying out a comprehensive propagation-of-error analysis of model uncertainty arising from estimated uncertainties in model inputs and process descriptions (i.e., an uncertainty analysis).

The discussion below describes two above-noted approaches, which appear to have received the greatest attention in regulatory applications: *operational evaluation* and *uncertainty analysis*; as well as a third emerging approach that has stronger evaluative powers: the *diagnostic evaluation*.

The Operational Evaluation is the most common type of test performed within both the scientific and the regulatory communities. For regulatory applications involving O_3 pollution, such evaluations tend to focus on ground-level O_3 concentrations, although other relevant measured species may be included. The guidelines developed by the U.S. EPA for evaluating models in regulatory applications involving O_3 mitigation provide one example of such an approach. Within these guidelines, the three primary criteria for using a given model application in the planning process are: 1) the mean normalized bias between measured and observed O_3 in the model domain is less than 15%; 2) the mean normalized gross error is less than 35%; and 3) the accuracy of the peak hourly O_3 concentration in the domain (unpaired in space and time) is within ±20%. The U.S. EPA also suggests an examination of other statistical measures, such as qualitative graphical comparisons of predicted and observed ozone concentrations in time and space. The basis for selecting these threshold values appears to be historical (i.e., they reflect the range of model performance that air-quality-models have typically achieved for O_3).

A number of problems arise when considering the information gathered from a typical operational evaluation. First, acceptance of a model through an operational evaluation derives from the lack of its rejection through a series of tests; i.e., it does not demonstrate the validity of the model application (Oreskes et al., 1994). For this reason, operational evaluations should be as comprehensive and rigorous as possible. The greater the diversity in meteorological conditions represented and the greater the number of model-simulated chemical species subjected to testing, the greater the level of confidence that can be developed. Some operational evaluations compare simulated and observed concentrations of organic compounds, CO, NO and NO_2, as well as O_3. In addition, simulated and observed concentrations aloft (i.e., from airborne platforms and tower measurements) may be compared. Observational data are themselves subject to error and can never be sufficiently complete to comprehensively characterize the state of the atmosphere. Thus, estimates of model error based on an operational evaluation are themselves approximations with an attendant uncertainty that is uncharacterized.

Second, a major source of uncertainty in operational evaluations arises from the incommensurability between model-calculated and observed parameters (see Textbox 4.4). In a grid-based or Eulerian model,

4.4 Dealing with Incommensurability

The problem of incommensurability can be addressed, in principle, either by minimizing or characterizing mismatches in scale. One way to minimize the degree of mismatch would be to reduce a model's grid cell dimensions to more closely approximate point measurements. This approach, however, would further stress the model by demanding that it simulate the positions of maxima and gradients very precisely. It would also impose a daunting additional computational burden. Finally, the approach would require considerable additional effort to create the more finely gridded meteorological and emission fields.

Another way to minimize the degree of mismatch would be to use remote-sensing techniques that make path-integrated measurements on scales similar to the grid cell dimensions. In this way, approximate grid-cell average concentrations could be measured.

A third approach might involve operationally estimating the actual variability of concentrations within a given grid cell. For example, one might estimate grid cell variability by placing many sensors in several grid cells selected to be representative. The estimated within-cell variability then provides an error bound on single measurements made in grid cells having conditions similar to those in which the high-resolution measurements were made. A model estimate might then be compared with a measurement made within its grid volume, bounded by the experimentally determined spatial variability for that type of cell.

All of these approaches, however, are experimental, and cannot yet be considered for operational AQMS applications at this time.

simulated concentrations represent grid-cell averages. Typically, a surface-layer grid cell ranges from about 4 km to 40 km in horizontal extent and 50 to 200 m in height. Measurements, on the other hand, are made only at one or more specific points within a grid cell. If substantial concentration gradients exist, the average concentration within a grid cell is usually not equal to that at the specific sampling point(s), leading to a mismatch or incommensurability between model-calculated and observed concentrations. Exacerbating this problem is the fact that some VOC and NO_x monitors have been sited to obtain maximum source-oriented concentrations and hence are clearly not representative of grid-cell averages.

A third error source stems from the occasional inability of meteorological models to predict accurately the locations of clouds and precipitation--phenomena that can have profound effects on the distributions of O_3 and its precursors. Thus, even if an air-quality model correctly represents the chemistry and small-scale physics, evaluation metrics based on point-to-grid-cell and spatial pattern comparisons may indicate poor performance.

For these reasons, operational evaluations can overestimate the actual model errors in pollutant concentrations at the sites where the measurements were made. The operational evaluation does not assess model skill in regions of the modeling domain where measurements were not made, nor does it assess in principle the ability of the model to accurately predict the O_3 response to a precursor-emission change.

The Uncertainty Analysis facilitates an internal examination of model uncertainty by propagating estimated errors in model inputs and parameters through the model computations to calculate corresponding uncertainties (i.e., error bounds and error distributions) in model outputs. With the growing power of computational facilities, this type of analysis is becoming increasingly tractable; in fact, some AQMS have incorporated this capability into their on-line codes (**CR19**). While the algorithms applied for uncertainty analyses can be quite robust, the model uncertainties generated from such analyses are nevertheless themselves subject to error and uncertainty. In the first place, an uncertainty analysis requires input of error probability-distributions in the inputs and parameters. These distributions can be poorly defined and often are estimated subjectively. Moreover, model uncertainty analyses of this type are limited inherently to examination of those processes and parameters that are specifically included in the model, and thus cannot account for potentially important excluded processes such as, for example, omission of a key chemical mechanism. For this reason, model uncertainties derived from this type of analysis may represent lower limits to the actual model error.

The Diagnostic Evaluation focuses on testing whether the physical and chemical processes within the model are acting correctly. This type of testing often requires an intensive field-study database, and considerably more analysis than simply comparing model inputs and outpurs with field data. Such a limited analysis allows only an inferential evaluation since it does not specifically identify the parts of the model contributing to any given deviation of the simulated values from the observations. On the other hand, more comprehensive diagnostic analyses, such as the process-analysis methodology featured in Textbox 4.3, have the potential to identify which model algorithms are responsible for the disagreement. Thus, diagnostic tests help to assess whether model (O_3) response to precursor changes is technically correct. In other words, these tests can tell if the O_3 concentrations are estimated correctly for the "right" reasons. This latter type of evaluation probes the model's ability to simulate individual processes within the model and thus can uncover the existence of compensating errors. Examples of diagnostic tests include investigating ratios of chemical species that are sensitive to specific processes within the model (i.e., O_3/NO_y, O_3/NO_z, and NO_z/NO_y) and comparing concentration changes from weekdays to weekends. However, diagnostic tests are subject to the same incommensurability issues as are the operational-evaluation tests, and they do not provide specific numerical estimates of model uncertainty.

4.2.5 Issue: Has the skill in the AQMS improved during last decade?

Finding: In terms of simple measures, the skill of operational air-quality models to simulate surface O_3 concentrations does not seem to have improved substantially, despite increases in model sophistication and complexity **(CR19)**.

Discussion: Table 4.1 presents results of O_3 episode simulations in the Los Angeles area from 1989 to 1997 and illustrates model performance for computing surface O_3 concentrations over the past decade (**CR20**). While the more recent simulations incorporate numerous improvements in model inputs and process simulation, the simple statistical measures of model accuracy do not suggest any significant improvement in the models' ability to simulate O_3. Operational evaluations of model applications in other geographical areas generally yield smaller normalized errors, but no clear trend toward lower simulation errors with model development and improvement.

A sobering lesson of this finding is that while the development of more advanced models may provide the scientific community with a more comprehensive tool for investigating atmospheric physical and chemical processes, it does not guarantee better model performance from an operational point-of-view. One explanation for the lack of improvement is that, although some sources of error in AQMS have been reduced through improvements in model formulations and inputs, substantial approximations, process omissions, and inadvertent errors remain and these have been and continue to be the primary contributors to model error.

There are, however, more optimistic explanations. The first of these relates to the practice of model "tuning." As noted earlier, it has been, and remains today, a standard practice to adjust model inputs and parameters within their individual uncertainty ranges to improve correspondence between observed and computed patterns. In the past, the opportunities for model tuning were somewhat greater than they are today. In the first place, past operational evaluations typically made use of fewer observations in both space and time than those carried out today. As a result, it was probably easier to use tuning to produce favorable comparisons between simulated and observed concentrations than today. Moreover, model parameters had more uncertainty and thus it was acceptable to adjust these parameters over wider ranges. The data necessary to characterize initial conditions, boundary conditions and meteorological fields used to drive the models were not as well characterized in both space and time as they are today, and hence, they also could be adjusted over wider ranges of values. It is probable that the relatively large amount of model tuning in the past provided an unrealistically positive assessment of model performance relative to present-day models. This may have, in turn, obscured the impact of recent model improvements on performance. On the other hand, it should be noted that tuning of present-day models may still mask significant model deficiencies. Including precursor concentration comparisons as well as O_3 comparisons in operational evaluations could place a more stringent chemical constraint on the models, and thus limit the tendency for model tuning to yield erroneously positive model-performance statistics.

A second explanation for the apparent lack of model-performance improvement relates to the limitation of the operational evaluation itself. As noted in Section 4.2.4, incommensurability between model output and observed data can result in significant

Table 4.1 O_3 performance statistics for simulations of Los Angeles episodes ordered by the year in which the modeling study was reported.

Model	Episode Day	Normalized Bias (%)	Normalized Gross Error (%)	Unpaired Accuracy of Peak (%)	Reference
EPA Guideline		< ±15	< 35	< ±20	
UAM	6/7/85	-12	33	-15	SCAQMD, 1989
UAM	6/7/85	-3	32	-7	SCAQMD, 1991
UAM	6/7/85	-24	30	-38	SCAQMD, 1994
UAM	6/25/87	-7	33	11	SCAQMD, 1991
UAM	6/25/87	-43	45	-19	SCAQMD, 1994
UAM	6/25/87	-40	42	-12	SCAQMD, 1996
CIT	6/25/87	-17	31	-24	McNair et al., 1996
UAM	8/28/87	-4	28	-2	SCAQMD, 1991
UAM	8/28/87	-36	52	-29	Wagner et al., 1992
CIT	8/28/87	-23	38	-54	Harley et al., 1993
UAM	8/28/87	-34	37	-34	SCAQMD, 1994
UAM	8/28/87	-26	29	-23	SCAQMD, 1996
CIT	8/28/87	34	50	-4	Harley et al., 1997

UAM = Urban Airshed Model, version IV; CIT = Caltech/Carnegie Institute of Technology model discrepancies between model results and observations that are not caused by model deficiencies. For example, a study conducted for Los Angeles, where the monitoring network is relatively dense, found that incommensurability alone can produce a normalized gross error for O_3 of 25 29% (McNair et al., 1996). It is conceivable, given the typical O_3 performance statistics indicated in Table 4.1, that O_3 model performance as measured by an operational evaluation has attained a level where further model improvements are being obscured by the discrepancies introduced by incommensurability and measurement error.

4.2.6 Issue: What is the reliability of the AQMS simulations used in policy formulation?

Finding: Definitive, quantitative knowledge regarding model uncertainty is an important prerequisite for reliability characterization. As noted above, however, uncertainties in specific AQMS applications are generally poorly defined.

Discussion: The reliability of air-quality models depends on the fidelity of their scientific formulations and their required input data. As outlined above, the

4.5 Kern County California 1985: A Cautionary Tale

Modeling results can be quite sensitive to the inputs provided and, in turn, to decisions made to construct those inputs based on available data, however rich or sparse. The experiences associated with modeling efforts conducted in 1985 in preparation for hearings to adopt NO_x emission control requirements for Kern County, California, provide an illustration.

UAM-IV modeling was conducted by two groups - state agency staff and a consulting company retained by the private sector. Both groups used the same model and the same database. Each was able to discuss matters with the other if they wished. One group determined that NO_x emissions were limiting in the area of highest O_3 concentrations, the other that VOCs were limiting, with NO_x reductions having adverse effects on O_3 in some subareas. Each recommended that the pollutant found to be limiting in its analysis should be controlled. How could such a difference arise when so much of the two analyses shared the same information and approaches?

Subsequent investigation has uncovered two primary reasons. First, a sub-area in the eastern side of the county was lacking surface meteorological data. It was believed that flow in this sub-area formed a portion of an eddy of considerable size. In order to “give guidance” to the wind field interpolation program, one group inserted a “phantom” meteorological station with prescribed “data.” The other group did not; they instead allowed the interpolation program to determine the shape of the flow in the area, including those portions in complex terrain. The net result was that each modeling effort produced somewhat different directions for wind flow passing through the general area where data were lacking. Since the emission fields were highly variable spatially, this led to mixtures of differing proportions of precursors being transported to the area of high O_3 concentration.

Second, one VOC surface measurement was available, made in the morning, in a direction generally upwind. To establish initial conditions, one group extrapolated the measurement aloft, in essence setting the concentrations of VOC aloft at levels as high as at the surface. The second group assumed that, since VOC is emitted at the surface and vertical mixing had yet to occur, VOC concentrations taper off sharply with elevation. Thus, the first group specified a high VOC-to-NO_x ratio aloft, the second, a much lower ratio. Moreover, the group specifying the high VOC-to-NO_x ratio modeled only a relatively short time period; thus, initial conditions displayed an exaggerated impact on O_3 concentrations. The net result was that from two slightly different and reasonable sets of inputs, two very different mixtures of precursors in the relevant downwind area were produced. This, in turn, generated control strategies that were different in kind and not just in degree.

4.6 New York Metropolitan Area 1988: A Second Cautionary Tale

Meteorology, through its indirect effect on atmospheric chemistry, can influence whether O_3 in a particular urban area is more responsive to VOC or NO_x controls. Usually there are insufficient observations of meteorological parameters at the surface and aloft to characterize required meteorological fields for air-quality modeling. Various diagnostic or predictive numerical models are applied in conjunction with the available data to produce the required meteorological fields. Each of these models carries with it a set of assumptions appropriate to the type of model, but which may potentially influence the atmospheric chemistry through meteorological interactions, producing different results for O_3 and O_3 responsiveness to emission controls.

An example illustrating this issue is described by Sistla et al. (1996) in which the UAM-IV model was used to simulate an O_3 episode in the New York Metropolitan area over the period July 5-8, 1988. Four different techniques for characterizing wind fields (3 diagnostic; 1 predictive) and three different methods for determining mixing heights (2 spatially varying; 1 spatially invariant) were used. All of these techniques were consistent with the existing guidance for applying the UAM-IV in regulatory modeling applications in the presence of a database consisting of only routinely measured parameters.

Model simulations included a base case as well as several emission control scenarios consisting of across-the-board reductions in anthropogenic emissions of: (a) 75% NO_x, 25% VOC; (b) 25% NO_x, 75% VOC, (c) 25% NO_x, 25% VOC. Results indicate that the various methods used for wind field development showed varying levels of skill in simulating the O_3 plume emanating from the New York Metropolitan area in the correct locations. Base-case plumes differing in location because of wind field differences produced differences in maximum ozone concentrations of as much as 20 ppb.

The model results also indicated that peak O_3 levels estimated under the NOx-focused emission control scenario were lower than those estimated under the VOC-focused scenario, using spatially-invariant mixing heights. On the other hand, peak O_3 levels predicted for the VOC-focused scenario were generally lower under spatially varying mixing heights. One of the lessons coming from this experience is that modeling uncertainties deriving from a sparse meteorological database may produce ambiguities in the emission control preference for reducing peak O_3 concentrations. In other words, two modeling groups, each following allowable procedures to model the same application with the same modeling system, may obtain different directional O_3 response to emission controls. A more dense meteorological network, especially for vertical profile data, and/or stricter guidance on developing meteorological fields in data-sparse regions may have avoided some of the ambiguity in this example. Also, the more recent practice of using advanced numerical predictive meteorological models helps to better characterize vertical profiles in data-sparse regions, although other parameters in the modeling system, with many degrees of freedom, may also lead to an ambiguous response scenario.

ability to test their fidelity has intrinsic pitfalls, and in an operational sense is often limited by uncertain and incomplete observational data. Moreover, experience suggests that model results can be significantly affected by what might appear to be relatively small changes in input data that had been inadequately constrained by supporting observations (see Textboxes 4.5 and 4.6)[c]. In instances where one or more categories of data needed for model inputs is sparse or lacking, the risk of generating compensating errors and model bias (see Textbox 4.7) is significant, and where these types of errors are present, reliability is questionable.

For these reasons, an assessment of the uncertainties in model output ideally should be a component of the information used in a regulatory context. Unfortunately, the uncertainty in model predictions depends on a variety of factors, including the type of prediction, the application, the uncertainties in model

[c] These examples are believed to be relatively rare, but they do illustrate the importance of critically examining modeling results when these results drive policy decisions.

input and parameters, and the model itself. Moreover as discussed, the methods currently available for assessing model uncertainty are themselves uncertain. Consequently, it is not possible to make universal statements about model uncertainty; e.g., air-quality model predictions are valid within ± X ppb or ± Y% . While such statements cannot be made, a number of comments on uncertainty are possible:

- The uncertainty in the absolute O_3 concentrations estimated by an AQMS is likely to be quite significant. In the first place, uncertainty analyses for each of the AQMS components — emissions, chemistry, transport, vertical exchange, plume dispersion, and deposition — typically indicate uncertainties in the range of 15-30% (see for example, Table 4.2). Moreover, statistics derived from uncertainty analysis may understate errors because the models themselves are necessarily incomplete and may have internal errors.
- The magnitude of the uncertainty in O_3 changes from emission changes is highly dependent upon a variety of conditions. For example, Table 4.2 illustrates the O_3 response to changes in VOC and NO_x for three values of the VOC/NO_x ratio. When VOC/NO_x = 8, a ratio that places the chemistry in the transitional region between VOC- and NO_x-limitation (see Section 3.2), the change in O_3 is small and equal to or less than the corresponding standard deviation. Under these conditions, there can be little confidence in the model predictions, or even the direction of the predicted O_3 change. For calculations removed from the transition region, on the other hand, the model-predicted O_3 changes appear to be quite robust. The implication is that, for areas having a clear and significant excess of one precursor, directional errors are quite unlikely. Conversely, when and where transitional conditions exist, identification of the relative efficacy of controlling one precursor over the other using an AQMS may be difficult and the results uncertain.
- There are no operational methodologies that comprehensively assess uncertainties associated with O_3 concentration predictions or the response of O_3 concentrations to emission changes Thus, the accuracy of AQMS estimates is not fully characterized.
- The difficulty of estimating emissions for

Table 4.2 Mean predicted O_3 concentrations, O_3 changes and their standard deviations for box-model simulations considering uncertainty in the chemical mechanism.[a]

VOC/NO_x (ppmC/ppm)	6		8		12	
Simulation	Peak Ozone (ppb)	Standard Deviation (ppb)	Peak Ozone (ppb)	Standard Deviation (ppb)	Peak Ozone (ppb)	Standard Deviation (ppb)
Base	155	57	198	48	188	41
25% VOC reduction	138	51	192	46	184	40
25% NOx reduction	180	54	194	45	180	39
ΔO_3(25% VOC)[b]	-17	9.3	-6.6	5.6	-4	2.4
ΔO_3(25% NOx)[c]	25	12	-4	10	-8	4.6

[a] (**CR19**)
[b] Case with 25% VOC reduction minus base case
[c] Case with 25% NO_x reduction minus base case

4.7 Model Error and Model Uncertainty: What Do We Mean When We Use These Terms?

The term “model error” is often used to quantitatively express the degree to which models, based on an imperfect mathematical characterization of the physical and chemical process under study, fail to accurately reflect reality. If, for example, we employ an air-pollution model to estimate the O_3 concentration at some particular time and location, we can expect the model-calculated concentration, c, to differ somewhat from the actual value, C. Usually we have no way of determining in advance what the numerical value of this difference, or “error’” will be. Instead, we can estimate the distribution of probable model error, based on a distribution of possible model outcomes expressed by a probability-density function (see Figure 4.2). As indicated here, the distribution of possible concentrations need not be symmetric about any axis; that is, it can be “skewed.” Moreover the mean value of the probability-density curve does not, in general, coincide with the “true” value, C. The difference between the true value and the mean model-calculated value reflects the bias in the model’s predictive ability. In practice, the probability-density function is difficult to determine, although one can attempt its estimation using error-propagation analysis. As a consequence, random errors (represented by the spread of the probability curve), and systematic errors (or “bias,” represented by the offset) are hard to segregate from one another. These two error classes behave in fundamentally different ways, however, and it is important to keep their distinction in mind when addressing model uncertainty. In view of these complexities, model uncertainty is often expressed in terms of some simple and approximate statistic, such as estimated standard deviation of the distribution about its mean. In view of the likelihood that skewness and bias nevertheless exist, however, such elementary statistics should be treated with due caution whenever encountered.

In addition to the model error discussed above (often referred to as “absolute error”), we can also define an error in the difference between two model predictions with slightly different input parameters (often referred to as “relative error”). For example, an air-quality manager might evaluate the impact of a proposed change in local precursor emissions by running the model with two sets of emission inventories: a base case that yields some mean O_3 concentration c and an emission control scenario that yields some new concentration c’. If we let C represent the “true” concentration for the base case and C’ for the emission control case, then $\delta = c' - c$ is the model-estimated value for the change in the O_3 concentration from the imposition of emission control and $\Delta = C' - C$ is the actual change. In performing this exercise, we would expect that the probability density function for c’ to be quite similar to that of c. Moreover, because many of the model’s deficiencies should be common to both simulations, it is reasonable to expect that the subtraction process will remove a significant component of the error in both calculations. The net result is that the error distribution in the calculation of δ should be significantly narrower than that of either c or c’ (see Figure 4.3).

In addition to concern about the magnitude in the absolute and relative errors in model predictions, it can be even more critical to know that the sign in the model-calculated value for δ is correct. This is often referred to as being “directionally correct” in that it relates to whether any given action will lead to a decrease or increase in O_3). One would certainly hope this to be true in a regulatory decision-making environment: in such cases imposed changes will at least move air quality in the model-predicted direction, as opposed to producing counterproductive results. The example depicted in Figure 4.3 suggests this to be generally the case, but there is often a small possibility - shown by the shaded area in the figure - that a “directionally incorrect” model prediction will be produced.

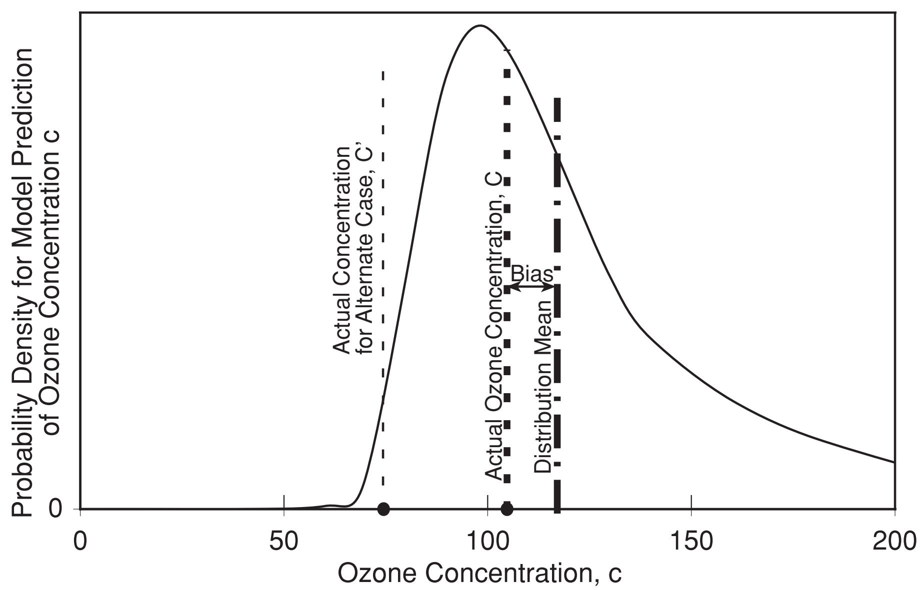

Figure 4.2 Hypothetical probability-density function for model-predicted surface O_3 concentrations, c. The mean model-predicted O_3 concentration is indicated by dashed-dotted line, while the "true value" in the O_3 concentration, C, is indicated by dashed line. The probability-density function for *model error* is simply the difference between the density function for c depicted here and the value C. The model bias is represented by difference between C and the mean value of c.

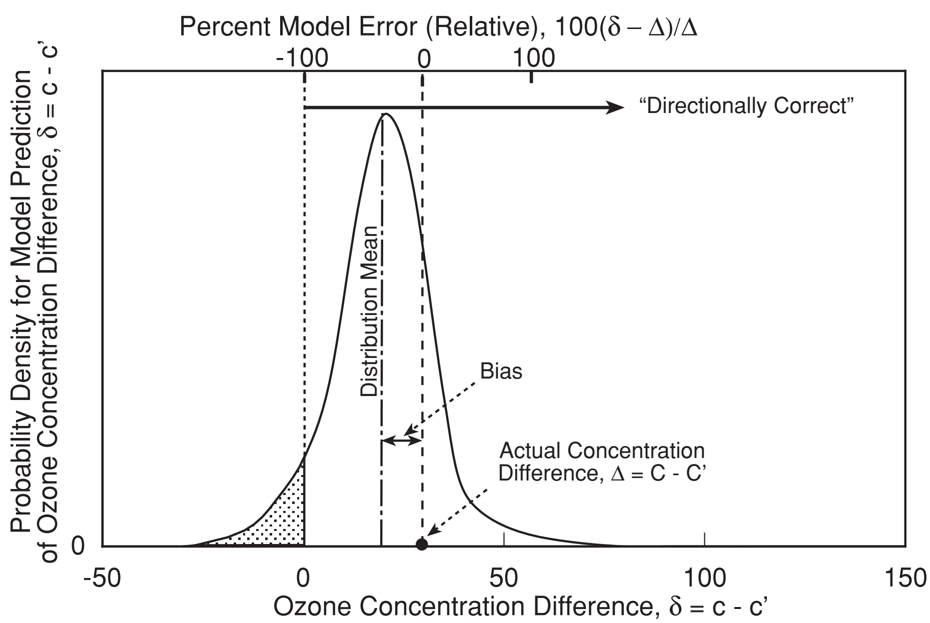

Figure 4.3 Hypothetical probability-density function in the change in model-calculated O_3 concentration from an imposed emission change. The difference between the heavy-dashed line and the mean value for the dashed-dotted line represents the relative bias in the model prediction. The shaded area indicates the region of directionally incorrect model predictions.

4.8 The Pitfalls of Model Simulations with Compensating Errors

Any operational evaluation that is based solely on a simple comparison of observed and model-calculated O_3 concentrations is unlikely to detect existing, compensating model errors. This situation is particularly problematic in the policy context, because models having such errors may accurately reproduce a particular observed O_3 field, but for the wrong reasons. This in turn can instill a misplaced confidence in a model's predictive capability, possibly leading to inappropriate policy decisions based on its use.

The schematic shown in Figure 4.4 gives a simple example of such a situation. Here point A_i represents the model's predicted O_3 concentration on an EKMA diagram using the actual (or true) NO_x and VOC emission inventories. Point B_i on the other hand, represents the model's result obtained with a NO_x inventory that is biased low, combined with a high-biased VOC inventory. By pure coincidence these emission inventories compensate one another, to produce the same model-predicted O_3 concentration for simulation B as for A.

Now consider what happens if these simulations are used to estimate the impact of a hypothetical reduction in NO_x emissions. In the case of the simulation using the correct inventories, the solution moves to the new point on the EKMA diagram indicated by ΔA. As can be seen from the figure, this results in a slight *up-gradient* displacement, resulting in a predicted *increase* in O_3. Conversely, the same NO_x-emission adjustment to the simulation originally using the incorrect inventories moves the solution a distance ΔB on the diagram. This is *down-gradient*, indicating an O_3 *decrease* and giving a *directionally incorrect* result.

Thus we see that the presence of compensating errors in AQMS simulations can pose a major pitfall for policy makers and could, in principle, lead to the adoption of directionally incorrect control strategies. For these reasons, AQMS applications in the policy-making arena should be carefully screened for compensating errors and related biases. This can be accomplished through rigorous operational evaluations involving precursor species as well as O_3, diagnostic testing of modules, independent modeling by two or more groups, and comparisons of AQMS results to those obtained using observation-based models.

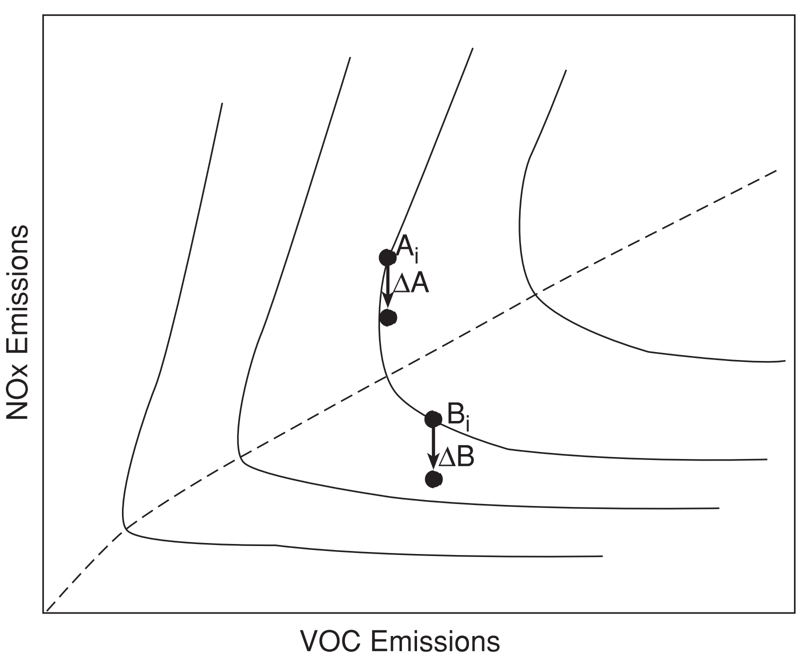

Figure 4.4 Hypothetical Schematic of O_3-Concentration Contours Resulting from Various Levels of NO_x and VOC Emissions

future periods and scenarios contributes to the uncertainties.

- If two model simulations differ only in the emission inputs, the errors in the predicted O_3 concentrations can be highly correlated between the two simulations. In such a situation, the uncertainty in the O_3 difference between the two simulations, ΔO_3, is less than the uncertainty in the predicted O_3 for either individual simulation (**CR19**). See Textbox 4.7 and Table 4.2. This suggests that models may better be used to predict ΔO_3 or other quantities where systematic biases can cancel, rather than to predict absolute O_3 concentrations. However, the uncertainty in ΔO_3 can still be larger than ΔO_3 itself, if ΔO_3 is small or the O_3 predictions are in the transition region between VOC- and NO_x-limitation.

Consideration of all factors reveals that model uncertainty can be significant and poorly characterized. Relying solely on the output of an AQMS to resolve control issues or to demonstrate attainment of an O_3 standard or objective is problematic. Regulators have recognized this limitation in the application of AQMS; as a consequence, they use additional information derived from air-quality and meteorological data and emission patterns.

Despite the limitations imposed by unknown uncertainties, air-quality models remain the only tools available for quantitatively simulating or estimating outcomes for other than historical or present conditions. Steps to alleviate some of the problems discussed above include:

- Performing diagnostic evaluations aimed at disclosing model formulation errors, using, for example, integrated process rate analyses, and correcting errors uncovered.
- Collecting quality-assured measurement data for key modeled inputs and for evaluating model performance in simulating a wide variety of diagnostic variables.
- Determining the spatial variability of modeled variables within areas equivalent to grid cell volumes.
- Comparing simulated and observed spatial

4.9 Proprietary Models

A proprietary model is one that is not available for review and testing by all members of the scientific, regulatory, and regulated communities. Access to the model is restricted by sequestering the source computer code and releasing only an executable version, by requiring a licensing agreement that prevents any modification of the code, and/or by requiring a substantial or excessive fee to acquire or use the model.

From the perspective of the scientific community, the major problem with proprietary models is that they cannot be evaluated as thoroughly and independently as models that are unrestricted in documentation and use. If the source code is unavailable, it is impossible to review the detailed physical, chemical, and numerical approximations made in the model or to check for coding errors. If modifications to the model are not allowed, it is impossible to conduct a rigorous diagnostic evaluation to understand why the model predicts as it does. It also is impossible to evaluate the limitations of the model by reducing the number or severity of the approximations made in its derivation. If a substantial fee is charged to acquire the model, not all members of the scientific community will be able to obtain it, and there will be fewer evaluations of the model and comparisons with other models.

The scientific community is inherently limited in the guidance that it can provide for applying a proprietary model in the regulatory process-and-interpreting the results. In particular, the confidence level associated with model predictions for future years or O_3 changes associated with emission changes is unknown. This uncertainty can undermine the credibility of the regulatory process among stakeholders.

patterns, using unbiased interpolation methods that also provide error estimates, using observed values representative of the spatial interpolation scales.

- Developing and using methods for quantifying modeling uncertainty, identifying its causes, and quantifying the magnitude of individual contributions.
- Simultaneously mitigating the more important causes of model uncertainty through investigations leading to a better understanding of atmospheric processes and their accurate representation in air-quality models, while developing or adopting methods for making air-quality management decisions, given uncertainty.

4.10 The Potential Benefits of a Community Modeling System

Comprehensive modeling systems have engineering attributes that position them to serve the function of true community air-quality models, analogous to the role that MM5 plays in the meteorological field and the Community Climate Model plays in the climate modeling field. Modular in design, a community modeling system can be readily updated with new modeling technology, provided that it adheres to simple coding conventions that ensure compatibility with the host framework. In general, several modules would be available to simulate any process in the official version of the model. Thus, one could choose from a menu of advection, diffusion, chemical mechanism and other types of schemes in configuring the model for a particular application. It would also have an integrated process rate analysis capability (see above) and a mechanism for conducting uncertainty analysis through repeated runs with modified inputs.

The administrative structure of a community model system might be composed of

- Dedicated staff that maintain the official version of the model, ensure that new modeling technology is properly evaluated and accepted before being adopted as a component of an official version, incorporate new modeling technology into the official version, and provide documentation to the user community relevant to the model's use.
- An executive committee and its advisory council to provide policy direction, establish rules for adoption of new modeling technology, and provide funding for the infrastructure.
- Representatives from a broad spectrum of stakeholders.

The user community might be segmented into two general types: developmental and operational. The developmental users would make changes and additions to the algorithms and architecture that would then be candidates for incorporation into the official version of the model. The operational users would use model configurations derived from the official version "as is." The opportunity to suggest modifications to the system, based on the experiences of both user communities might be facilitated through formal annual meetings where users would present their ideas for modifications and additions to their peers and to the managers of the system. A mechanism for assessing and accepting proposed modifications into the official CAQMS version would be developed.

Benefits of a Community Air-Quality Modeling System (CAQMS):

- Provides an open modular model framework accessible to a broad user community.
- Facilitates routine model assessment and peer review at the developmental and operational level.
- Encourages periodic updates of regulatory models with state-of-the-art technology.
- Facilitates integrated assessments across multiple air-quality issues and assures self-consistency.
- Affords greater flexibility in the construction and comparative evaluation of process modules and modeling configurations.
- Provides periodic peer review of new technology at the evaluation and acceptance stage, enhancing model credibility in the community at large.

The establishment of model reliability, as highlighted in the last three bullets above, requires open modeling systems and indicates that the use of proprietary models (see Textbox 4.9) be discouraged. The development and use of a Community Air-Quality Modeling System (see Textbox 4.10) is likely to be an effective means for establishing reliability through shared experiences of the scientific and user communities.