Protocols for Evaluating Oxidant Mechanisms Used in Urban and Regional Models

H. E. Jeffries, M. W. Gery, W. P. L. Carter

Department of Environmental Sciences
and Engineering
School of Public Health
University of North Carolina
Chapel Hill, NC. 27514

Atmospheric Research Associates
4th Floor
729 Boylston Street
Boston, MA 02116

Statewide Air Pollution Research Center
University of California
Riverside CA 92521

Cooperative Agreement 815779

Project Officer
Marcia Dodge

Chemical Processes and Characterization Division
Atmospheric Research and Exposure Assessment Laboratory
Research Triangle Park, NC. 27711

ATMOSPHERIC RESEARCH AND EXPOSURE ASSESSMENT LABORATORY
OFFICE OF RESEARCH AND DEVELOPMENT
U. S. ENVIRONMENTAL PROTECTION AGENCY
RESEARCH TRIANGLE PARK, NC. 27711

DISCLAIMER

This report has been reviewed by the U. S. Environmental Protection Agency, and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the U. S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

Abstract

In this project, a task force of three chamber operators and three modelers was assembled to address needs raised at a prior workshop on the procedures and practices that should be followed when evaluating photochemical reaction mechanisms for their suitability for use in EPA's urban and regional air quality models. In addition to the members of the task force, two workshops were held that were attended by 18–20 researchers in the field including scientists from England and Australia. At these meetings, issues raised by the task force were discussed and commentary on the approach was provided.

In this report, based on our work and the community input, we describe how to create a protocol for the evaluation of photochemical reaction mechanisms. Rather than prescribe a set of specific actions, we present instead criteria that influence decisions without specifying what those decisions must be. These specify what the evaluator must consider, what is and is not relevant, and what must be reported as the basis of decisions made.

Based on general scientific principles, we describe five characteristics that reaction mechanisms must have if they are to be acceptable. Mechanism evaluation is more than just establishing agreement of the model with observations. Our approach is based on a complex argument in the form of a cascaded inference chain showing how to proceed to establish that a candidate mechanism might exhibit all five characteristics. The evidence in this argument is mostly chamber data. The argument rests on the credibility of the chamber data and thus, the data's credibility must be communicated by the evaluator by reporting how he was convinced the data are credible.

In five chapters the report details the elements that must be considered and describes the general content of reports the evaluator must produce.

While we now have a well reasoned and logically supported procedure for evaluating a mechanism, making the case that a particular mechanism is accurate is presently a somewhat difficult undertaking mainly because of the data incompleteness problem (discussed in Chapter 6). Although a large number of experiments are available that were produced over more than 15 years by three different chamber groups, the number of experiments that are capable of providing a compelling evaluation of the various mechanism components is quite limited. Thus, at present, it is not possible to make a compelling case for accepting a mechanism for ambient air use, only to vindicate its use as having been evaluated the best that can be currently done, given the available data. Clearly, additional chamber data are needed for the "standard chamber data base."

Contents

Figures

Acknowledgments

We would like to thank the many people who helped us on this project and especially the participants of the two workshops which we held to discuss the purposes and contents of protocols for mechanisms evaluations.

Participants in the June 1990 workshop were:

- Jeffrey Arnold, University of North Carolina
- Roger Atkinson, University of California, Riverside
- William Carter, University of California, Riverside
- Richard Derwent, AEA Technology, Oxfordshire
- Marcia Dodge, Environmental Protection Agency
- Thomas Fister, University of North Carolina
- Michael Gery, Atmospheric Research Assoc.
- Harvey Jeffries, University of North Carolina
- Nelson Kelly, General Motors Research Labs
- James Killus, Atmospheric Research Assoc.
- Bill Long, University of California, Riverside
- Frederick Lurmann, Sonoma Technology Inc.
- Steve McDow, University of North Carolina
- Kah Eng Pua, University of North Carolina
- Kenneth Sexton, University of North Carolina
- Romualdas Simonaitis, Tennessee Valley Authority
- William Stockwell, State University of New York

Participants in the December 1991 workshop were:

- Jeffrey Arnold, University of North Carolina
- Joseph Bufalini, Environmental Protection Agency
- William Carter, University of California, Riverside
- Tai Y. Chang, Ford Motor Company
- Basil Dimitriades, Environmental Protection Agency
- Marcia Dodge, Environmental Protection Agency
- Thomas Fister, University of North Carolina
- Michael Gery, Atmospheric Research Assoc.
- Harvey Jeffries, University of North Carolina
- Graham Johnson, CSIRO, Australia
- Nelson Kelly, General Motors Research Labs
- James Killus, Atmospheric Research Assoc.
- Steve McDow, University of North Carolina
- Kenneth Schere, Environmental Protection Agency
- Kenneth Sexton, University of North Carolina
- Joseph Sickles, Environmental Protection Agency
- Gary Whitten, System Applications Inc.

We especially thank Jeffrey Arnold for his contributions to the concepts and ideas expressed herein; he introduced us to a whole field of inquiry that most chamber chemists have never thought about. His active participation in the formulation of the approach and his critical analysis of the various drafts is greatly appreciated. We are grateful to Kenneth Sexton and Sharon Pickard for their organization and hosting of the two workshops. Kenneth Sexton was also instrumental in understanding important differences between the UNC and the UCR chamber operations; he also provided important measurements in the UNC and UCR chambers. James Killus was an active member of the task force. He even moved temporally to UCR to gain access to the original UCR data. He made significant contributions to our understanding of the importance of subtle factors in the data. Steve McDow provided valuable insight into how surface reactions might be occurring on Teflon film in chambers. A number of students also made contributions to this project including Tom Fister, Wen Li, and Adom Smith.

Foreword

Why does EPA need protocols for testing oxidant mechanisms?

In his introduction to the Proceedings of the 1986 Workshop on Evaluation/Documentation of Chemical Mechanisms (EPA-600/9-87/024), Basil Dimitriades said

> The underlying problem and the need for standard procedures for documenting and evaluating chemical mechanisms were first recognized and discussed during a 1983 workshop on the Empirical Kinetics Modeling Approach. At that time, EPA was disturbed by the fact that models of insufficiently documented validity were offered as official Agency guidelines for development of costly control strategies. Furthermore, the plethora of ozone mechanisms in existence, their differences in terms of ozone predictions, and the difficulties in documenting even the relative validities of the various models caused another problem: they encouraged the inappropriate practice of State government officials responsible for development of State Implementation Plans to select for application those mechanistic models that happen to support pre-conceived control targets. Therefore, for EPA to be able to defend its control policies and regulations, and for the States to be able to develop objective and effective SIP's, it is imperative that the model-type guidelines issued by EPA be of well documented utility and validity.

The purpose of the 1986 workshop was to examine critically approaches for evaluating the accuracy of gas-phase chemical reaction mechanisms for use in urban and regional oxidant models and to determine if a "standard data base" for conducting evaluations could be developed. A consensus was reached at this workshop that "there is, and has been, a generally accepted procedure for testing the extent of 'reasonable agreement' between model predictions and experimental measurements" and "this procedure involves the use of laboratory kinetic, mechanistic, and product data, the use of smog chamber data, and the use of other test data." Further, there was consensus that "although there certainly were problems with their data, smog chambers still provided the most unambiguous data for the testing of urban chemical transformation mechanisms." Recommendations included (a) the creation of task forces to resolve a number of measurement issues and to assemble and review the best chamber data for evaluating mechanisms, and (b) the creation of a review group to determine how best to conduct mechanism tests.

Subsequently, the University of North Carolina (Jeffries, Sexton, Arnold), Atmospheric Research Associates (Gery, Killus), and the University of California, Riverside (Carter, Long) were funded by EPA as a task force to carry out part of the 1986 recommendations. Included in the design of the project were review meetings with other scientists familiar with chamber data and mechanism evaluation.

In the first year of this effort we were influenced by the work of Robin Dennis at EPA, who was producing a protocol for evaluating the RADM model, and by Philip Roth, who was a participant in both the RADM protocol and the Urban Airshed Model guidance group (Jeffries was also a member of the UAM group). Several key ideas from these programs were influential on our thinking; these were: (i) that protocol development is a process, not a product, (ii) that acceptance should come slowly after a long series of stressful tests with no failures, and (iii) that given the complexity of the issues under study, scientists are worried about "substituting an answer for a judgment."

We slowly recognized that the best we could do would be to produce guidance on how an individual researcher might formulate a protocol; we would not be producing a cookbook for evaluating mechanisms. That is, we wished to establish criteria that influence decisions without specifying what those decisions must be. Such criteria do specify: (a) what each evaluator must consider in reaching his conclusion that a mechanism is satisfactory for EPA use, (b) what he may and may not consider relevant, and (c) what he can legitimately

be required to report as the basis of his choices. In this report, we present our best ideas about these criteria and thus we set the stage for an individual evaluator to create his own protocol accounting for the specific goals of his particular project, yet meeting the criteria given here. No prior evaluation even comes close to meeting these criteria.

Other reports about data quality and the documentation of chamber experiments that constitute the best available chamber data base from multiple laboratories are forthcoming.

The three authors are not in agreement about some details for carrying out a particular evaluation, but we do agree on the approach recommended here. We look forward to the next round of mechanism evaluations to see how these ideas will work in the practical world of limited budgets and firm deadlines.

1

Protocols

A protocol is a signed document containing a record of the points on which agreement has been reached by negotiating parties preliminary to a final treaty or compact (Webster's Twentieth Unabridged Dictionary). *The document you are reading is not a protocol.* It does describe *how to create a protocol* for a given photochemical reaction mechanism evaluation task, where the negotiating parties are the sponsor (assumed to be the EPA, but the guidance is not limited by this assumption) and an evaluator who will implement the proposed actions. The evaluator may be the original mechanism developer or another person. The guidance given here describes what must be reported for mechanisms and data used in evaluations, what must be included in the evaluation process, and what must be reported as the basis of decisions for choices made in carrying out the evaluation.

Scope of an Evaluation Protocol

We assume that the purpose of a mechanism evaluation study sponsored by the EPA is to determine whether a particular chemical reaction mechanism satisfactorily describes the transformations of specific chemical species that occur on urban and regional scales. These results may then be used by EPA officials—along with other information not directly related to the quality of the chemistry—to judge that, at this time, it is reasonable to adopt this mechanism for use in EPA air quality models. Or, the study may conclude the opposite, in which case the mechanism would most likely not be used in EPA air quality models. In either case, we must have a persuasive argument that requires a clear chain of reasoning and credible evidence.

Acceptable Mechanisms

Mechanism evaluation is more than just establishing the agreement of the model with observations. The notion of acceptance is very complex and perhaps ambiguous; there is no single criterion of merit. In part this is because knowledge is incomplete and somewhat uncertain and therefore one must accept a particular explanation as the best one *now* available.

As with other scientific theories, acceptable chemical reaction mechanisms have several characteristics* (this list is not exhaustive, but the items are most important and certainly show the type of elements that enter into such judgments):

i) *consistency* – a mechanism's formulation must be consistent, not only within itself, but with currently accepted theories and facts applicable to these theories; the most obvious example of this characteristic is that acceptable mechanisms must incorporate consensus kinetics values such as rate constants and reaction products;

ii) *accuracy* – the mechanism's consequences or predictions should be in demonstrated agreement with the results of existing experiments and observations; such agreement must not arise from compensating errors among a mechanism's inner workings;

iii) *simplicity* – the mechanism should not be burdened with a large amount of mathematical apparatus required to explain the basic features (i.e., be simple and elegant in symbolic expression), yet should contain all the terms that are needed in explanation, so one does not have to substitute a large number of external facts; it should have the "feel" of a good solution—there should be a certain sense of congruity or recognition. Thus, we prefer explanations of events in terms of necessary causal forces to explanations of events in terms of other empirical generalizations.

iv) *scope* – the mechanism's predictions should extend beyond the particular observations it was initially designed to explain; for example, a mechanism's predictions should show agreement with chamber data that were not used in its development and, in another form, a mechanism initially intended for the prediction of ozone, should also be capable of predicting other associated intermediate or final products such as formaldehyde or hydrogen peroxide;

v) *fertility* – the mechanism is expected to predict novel phenomena, that is, phenomena that were not parts of the set to be explained. The display of such fertility reduces the likelihood that this mechanism is just an *ad hoc* one— that is, invented just for the original occasion but of limited use. Or, where an anomaly is encountered, a good model should itself serve to suggest possible modifications or extensions.

Individually these characteristics as criteria are imprecise: individuals may legitimately differ about their application to concrete cases. In addition, when used together, they conflict with each other, i.e., a simple mechanism may be limited in scope. Nevertheless, these characteristics must be used to warrant the arguments that mechanism developers and evaluators must make in building a case that their mechanism should be used by EPA in ambient air modeling.

All existing chemical reaction mechanisms—some now known to be very wrongly formulated—have "fit the data," but only more or less. It is just this incompleteness and

* These are adapted from Thomas Kuhn's "Objectivity, Value Judgment, and Theory Choice," in The Essential Tension, Univ. of Chicago Press, 1977.

imperfection of the theory-data fit that has defined many of the research topics in this field in the last 20-years. These past results suggest that accuracy of mechanisms alone cannot always discriminate the goodness of mechanisms.

Thus, objective criteria may be insufficient to determine in full any algorithm for mechanism choice. If we cannot set out an algorithm to which all will agree then how should we proceed? The answer is we seek criteria that influence decisions without specifying what those decisions must be. Such criteria can specify a great deal:

- what each evaluator must consider in reaching a decision that a reaction mechanism can be used by EPA in ambient air simulations,
- what each evaluator may and may not consider relevant in the process, and
- what each evaluator can legitimately be required to report as the basis for the choices they have made.

These are the principles used in formulating our protocol guidance.

What Must Be Considered In Accepting A Mechanism

If mechanism evaluation is more than just establishing agreement of the model with observations, then how does one go about building a case that a mechanism should be used by EPA? Here we answer this question by constructing a cascaded inference chain based upon the criteria given above. Such a chain is built from a series of single-step *arguments*, each seeking to establish a single point or claim, and arranged in multiple paths to bridge between the physical evidence and the final conclusion. Formal arguments have necessary parts and a definite structure (see Figure 1). Data become evidence when they are relevant to the claim being asserted. A datum is relevant to the extent it causes you to change your mind about the likeliness of the claim. Arguments occur in two stages: (1) the foundation of the argument, which establishes the existence and validity of the data (evidence) in the argument, and (2) the argument itself, i.e., the parts in Figure 1 and their relationship. The former steps are concerned with credibility and the latter with issues of relevancy. In an inference chain, the claim or conclusion of one argument becomes the evidence (data) for the next higher argument (see Figure 2). Some steps may require two or three pieces of evidence be established to support the higher claim. Each of these pieces of evidence would have associated issues of credibility, and thus determining the credibility of each higher claim is a complex judgmental task for the evaluator. It is just this judgment that has to be explained in the reports.

Arguments also require *warrants*. A warrant registers explicitly the legitimacy of the steps involved in the argument and refers to a larger class of steps whose legitimacy is being presupposed. That is, a warrant tells us why the evidence is relevant to the conclusions,

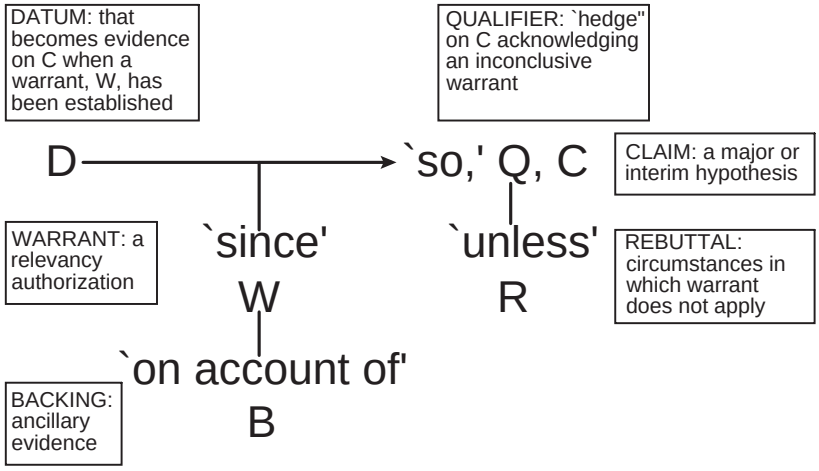

Figure 1. Parts of a formal argument.

i.e., why a person should accept the claim, given the data. Some warrants authorize us to accept a claim unequivocally and, given the appropriate data, these warrants entitle us to qualify our conclusion with the adverb 'necessarily'; others are more tentative, or else subject to conditions, exceptions, or qualifications, such as 'probably' and 'presumably.' Thus, we generally need to add some explicit reference to the degree of force that our data confer on our claim by virtue of our warrant. There are also circumstances in which the warrant does not apply (stated in the rebuttal). Finally, while warrants may be applicable in a particular case, we may have to explain why *in general* we should accept a warrant as having authority. That is, the warrant itself may rest on some other condition not explicitly given in our argument. While warrants are often hypothetical, bridge-like statements, the backing for warrants is in the same form as the data on which the claim rests. Thus, backing for warrants appear as ancillary evidence in the arguments and are therefore subject to the same credibility concerns as are the mainline data.

Many arguments that appear in evaluation studies omit the backing and rebuttal components; some arguments even omit the warrant, counting on the reader to fill it in from his own experience. We believe that this is a poor practice that often leaves the reader not really convinced that a sound case has been made. We will supply warrants for the major steps of our argument in a following section.

The Principal Argument

Figure 3 shows the structure of a cascaded inference chain containing a series of arguments that justify the claim we originally described above:†

† This claim appears at the top left in Figure 3 and because of space limitations in the figure, it is written

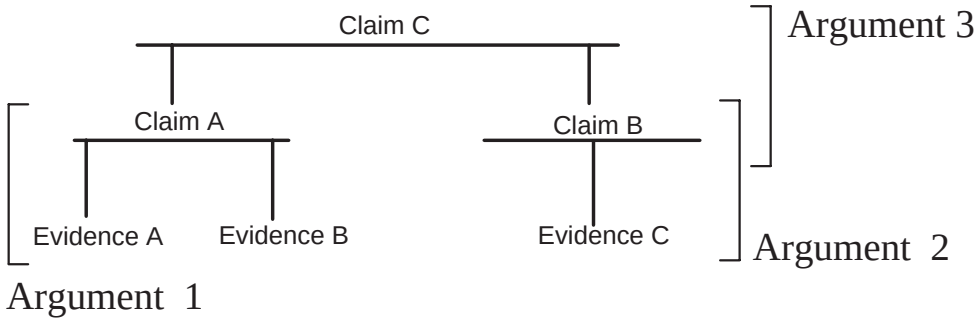

Figure 2. Parts of a cascaded inference chain. This chain has three arguments: Argument 1 is that (Evidence A and Evidence B) imply (Claim A is true); Argument 2 is that (Evidence C) implies (Claim B is true); Argument 3 is that (Claim A being true and Claim B being true) implies (Claim C is true). Warrants are not shown in these diagrams.

"It is reasonable to use a particular chemical mechanism in EPA urban and regional models."

This conclusion is not the focus of our protocol guidance because it involves factors other than the quality of the chemistry. Instead, we accept as the focus of an EPA-sponsored mechanism evaluation study the next lower necessary claim:*

"It is reasonable to act as if the principal reaction mechanism describes, with a stated degree of uncertainty, the chemical transformations of relevant species on urban and regional scales."

In the upper half of the diagram, we are concerned with whether relevant data (evidence) exist to build an argument that reaches the conclusion, and we are concerned with the sources and magnitudes of the uncertainties that arise in constructing this argument. The warrants for these arguments will be supplied below.

In Figure 3, each claim depends upon lower claims as evidence for accepting the higher claim. Thus, to establish the validity of the claim cited above requires that we make two further intermediate and necessary claims:

"It is reasonable to act as if the principal reaction mechanism is well-formulated."

which covers the first three criteria for mechanism acceptance (accuracy, consistency, and simplicity) from the set discussed above, and

"It is reasonable to act as if the principal reaction mechanism has predictive capability."

and this covers the other two criteria (scope and fertility).

in a shorter form, "Use Mechanism in EPA Model".

* This claim appears second from the top left of Figure 3 and it is written in the shorter form, "Mechanism describes relevant species on urban/regional scales".

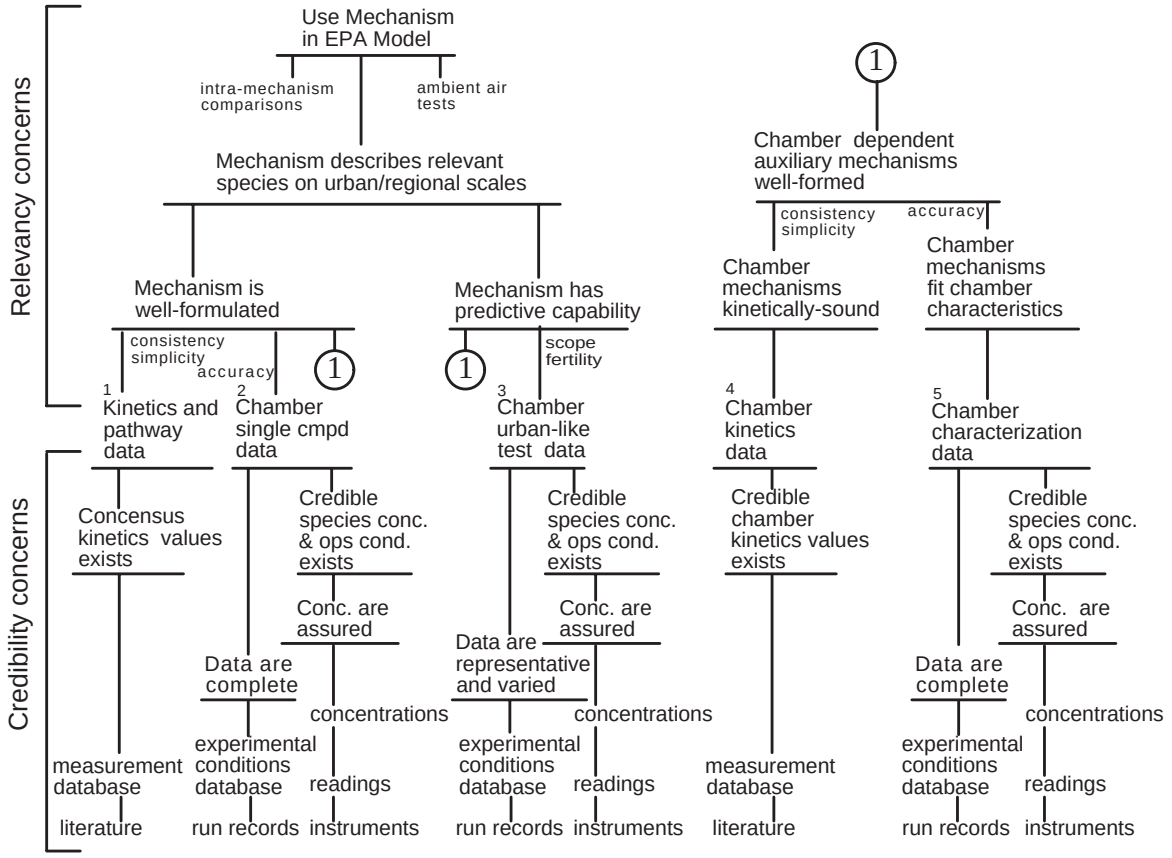

Figure 3. A cascaded inference chain of arguments for why a mechanism should be adopted by EPA. The items numbered 1 to 5 in the middle of the figure are observational data sets that make up the standard data base for mechanism evaluation.

Basically the first of these two conditions says that the mechanism *explains* atmospheric transformation chemistry as best we know it, and the second condition says that the mechanism also *predicts* atmospheric chemical events with acceptable accuracy without compensating errors. The warrant that makes these two conditions relevant to the higher claim is that prediction techniques which are *not* based on fundamental, cause-and-effect, time-dependent physical laws may *not* be reliable in all circumstances. As Stephen Toulmin said, "we have a forecasting technique which not only works, but works for explicable reasons."†

† Stephen Toulmin, *Foresight and Understanding, An Enquiry into the Aims of Science* Harper and Row, 1961

The Primary Evidence

As we cascade downward through the argument chain in Figure 3 we come to observational data (consensus laboratory findings and chamber data) which are used as evidence for the lowest claims in the argument. These are shown in a row in the middle of the figure, and they are the components of the "standard data base" needed to test photochemical reaction mechanisms. The data base consists of five different kinds of data:*

1. kinetics and pathway data that support the formulation of the principal reaction mechanism;
2. chamber data produced from experiments with single VOCs which test explicit representations in the principal reaction mechanism;
3. chamber data produced from experiments with complex, urban-like VOC-mixtures which test the generalization in the principal reaction mechanism;
4. kinetics and reaction data that support the formulation of the chamber-dependent reaction mechanisms for each chamber used in tests;
5. chamber data produced from experiments that reveal unique attributes of each chamber, or chamber characterization experiments, which test the qualities of the chamber-dependent reaction mechanism.

Each of these will be described more fully below, but first we must explain why such data are relevant to the argument.

Warrant for Use of Observational Data As Evidence

In Figure 3, observational data are used as evidence to establish several necessary intermediate claims. What makes kinetics or smog chamber data legitimate for this purpose? The answers to these questions supply the warrants for our argument.

Warrant For Kinetics Data Being Evidence

Beginning in 1978 with the NASA publications on photochemical kinetics data, and later in 1984 with Atkinson and Lloyd's review of kinetic and mechanistic data for selected reactions, there has been a growing set of consensus kinetics data directly relevant to the formulation of photochemical reaction mechanisms. These reviews, conducted by a group of research scientists on a regular basis, have established the accepted body of kinetics knowledge (essentially the "truth" about many individual atmospheric reactions as we now know it). These are type 1 data in the list above and in Figure 3. Because these data represent

* Several new terms are introduced here. They will be formally defined in the next chapter. If you are having difficulties understanding this section, perhaps you should skip to the next chapter and review the formal definitions before proceeding here.

the current scientific understanding of the atmospheric chemical transformation world, and because there is only one world, mechanism developers must create photochemical reaction mechanisms that are entirely consistent with this body of knowledge. Evaluators must assure that this is so.

A rebuttal for this warrant is when the developer presents newer kinetics data than were in the reviews at the time of formulation. These newer data, however, may have less force than if they appeared in one of the consensus reviews, because they might not subsequently be repeatable by others or they might not be totally reliable in the opinions of the independent kinetics reviewers. Nevertheless, any kinetics data for a process cannot be ignored by a developer and must either be included in the formulation or be explicitly accounted for if rejected.

Kinetics data, therefore, have immense force in determining mechanism acceptance. Such data are *direct* evidence about processes in the mechanism. For the most part, however, this force only works in one direction. That is, a mechanism that is at odds with kinetics data is almost certain to be judged as ill-formed, while one that is totally consistent with *all known* kinetics data may still be thought of as ill-formed because the parts of the mechanism not directly derived from kinetics data may be incorrect and in particular the mechanism may suffer from unintentional deletion whereby important but unknown reactions that occur in the air are missing in the mechanism. Thus, agreement with data of type 1 is a necessary but not sufficient condition for accepting a mechanism.

Warrant For Smog Chamber Data As Evidence

The ideal reaction mechanism would be one in which *all* the model entities represent real molecules with well-understood properties, and in which all relevant processes among the entities would be completely specified. Such a mechanism, it can be argued, could reliably be trusted to provide accurate answers in the costly decision making cases because it is true. While there are a number of convergent sorts of causal lines that lead to the conclusion that many entities in the mechanisms do exist and do have the properties as given, right now, *our knowledge of atmospheric chemistry is sufficiently incomplete that modelers have no other choice but to postulate molecular entities, their properties, and interactions that allow the model to account for observed phenomena.* That is, modelers must propose what is otherwise a still hidden-structure to explain observations.

Smog chambers are the most useful instruments for producing observations that are likely to reveal clearly and unambiguously the effects of this hidden-structure for atmospheric chemical transformations. That is, chambers have a large signal-to-noise ratio for the creation of chemical transformation phenomena. This is because, in chambers, events can be arranged so that to a high degree of certainty only chemistry is working to change

concentrations of species; other causes of changes such as emissions and deposition, dilution and entertainment, and transport have zero or very low rates compared to the chemical rates of change. Thus, by explaining the chemical transformation phenomena in chambers, we can at least gain some understanding of an underlying structure that is only ambiguously revealed given just the naturally occurring events in the ambient urban atmosphere where several competing processes change the concentrations. Further, it is only in chambers where sufficient observational data are available to aid modelers in building part of the mechanism.

A rebuttal to this warrant, however, is that the uses of chamber data are complicated by the fact that chambers can and do introduce distortions into the transformation processes as observed in the chamber. That is, events which occur only in chambers influence the observations, and thus explicit actions are needed to avoid introducing derivational biases in the mechanism formulation. Care must be taken to see that we account for the distortions introduced and that these do not influence the properties of the postulated entities in a model of the transformations. If such separations are successful, we can then argue on strong grounds that the distortion-free part of the model can be applied to cases other than the chamber, e.g., the ambient atmosphere. If the evidence is weak on this matter of separating the chamber-dependent reactions from the principal reaction mechanism, the belief of the well-formedness of the principal mechanism is substantially weakened.

Chamber data are seemingly weaker evidence than kinetics data for rejecting mechanisms. Clearly, chamber data are always *circumstantial* evidence in the matter because, in a chamber we always observe only the macro-effects, not the individual processes that are represented in the mechanism. But even so, chamber data do provide a rather upper constraint on the formulation of reaction mechanisms. In a similar fashion, kinetics data, which are direct evidence about the individual processes, are incomplete and therefore provide only a lower constraint on mechanism formulation. Having assured himself and others that a mechanism is consistent with kinetics (thus establishing that the mechanism's formulation is within the lower bound), it is the responsibility of the evaluator to assure himself and others that the mechanism's predictions are also consistent with chamber observations for explicit species included in the mechanism (thus establishing that the formulations of various parts of the mechanism are within the upper bound also).

Thus, it is mostly the degree of success of the model's fit to the chamber observations that in part warrants the degree of its acceptance as the best one available for simulating ambient atmospheres, but only if the demonstrated agreement is also shown to be free of compensating errors, especially those arising between the chamber-dependent reactions and the principal reactions. Nevertheless, if we cannot insist that a chemical reaction mechanism be proven real, then we must demand that it be at least *instrumentally reliable*, and thus reproduces the outputs of the chambers as sophisticated instruments.

Need For Chamber-Dependent Auxiliary Mechanisms

In Figure 3, we wish to claim that a particular mechanism is well formulated and that it has predictive power. Unfortunately, the mechanism we wish to make these claims about is not the mechanism that can be tested using observed data. This is because, as explained above, each chamber has a set of perhaps unique reactions between some gas-phase species and wall surfaces (and probably additional reactions between wall-adsorbed species and other wall-adsorbed species). Mounting evidence suggests that oxides of nitrogen species are central players in these wall-mediated reactions and that a gas- phase reaction product is most likely nitrous acid (HONO) which rapidly photolyzes to produce hydroxyl radicals, the most central radical in all VOC oxidation mechanisms. Thus, the chamber walls have a great potential to influence the predictions made by the principal or chamber-independent mechanism undergoing tests with chamber data. This coupling between chamber-dependent reaction mechanisms and the chamber independent reaction mechanism can result in compensating errors, whereby a wrongly formulated wall-reaction mechanism compensates for a wrongly formulated and supposedly chamber-independent mechanism for say aromatics, giving seemingly accurate predictions for the chamber, but when used in ambient air, giving entirely wrong predictions.

The solution to this troublesome problem is to require the formulation† of separate chamber-wall-mediated reaction mechanisms that can be tested in each chamber (see the right-hand side of Figure 3). This requirement introduces the need for additional kinetics and chamber data to be included in the standard data base.

The second type of kinetics and reaction product data in Figure 3—type 4 data, are those related to reactions that occur between the gas phase and the chamber walls or on the chamber walls. This body of knowledge is much less advanced than consensus kinetics data, in part, because the reactions appear to be somewhat unique to each chamber. Nevertheless, most chamber operators have taken special measurements intended to explain (or at least account for) these transformations.

While many of the transformation processes mediated by walls have not yet been explained in a satisfactory way, sufficient observations have been made to allow the construction of chamber-wall-mediated reaction mechanisms that can simulate the observed effects reasonably well. The developer of these chamber-dependent mechanisms should present evidence as to the identities of the species that might be in common with the principal mechanism. This evidence would be used to support the completeness of the chamber tests described below.

Because the principal mechanism's formulation is supposed to be independent of the set of chamber-wall-mediated reactions used by the principal mechanism developer, the

† *Development* of chamber-wall-mediated reaction mechanisms is not included in this evaluation protocol.

evaluator does not have to use the same set of chamber-dependent reaction as the developer. The evaluator may have formulated reaction sets of his own. But wherever the evaluator obtains his reaction set, its quality is very important to the outcome of the evaluation, therefore, *the chamber-wall-mediated mechanisms must be evaluated.* This must occur before the principal mechanism can be evaluated.

The predictive accuracy of the various chamber-wall reaction mechanisms must be established by evaluating them with chamber data from experiments using the simplest chain-propagating species (e.g., carbon monoxide, formaldehyde, methane), systems with principal chemistry assumed to be so well-understood from fundamental kinetics data that the cause of any lack of fit between predictions and observations for these simple systems can be assigned to the chamber-wall reaction mechanisms. A second demonstration of the goodness of the chamber-wall reaction mechanisms would be that the principal mechanism, when combined with several different chamber-wall mechanisms, could give accurate simulations in all the different chambers. This result can be used to argue that there is no "contamination" of the principal mechanism by the different chamber-wall mechanisms. Making this argument is a critical step in supporting the claim that the principal mechanism describes relevant species concentrations on urban and regional scales.

Acceptable Chamber Data

Completeness

The term 'chamber data' will be defined in more detail later, but here it is important to recognize that chamber data includes more than just the concentration values at a series of times for selected species. For chamber data to be acceptable for use in the testing described here, they also must include other information needed to simulate a particular experiment. Thus, chamber data include:

- data which specify the *physical operating conditions* of the chamber (these are mechanism-independent values), for example,
  - ▲ data needed to convert the species absorption cross-sections included in a particular mechanism into time-dependent photolysis rates for a simulation with that mechanism, i.e., data needed to compute actinic fluxes,
  - ▲ data needed to provide time-dependent chamber air temperatures and chamber water vapor concentrations to the simulation program, and
  - ▲ data needed to provide time-dependent chamber dilution to the simulation program;
- data which specify the *initial chemical conditions* in the chamber, for example,
  - ▲ a set of observations at one time which are declared the best estimate of the "starting" concentrations of material deliberately injected into the chamber (note that these can be directly used in the simulation program input only if the mechanism contains an explicit species representation, otherwise, the "starting" concentrations of real species must be converted, by some conversion model, to starting concentrations of model species; this is obviously a mechanism-dependent process),
  - ▲ data from some chambers may include a set of observations at some time before the injection of the test species which provide the background species concentrations in the chamber filling air, e.g., methane, carbon monoxide, trace amounts of various VOCs, and sometimes even formaldehyde;
  - ▲ some data sets may contain *estimates* of some *unmeasured,* but critical, initial concentrations such as initial nitrous acid.
- data which specify the *quality* of the concentration time-streams; and
- data that are useful in measuring "fit" between observations and predictions.

Credibility and Representativeness

In the bottom half of the diagram in Figure 3, we are concerned with credibility: how sure are we that the events represented by the reported data actually occurred, how accurately are they described, and do the data reports provide a complete enough representation. These characteristics can be established for each data stream in each experiment and for the whole data set through use of recognized quality assurance principles and methods. Additionally, these quality assurance methods must be backed by supporting evidence that includes calibration histories, histories of instrument performance and operating conditions, and sufficiently precise replicate results. Necessary completeness of the data set can be established by creating data bases of experimental conditions that allow ready matching of existing experiments to testing needs.

The Basic Steps in Carrying Out An Evaluation

Reach agreement on the purpose and scope of the evaluation

Different protocols will be needed depending upon the purposes of the evaluation. Purposes influence which principal mechanism is to be used, which species are relevant, and what data bases can be used.

Select principal reaction mechanism

Each principal mechanism will require different work to document its design qualities and will require different work to set up the evaluation tests. Each principal mechanism will have different explicit species represented and this influences which detailed comparisons can be conducted. A mechanism containing a larger number of explicit species allows the possibility of a much more extensive documentation of the comparison tests, and ultimately a complete evaluation outcome.

Select chamber-wall reaction mechanisms

Principal reaction mechanisms are supposed to be independent of chamber characteristics. Developers have not always tested this attribute. As with the basic kinetics data described above, knowledge about chamber-wall-mediated reactions improves over time. Thus, the evaluator may have a choice of chamber-wall-mediated reaction models. In any case, these models must be documented and quality assured in the same manner as the principal reaction mechanism.

Determine the relevant species

The purpose of the evaluation will determine what species are relevant in the evaluation and the purpose will influence what chamber data can be used to test these species. While all high-quality chamber experiments can be used to test the mechanism for ozone, oxides of nitrogen, and loss of principal VOCs species, the testing of say the predictive accuracy for PAN or HCHO can only be done with a more limited set of experiments because reliable observations for these species are not always available. Observations for hydrogen peroxide, organic peroxides, and organic acids, for example, are essentially non-existent, so these cannot be relevant species in a protocol because we cannot establish predictive accuracy in these cases.

Identify the relevant chamber databases

Given a principal mechanism and agreement on the relevant species, the evaluator should reach agreement with the sponsor on which chamber data bases are most relevant and which

ones might receive little or only cursory treatment. The data bases are also classified into "tiers of competency" for mechanism testing and the sponsor may only want an evaluation with the most competent data or he may desire that all available data be used.

Determine that chamber-wall reaction mechanisms are well-formed

The evaluator must select chamber-wall reaction mechanisms for each chamber he will simulate. The evaluator does not have to use the same wall reaction mechanism as the developer used, but he must use one that is well-formed and gives accurate predictions. In any case, the evaluator must establish and report that the wall mechanisms he will use have these characteristics.

Report on chamber-wall-mediated mechanism formulations

The evaluator must assure himself and others that chamber-wall-mediated reaction mechanisms are as well formulated as the laboratory data allow. This would be accomplished by reviewing the original formulations, confirming that these are consistent with present knowledge. There may be competing formulations in the same chamber and the evaluator must then assemble an argument and evidence to determine which if either is better formulated.

Evaluate chamber-wall-mediated mechanism predictive power

All chamber data bases contain special characterization experiments that allow limited testing of the predictive power of chamber-wall reaction mechanisms. A fully explicit principal reaction mechanism for the appropriate chain-propagating species (e.g., CO, HCHO, CH_4) in these characterization experiments should be combined with each chamber-wall reaction mechanism and evaluated on all characterization experiments. The evaluator must reach a conclusion about the "well-formedness" and predictive power of each chamber-wall reaction mechanism, and must conclude which ones will be used in subsequent testing of the principal mechanism. Depending upon their state of development, it may be appropriate to proceed with more than one chamber-wall mechanism for each chamber; this should be considered in negotiating a particular protocol.

Determine that the principal mechanism is well-formed

It is assumed that mechanism developers receiving new EPA-funds for producing a new mechanism would be required to meet the reporting and documenting requirements set forth here, in which case, the evaluator may elect to draw heavily upon the developer's reports in establishing that the mechanism is as well-formed as we currently know how to make it. Under these circumstances, the evaluator's task is made easier by the work of the

developer, and it may be quite simple to justify proceeding to comparisons of mechanism predictions with chamber observations.

Because new kinetics data are constantly being produced, and reviews are conducted at multiple-year intervals, it is common for at least rate constants in previously formulated reaction mechanisms to become "out of date." In addition, sometimes whole new reaction processes are elucidated. Thus, older mechanisms may contain rate constants and reaction formulations that are no longer accepted as representing the atmospheric chemical transformation processes. In carrying out a mechanism evaluation for these older mechanisms, a first step is to determine if a mechanism is consistent with this consensus "truth" and if not, to decide what to do about this flaw.

To evaluate older existing mechanisms, or to evaluate non-EPA-sponsored mechanisms produced without adequate reporting, the mechanism reporting and documenting responsibility (or quality assurance) falls entirely to the evaluator, who must meet the mechanism reporting and documenting requirements as described in a later chapter. These documentation requirements are extensive, and for some of the older principal mechanisms, the efforts to document the mechanism formulation may be quite large. A particular protocol should include an agreement between the sponsor and the evaluator for how this mechanism documentation and quality assurance will be accomplished. For some mechanisms, it may be impractical to meet the assessment and reporting requirements outlined here because the formulation information is just not available and could only be reconstructed with difficulty. Because the well-formedness of the principal mechanism is such a critical factor in the final acceptability, we strongly recommend that sponsors and evaluators *not act as if* these mechanisms are formulated correctly when they know they are not. They *should not continue* with comparisons of predictions with observations for such mechanisms because even if the mechanism showed high predictive accuracy in a limited set of test cases, this alone is insufficient evidence to decide that a mechanism is acceptable for ambient air use.

A protocol should include agreements about how to proceed depending upon the outcome of the assessment of the mechanism's "well-formedness," that is, how will the evaluator and sponsor agree on which one of the paths above will be taken and what sort of evidence must be presented to support the selected path.

Determine that explicit species, single species, and simple mixtures are accurately predicted without compensating errors

The accuracy of formulation of the explicit and model species in the principal mechanism is assessed by comparing predictions for relevant species with corresponding observations in experimental cases conducted with only a single organic species or with simple mixtures of organic species. Explicitly represented species can be directly tested. Model species can

be tested in two ways: 1) by using a single species that is a member of the class represented by the model-species, and 2) by using a simple mixture of species that are all members of the model-species class.

Determine that complex urban-like mixtures are accurately predicted without compensating errors

The results of comparisons of a particular mechanism's predictions and chamber observational data for complex urban-like mixtures must now be used to argue that a particular mechanism has predictive power.

Reach conclusions about use of principal mechanism

The results of the well-formedness conclusions and the predictive accuracy conclusions must be assembled into evidence for the argument that it is reasonable to act as if the principal reaction mechanism does or does not describe chemical transformations of relevant species on urban and regional scales.

Structure For The Rest This Document

Chapter 2 discusses definitions in detail.

To be tested under protocols negotiated using the guidance presented here, reaction mechanisms must be reported and documented in a certain manner. These requirements are described in Chapter 3. The evaluator must perform quality assurance on the mechanism's formulation. The most important evidence relates to the origins of the reactions and rate constants used in mechanisms. The evaluator must reach a decision about the well-formedness of the principal mechanism and decide whether he should proceed to the comparison of the mechanism's predictions with observations for a large number of chamber experiments.

Likewise, to provide observations for comparison with model predictions, environmental irradiation chamber data must meet certain criteria and must include certain documentation. These requirements are described in Chapter 4. The most important criteria are accuracy and completeness. The most important documentation is quality assurance reports. The evaluator must believe, and must be able to convince others, that the chamber data used in comparison with the mechanism predictions are credible. The strength of the conclusions about the well-formedness and predictive ability is directly dependent upon the credibility of the chamber data. If the test data have large uncertainties, the comparisons become useless. Thus, another task for the evaluator is to organize comparison tests according to the classification of the chamber data uncertainty.

Chapter 5 describes what must be reported about the creation of the simulation inputs. It also describes what must be reported about simulation outputs. Certain statistics must be computed and reported in certain ways. These are also described in Chapter 5.

Chapter 6 discusses problems in reaching conclusions about the quality of mechanisms and describes some strategies for coping with these difficulties. It also explains what must be considered when comparing mechanism predictions and chamber observations.

An Appendix provides a summary of all the required steps and necessary reports in a protocol.

2

Names and Definitions

There are so many pieces of data and so many different models and so many kinds of operations that it is necessary to have a chapter that gives formal names and definitions to these entities so that later discussion will not be burdened with interruptions for naming and for definitions.

Figure 4 shows most of the items we need to name and define. Before we give formal definitions, however, it will be useful to discuss the major items in the figure in a general way. Terms in bold in this general discussion will be formally defined in the next several sections.

One theme running through Figure 4 is identifying which parts are *mechanism-dependent* and which parts are *chamber-dependent* (the evaluation process is independent of both). At the top left is the **principal mechanism** reaction listing with its **kinetics data** and **species photolysis data**. This mechanism is the one being evaluated and it must be found to be independent of all chambers. As described in Chapter 1, however, the principal mechanism cannot be directly tested with smog chamber data because chambers have unique **wall-mediated reactions** for species that appear in the principal mechanism. Thus, each chamber must have its own set of *reactions* which compose the **chamber-dependent auxiliary mechanism**. These auxiliary mechanisms must be added to the principal mechanism to produce **combined mechanisms**, one for each chamber. The combined mechanisms are thus dependent upon both the principal mechanism and the particular chamber. These combined mechanisms are the ones used in chamber simulations, and therefore, are the mechanisms that are tested. It is a separate step in the evaluation to conclude that when the combined mechanisms are accurate in each chamber, then the principal mechanism is accurate in ambient air.

Below the mechanism part in Figure 4 are the component parts and operations that are concerned with converting **chamber data** into **simulation data**. Chamber data are

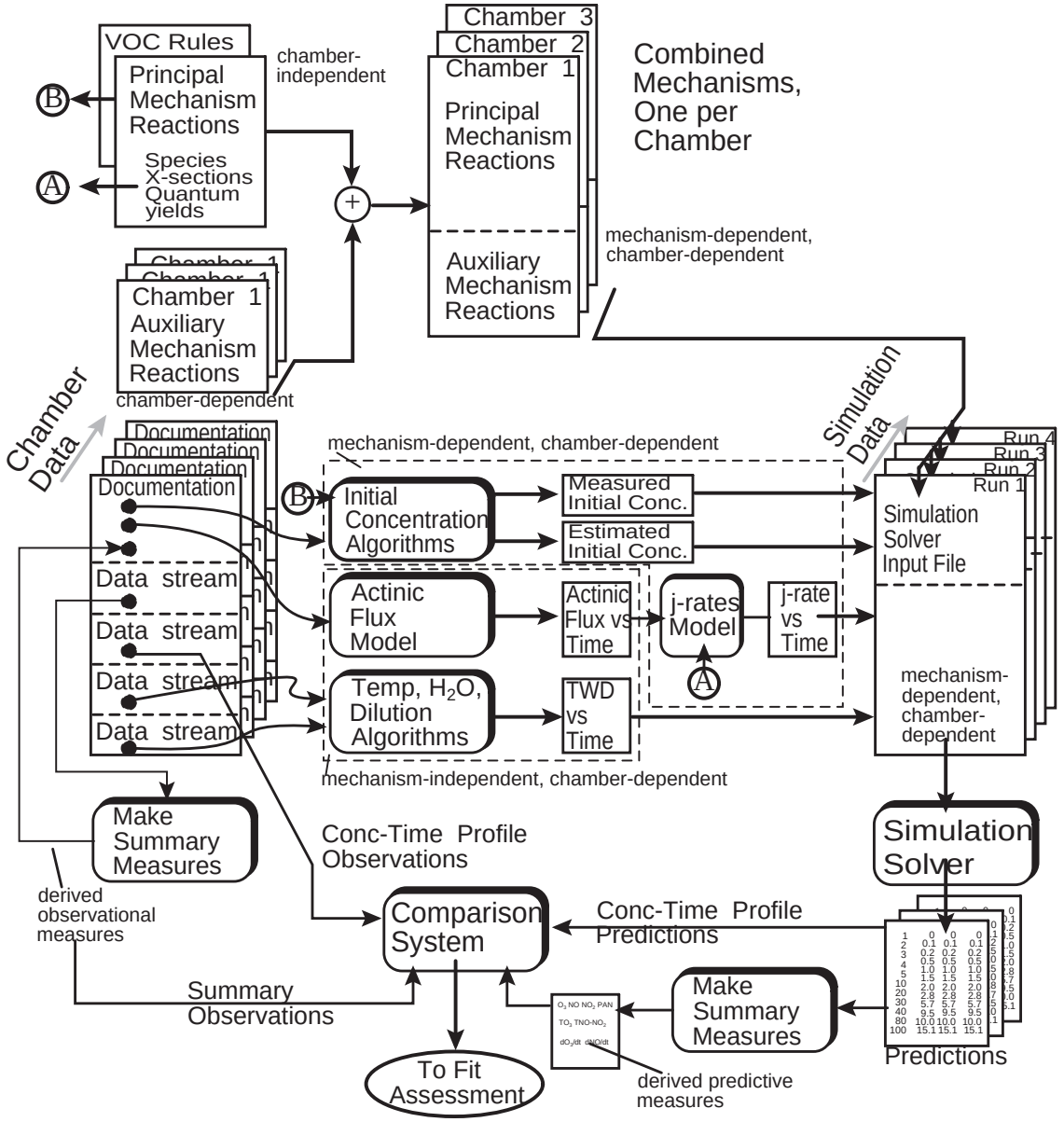

Figure 4. The relationship among mechanisms, chamber data, initial conditions models, chamber operations models, simulation data, solver system, and comparison system.

computer-readable text files containing observations of chemical and physical conditions at the beginning and during an experiment. Chamber data also include textual documentation needed as input to some of the **chamber conditions models**, including **initial concentration algorithms**, **actinic flux models**, **j-rate models**, and **TWD algorithms**. These algorithms

and models are used to convert observations in the chamber data files into data values that are components of the simulation data files. The initial concentration algorithms produce the initial simulation concentrations of **explicit** and **model species**. To convert chamber data to model species requires the VOC **Conversion Rules** which are part of the principal mechanism. Note that while most initial concentrations are measured, some have to be estimated. The latter frequently include contamination in the chamber filling air (e.g., CH_4, CO, H_2, background VOCs, HCHO, low levels of NO_X), as well as initial HONO concentrations generated from the NO_X-injection and which are almost never measured. The actinic flux and j-rate models produce the **species photolysis rates**, and the TWD algorithms produce temperature, water vapor concentration and dilution rate as a function of time in a form the simulation solver program can read. Note that the initial conditions and j-rate simulation data are both mechanism and chamber dependent, and the actinic flux and temperature, water vapor concentration, and dilution rates are mechanism independent, but are of course chamber dependent.

Finally, at the bottom of Figure 4 is the simulation solver that reads the simulation data input files and produces predictions. Two other procedures operate on the two sets of data (observations and predictions). One of these procedures, usually in the form of a computer program, produces **derived summary measures** (e.g., peak or maximum [O_3], time of maximum, average production rate) from both the observations and from the prediction time streams. The last procedure, a **comparison system**, which is usually part computer program, part human-based, performs comparisons between the observations and the predictions and produces output for the assessment of **goodness-of-fit**.

Mechanism-Related Definitions

explicit species–those chemical entities in a reaction mechanism that represent molecular or radical entities that are believed to exist in the real world, e.g., the inorganic chemical species in urban mechanisms.

model species–those chemical entities in a reaction mechanism that are an idealized, or hypothesized, or generalized representation of a molecule or radical, but which do not exist in the real world, e.g., PAR in Carbon Bond Four, or RCHO in SAPRC90 mechanism.

kinetics data–a collection of reaction rate expressions for elementary reactions as a function of temperature and pressure. Consensus kinetics data are kinetics data that are recommended for use after review by a recognized authority or group of authorities, e.g., the NASA or IUPAC reviews.

non-confirmed kinetics data—reaction rates, product yields, mechanisms, and stoichiometries based upon adjusting a mechanism under development by 'fitting' observed chamber concentration-time profiles.

species photolysis data–a set of molecular absorption cross-sections and process quantum yields as a function of radiation wavelength for each species in the mechanism that undergoes photolysis. These are often included in the consensus kinetics data reviews.

mechanism–a set of chemical reactions including rate expressions and species absorption cross-sections, which is intended to represent the complete homogeneous chemical transformations of a set of explicit and/or model species relevant to certain atmospheric processes, e.g., the ethene mechanism, or the aromatics mechanism, or an urban VOC mixture/oxides of nitrogen/air mechanism.

core reaction set—inorganic reactions plus the reactions of very simple organic species. These reactions are elementary in form and the kinetic and mechanistic parameters are usually more certain than for the larger organic compounds.

principal mechanism–a complete mechanism for an organic compound or mixtures of organic compounds combined with oxides of nitrogen and air; these are reactions that all chambers and the ambient air are thought to have in common at present; it includes (1) the reactions among the explicit and model species, and all information needed to compute their rate constants and variable coefficients or other parameters, given the temperature and pressure, (2) absorption cross-sections and quantum yields for photolyzing species so that photolysis rates can be computed from a set of chamber actinic fluxes, and (3) rules for converting observed VOC species into model species. It specifically does *not* include any heterogeneous reactions, "hardwired" pressure or temperature rate constants, or fixed-ratios of photolysis rates of one species to another.

wall-mediated reactions–heterogeneous reactions between gas-phase species and chamber surfaces which result in new gas-phase species or reactions among previously absorbed species on the walls that result in a gas-phase product.

chamber-dependent auxiliary mechanism–the *reactions and rate constants* required to represent a specific chamber's contribution to chemical transformations in the chamber, i.e., a mechanism intended: (1) to represent the effects a chamber's walls have on chemical transformation processes; (2) to include any additional gas-phase reactions of species required to represent gas-phase contaminates in the chamber's filling air that might not be included in the principal mechanism (e.g., background VOC) but which will make a contribution to the system's reactivity.

combined mechanism–the combination of a principal mechanism and a particular auxiliary mechanism; the combined mechanism is used to simulate a given chamber. These are the mechanisms that are tested using procedures recommend here.

Data Definitions

These definitions give the general contents of the two sets of computer text files used in an evaluation.

chamber data–computer text files that are mechanism-independent and providing:

1) all the information the chamber operators have available that can serve as inputs to chamber operations models used to produce simulation data input files for the experiment, for example,
    a) the chamber operator's best calculation of the initial concentrations of the injected compounds,
    b) chamber pressure during run,
    c) for outdoor experiments, values useful as inputs to radiation transfer models, e.g., total O_3 column, total H_2O column, and aerosol optical density,
    d) time of events and clock corrections.
2) information useful in displaying or graphing the measurements, e.g., titles, labels, scales;
3) warnings, cautions, and other notes about unusual conditions;
4) a data map of what is in the file and where it can be found;
5) accuracy and precision assessments for each concentration data stream;
6) summary measures derived from the concentration-time data streams and used to compare with similar summary measures derived from simulation results;
7) concentration-time numerical data streams for measured species suitable for producing concentration-time plots; these must be listed using formats that are easily read by several different computer languages.

simulation data–computer text files that are both mechanism and chamber dependent providing the simulation solver program with:

1) identity of combined mechanism to be used by the solver;
2) the initial concentrations of all relevant species present in the chamber at the beginning of the experiment;
3) the in-chamber pressure;
4) and, given as a function of time for the duration of the simulation:
    a) either the in-chamber actinic fluxes or the species photolysis rates;
    b) the in-chamber air temperature and water vapor concentration;
    c) the chamber's dilution rate;
    d) the variation of any chamber-wall mediated reaction rate or parameters.

Chamber Conditions Models Definitions

These are the component parts and operations that bridge between the chamber data and the simulation data. These are collectively referred to as "chamber conditions models." There are usually several different models: some are mathematical, consisting of equations based on theories and appropriate input observations and parameters, and some are empirical, consisting of a series of ordered observations extracted directly from the chamber data sets.

initial concentration algorithms–these algorithms compute the initial values of all relevant explicit and model species at the beginning of each experiment. There are two types of initial concentration conditions: measured and estimated. Measured initial species are reported in the chamber data files. Species that are present in the chamber filling air are often not measured, which is also true of species that form during the injection process. Because these sometimes influence the simulation predictions, methods for estimating these concentrations must be produced and these species must be included in the simulation input file. Methods are needed to convert both types of initial conditions to the form needed by the simulation program. That is, algorithms for converting from explicit organic species into the model-species representation in the mechanism must be applied to the chamber species concentration values to produce values for use in the simulation input files.

physical conditions models–the methods used to specify the time-dependent in-chamber actinic flux, temperature, water vapor concentration, and dilution rate for each experiment.

chamber actinic flux model—procedures used to produce the time-dependent values of actinic flux inside chambers. Such values are never measured in a chamber and they must be calculated from other measurements, typically spectral irradiance measurements, or particular species photolysis rates (e.g., NO_2-photolysis rate or O_3-photolysis rates).

photolysis or j-rate models—these are calculation methods for combining the species absorption cross-section and quantum yield data from the mechanism with the in-chamber actinic flux from the actinic flux model to compute the in-chamber photolysis rates of all photolytic species in the reaction mechanism.

chamber TWD algorithms—these are methods for specifying the time-varying in-chamber air temperature, water vapor concentration, and chamber dilution rate. It is common for all three to be measured and included in the chamber data. In this case, the model is usually simple—merely a tabular list of the time and measured value. Sometimes the observations are "noisy" and the observations may require smoothing before being made into a table.

Observation and Prediction Comparison

Once the simulation data file is complete it can be used as input to the simulation solver program that builds and solves the differential equations that describe the evolution of each

species concentration. These values are usually written to an output file of simulation predictions and these serve as the analog to the chamber observations. A computer program is usually used to calculate further summary measures from the predicted values. The chamber data files are the source of the concentration-time profile data and most files would contain pre-computed summary measures useful for comparison with the equivalent predicted values. Finally, some type of system is used to compare observations and predictions and to provide information useful to determine the extent of agreement between the simulation predictions and the chamber observations. These become input into a judgmental process that will be described in Chapter 5.

predictions—the explicit and model species concentrations at a sequence of time steps produced by solving the mechanism's differential equations given the initial conditions and time varying boundary conditions of the chamber.

derived summary observational measures—these are performance measures that are derived from calculations performed on the concentration-time profile streams of the chamber data sets. They include measures such as time to NO and NO_2 cross-over, magnitude of maximum [O_3], time of maximum, and rate of [O_3] production.

derived summary predictive measures—these are performance measures that are derived from the concentration-time profile streams of the simulation predictions. They are the predictive counterparts of the derived summary observational measures.

experimental series classes–ways to cluster chamber experiments to make narrower, more meaningful comparisons between observations and predictions, e.g., all ethene experiments, all propene experiments, all cooler-temperature, urban mixture experiments.

data quality classes–ways to cluster chamber experiments according to the strength of the supporting evidence on data quality and assurance activities.

relevant species–explicit and model species in the principal mechanism that (a) can be compared with observed concentrations from the data files (model species will have to be transformed to be compared with experimentally measured species), and (b) are likely to cause the evaluator to change his mind about whether the principal mechanism would make accurate predictions on urban and regional scales.

'goodness-of-fit'–an expression, or set of expressions, (ideally quantitative) of the extent of reasonable agreement between available experimental observations and their correlated simulation predictions.

3

Reporting Mechanisms

For an evaluator to test a principal mechanism using protocols developed by procedures given here, certain information about the principal mechanism must be reported before the simulations take place. These reports become part of the basis for choices the evaluator must make in judging the mechanism. In this chapter we describe this mechanism-related information and outline how the evaluator must report it as part of his evaluation documentation.

For new mechanisms formulated under EPA sponsorship, the mechanism developer would have the responsibility for creating most of the information that needs to be reported, and the evaluator's role would become one of subsequent quality assurance of the material supplied by the developer. For existing mechanisms, however, it is the responsibility of the evaluator to gather, organize, and report the needed information. For most existing reaction mechanisms, the evaluator can glean the required information from final reports on the development effort and perhaps, if needed, even by discussion with the mechanism's authors.

There is also the situation in which the evaluator *is* the mechanism developer. That the evaluator is evaluating his own mechanism does not change what he must report about the mechanism. And in fact, because the evaluator is so familiar with the development of the principal mechanism, he should be even more careful to follow the guidelines to avoid *under*-reporting the principal mechanism.

The Information Needed

Operational Information

To test a mechanism for a particular application, an evaluator must obviously know a number of *operational* items such as:

i) contents of the mechanism:
   a) the reaction list including all parameters needed for reaction rate constant evaluation (for photolysis reactions these are the absorption cross-sections and quantum yields);
   b) an identification of the species in the reactions, whether they are explicit or model species, and what they represent in the real world;

ii) instructions for using the mechanism:
   a) how to evaluate the reaction rate constants as given;
   b) how to evaluate any other parameters that appear in the mechanism;
   c) how to convert real-world species into the mechanism's explicit and model species; and

iii) methods to allow the evaluator to determine if he has correctly implemented the principal reaction mechanism in his own solver code.

The evaluator can more easily meet many of these needs if the developer has put forth a number of complete example simulations (both simple and complex) which illustrate the mechanism's use and performance. The evaluator must demonstrate that he can reproduce the developer's original results for initial, intermediate, and radical species in the mechanism, as well as for final products like ozone and PAN. Details of these documentation needs will be described below.

Formulation Information

In his first major decision, the evaluator must conclude whether the principal mechanism is "well-formed." To make such a decision the evaluator needs to know a great deal about how the principal mechanism was formulated originally.

To gain such knowledge, sufficient documentation from the mechanism's formulation must be available or must be re-created by the evaluator so that he can report as part of the evaluation the *origin* of each reaction and the *source* or *calculation method* for every parameter in each reaction. The evaluator must be able to review or re-assess these as part of the first stage of the evaluation. That is, the reporting on the construction of the mechanism must be sufficiently extensive that the mechanism can be completely re-created from (a) the cited consensus kinetics data, (b) selected chamber data, and (c) a concise description of the creation of the model species included. Furthermore, the reporting should be sufficient that any test simulations used in the formulation of the mechanism can be repeated by the evaluator so he can judge for himself the qualities of the mechanism's predictive accuracy in these "tuned cases."†

† Two points to be made about these tuned cases: (1) they may have been performed on an interme-

Finally, the evaluator should abstract from the documentation on the principal mechanism's formulation any conclusions the developer reached about the mechanism's limitations and uncertainties. The evaluator will review any available performance evaluation statistics and any discussion of limitations presented by the developer. These may lead the evaluator to perform additional chamber simulations during the comparison phase of the evaluation to confirm or to bound the limitations and uncertainties already recognized.

The following discussion of required supporting information is organized around the above subjects. First, the reporting requirements for the principal mechanism are specified. Next, the details needed to explain the basic information and reasoning behind the mechanism formulation and condensation choices are summarized. Then the reporting requirements for the auxiliary mechanisms are described. The final section discusses information needed by an evaluator to determine limitations of the mechanism.

Reporting the Principal Mechanism

The principal mechanism contains chemical reactions that all chambers and ambient air should have in common. It is the portion of the overall mechanism that should be capable of describing homogeneous, gas-phase chemistry in any air mass. The principal mechanism and its associated data structures should have sufficient utility to allow practical applications, yet should contain all the terms that are needed in explanation so that one does not have to substitute a large number of external facts. All of these technical details must be reported for the final principal mechanism to provide a concise summary of mechanism content, sufficiently complete to allow computer simulations to be performed by the mechanism's evaluator to diagnose mechanism operations and performance.

Principal Mechanism Operational Reports

The evaluator's report on a principal mechanism must contain at least the following tables and text:

1) A list of symbols used for molecular entities in the mechanism and a description of what they represent.
2) A list of reactions that occur between mechanism entities, where each reaction in the list includes:
    a) Reactants, products, and product stoichiometry.

diate form of the principal mechanism that was then transformed into the final form; this may require documenting and perhaps producing the intermediate forms of the principal mechanism; and (2) in the subsequent comparisons of mechanism predictions and chamber observations, high accuracy on the cases used to adjust parameters should not carry too much weight, but should be somewhat discounted relative to test cases never used in the adjustments.

b) In the case of thermal reactions, the specific reaction rate in molecules, cm^3, and seconds units in as complex a form as needed to express the value as a function of temperature, pressure, and if needed, other third-body interactions.

c) In the case of photolytic reactions, a set of absorption cross-sections and quantum yields in as complex a form as needed to express the values as a function of wavelength, and possibly, temperature and pressure. The values themselves can appear in separate listings from the reaction listing but are to be considered an inherent part of the mechanism. Text included in the cross-section and quantum yield listings should give the units and the form of the spectral data (e.g., points, values averaged over intervals, etc.).

3) The aggregation methodology used for specifying how all emitted and atmospheric organic compounds are represented by the entities of the mechanism. This documentation should be composed of:

- In the case of a mechanism with fixed characteristic species:
  a) A table listing the mechanistic entity representations for as many ambient and emitted compounds as can be included. This may be included in the mechanism report as an appendix.
  b) An explanation of the encoding rules used to derive the table.
  c) A discussion of limitations in the encoding rules, and notification of exceptions where a molecular entity cannot be used for a specific compound.
- In the case of a mechanism with episode-specific computed rates and parameters (i.e., kinetics or stoichiometry linked to a chemical aggregation scheme):
  a) clear and complete presentation of the necessary formulae must be given, including complete examples.

4) Technical documentation attached to the above listings, concisely explaining the following:

a) The types of general formulae used in the kinetic and photolytic rate calculations.

b) In some cases, any special methodology for encoding mechanism reaction properties that do not fit into the general structures but are needed to account for real-world phenomena.

c) In the case of variable stoichiometry, a clear and complete presentation of the formulae needed to determine product yields must be given.

d) Text or tables that provide, for each reaction and its components:

i. In the case of elementary or simple reactions, references to the sources of all kinetic and mechanistic data, and description of the type of data sources (e.g., consensus review, independent paper merged with review, superseding

review containing new data, unpublished laboratory data, estimate, assumption, and the like).

ii. In the case of complex, condensed, and lumped reactions, reference to the section of the development report that describes the formulation of the reaction, or a brief appendix that re-creates the formulation of the lumped reaction.

Other Operational Items Which Must Be Judged And Reported

Inevitably, a point was reached in the mechanism's formulation when no new kinetic data were added because the required re-evaluations would have been too time-consuming. Therefore, the evaluator should determine and report the date when the developer no longer included new kinetic information in the reaction sets.

Finally, the evaluator must demonstrate he has correctly implemented the principal mechanism by showing replicate simulations for the examples supplied by the developer. If these examples are not available from the mechanism developer, the evaluator must make provisions for an independent party to implement the mechanism separately. Then a set of example simulations must be performed by both the evaluator and the independent party. It may be that the definition of the mechanism is ambiguous and the two implementations may therefore initially differ. In this case, the evaluator and the independent party must agree about which implementation is the preferred one. This agreement and the results of the example simulations must be reported.

Principal Mechanism Formulation Reports

An evaluator can only be convinced that a mechanism is well formed if the principles, methodology, logical arguments, and supporting scientific data used to formulate different sections of the mechanism are understandable and complete. This includes comprehending the relevant, evaluated kinetic data, along with the approximated chemistry that was devised without evaluated data. Generally, a core reaction set can be formulated almost completely from consensus kinetics data. Because of missing kinetics data and the current need for condensed organic mechanisms, however, the formulation of the organic reaction schemes must advance through numerous steps.

Until the different organic reaction sets are organized into the final principal mechanism, the progressive development of the organic reaction set is usually presented by segregating it into individual sections. In each section and, at each important formulation step, the developer should have explained and justified mechanism refinements through either theoretical arguments or by adjusting a theoretically based mechanism to improve the calculation of a set of smog chamber observations. The evaluator must understand the

developer's rationale for mechanism modification and the reasons for selecting particular smog chamber experiments and measures of fit.

In many cases, especially for more complex organic compounds, smog chamber experiments for particular species might not be available or might be too poor to offer real tests of performance. The evaluator must determine what necessary assumptions and estimates were made at each step, and should document the areas of associated uncertainty. The evaluator needs to convince himself and others that the mechanism employs the highest quality chemical data presently available, and is based on accepted chemical principles and observational evidence. If the mechanism to be evaluated falls short of this standard, the evaluator must document its shortcomings.

Specifically, the evaluator should:

i) determine that, to the maximum extent possible, consensus kinetics data were used in the formulation of the principal mechanism;

ii) determine if the data that were used are still consistent with the current consensus kinetics data and if not, to report the mechanism's shortcomings for later analysis as to their significance;

iii) determine that the parts of the principal mechanism that were not based on explicit consensus kinetics data were:

   a) formulated by a reasonable extension of chemical knowledge, i.e., what was the rationale applied and how were decisions made during the formulation of these parts,

   b) determine where parameter adjustments were needed to complete the formulation of a reaction set and determine which chamber experiments were used for this purpose,

   c) determine that the chemistries of all explicit species used in the mechanism were adequately tested using data from several chambers, and

   d) determine if the testing procedures for the explicit species represented in the mechanism specifically included techniques to seek out compensating errors among parts of the principal mechanism or between the principal mechanism and the chamber-dependent auxiliary reaction mechanisms used in each chamber.

The evaluator must assure that the basic construction of the principal mechanism is sound; that is, the evaluator should determine that the principal mechanism does *not* include:

i) heterogeneous reactions involving walls;

ii) chamber-related parameters or conditions that are rightly part of either the chamber-dependent mechanisms or chamber conditions models (for instance, there should be no

assumption of constant actinic flux, temperature, pressure, or water concentrations); and

iii) photolysis *rates* or *ratios of photolysis rates* that have been determined for a specific actinic flux.

Furthermore, the mechanism should be limited in empirical representations. It should not be burdened with a large amount of mathematical apparatus required to explain the basic features (i.e., it should be simple and elegant in symbolic expression). The evaluator must comment on this aspect of the mechanism's formulation.

The following section discusses the key elements related to reporting of the principal mechanism's formulation qualities. In the case of supporting smog chamber data and chamber-dependent chemistry, further details are given in sections following this list.

Items Which Must Be Judged and Reported

I. **Knowledge of the Development Protocol.**
The formulation of a complex mechanistic structure must follow some set rules, sometimes referred to as the development protocol. The evaluator needs to understand the guidelines followed by the developer because these set the boundaries of the development process and therefore greatly influenced the formulation of the principal mechanisms. The evaluator should briefly report his understanding of:

1) How was the process of making mechanism development decisions segregated into individual development steps, and how were these steps linked in a structured manner?
2) Were efforts made to minimize compensating errors in the mechanism formulation (e.g., did the developer use a stepwise or hierarchical development process, or simulate multiple chambers with different chamber-dependent mechanisms? Did the developer allow off-setting statistics such as averaging across different chambers? etc.)
3) What was the standard by which a chemical mechanism was shown to have successfully assimilated the known chemistry at a particular development step or, in situations where subjective judgment was the only development option, how did the mechanism development process move on without good chemical data at a particular development step? In the case of smog chamber simulations, this amounts to a description of when and under what conditions the model calculations and observational data were said to be in 'adequate' agreement.
4) What was the use of the different types of development data? Which data were given higher priority? For example:

a) Could consensus kinetic data be superseded by other types of data, and if so, under what conditions?

b) After they were formulated, were the reactions and kinetics of simpler, more certain chemistry held constant, with modifications to only the new chemistry?

c) If new kinetics or new smog chamber data were included after the study was in progress, did re-evaluation occur for all affected steps to confirm that the inclusion did not cause a deviation from previously documented results?

II. **Knowledge of Supporting Data.**

Along with consensus kinetics data, smog chamber observations are a major source of chemical information incorporated into a chemical kinetics mechanism. Simulation of smog chamber observations is required in the development of chemical kinetics mechanisms, i.e., for a developing mechanism, non-confirmed kinetics data must often be extracted from the chamber data through the simulation process. Unlike consensus kinetics data (which has already been compiled from experimental data that have been peer-reviewed and published in the scientific literature), smog chamber data presents a relatively unfiltered form of data that requires the developer to make significant choices regarding its use. The evaluator should determine what supporting data were used by the developer, and as far as reasonably possible, all such data should be accumulated and sources reviewed. The evaluator should:

a) assemble in one location a list of all smog chamber data that were used in the mechanism's development;

b) identify the last revision date for the development chamber data base used in the formulation;* and

c) collect the chamber-dependent auxiliary mechanisms and other chamber operations models used by the developer for each smog chamber.

Furthermore, a significant aid in judging the performance of the mechanism would be the availability of the simulation data input files used in any chamber simulations the developer might have performed. Ideally, electronic copies of input files to the developer's solver program would be available to the evaluator.

Finally, evaluators should discuss any perceived limitations and strengths of each data set. This can go as far as to describe limitations of facilities, measurements,

* Chamber data are and will continue to be revised as new information about these data become available. Some of these changes may invalidate a conclusion reached by the mechanism developer or a previous evaluator. The results of work done as part of this study, for example, show that *all* mechanisms developed with chamber data from UNC and UCR before 1992 will have been tested with older and often incorrect chamber data.

chamber conditions, and the quantity and types of experiments in each data set. Any significant limitations should be listed, along with a discussion of how these limitations might translate into application uncertainties. If possible, the developer should suggest further smog chamber tests to resolve these uncertainties.

III. **Understanding the Formulation and Condensation Rationale.**
Any chemical kinetics mechanism purporting to represent hundreds of organic compounds with a few lumped species will require a highly condensed format. To be coherent, such a mechanism must be based on specific organizing principles. In evaluating a candidate mechanism, the evaluator must establish an understanding of at least the following areas:

1) The overlying principles of chemical organization in the mechanism, including the relationship between different lumping groups and explicit species of the principal mechanism.
2) The method used to select individual lumping groups (i.e., characteristic reaction times, molecular weight, similar chemical structures, etc.).
3) The approach for handling generalized organic radicals and chain degradation kinetics for each lumped entity.
4) For each intermediate of final lumping group, or any non-chemical operator, the procedure, choices and decisions used to form a lumped reaction scheme by combining more explicit reaction schemes.
5) Any mechanism character specifically resulting from the time scale (characteristic reaction time) of the chemistry selected. For example, are the choices made in lumping organic species and kinetics more representative of urban or continental time-scales?
6) Any mechanism character specifically relating to the intended chemical application. For example, is the mechanism designed for ozone only, or is it intended also to predict other oxidation products such as peroxides and acids?

It is not necessary that the evaluator document the intricate details of the above items. Rather, it is expected that the developer's documentation would convey the rationale behind all important mechanism formulation decisions. Therefore, the evaluator should report his general level of comfort in comprehending this type of information, but he should point out any particular areas where the developer's documentation is insufficient to convey understanding.

In some particular instances, an evaluator may need to resort to detailed reporting of specific formulations and choices. This will occur most often when there is difficulty comprehending the theoretical basis of a specific formulation step, or when serious

inconsistencies or inaccuracies occur in mechanistic simulations of smog chamber data. In such cases, it will be necessary for the evaluator to explore the developer's decision process and his supporting data to determine why specific choices were made and how he supported the choice.

Depending on the problem, the evaluator may have to report details such as:

i) An accounting of all relevant consensus and non-evaluated kinetics data used at the particular development step, along with consideration of what information is missing. At such points it may also be useful to provide a listing of intermediate reactions to describe the status of the developing mechanism under discussion. This includes

   a) a listing of the relevant consensus kinetics data (as described above for the principal mechanism);

   b) a discussion of any refinements made to consensus kinetics data (e.g., to alter kinetics, eliminate reactions considered unimportant under all chamber and ambient conditions or possibly, to add some reactions).

   When possible, the evaluator should use cited chemical principles, analogous chemistry, or smog chamber simulations to discuss approximated chemistry or empirical representations. The evaluator should also estimate the range of conditions over which such approximations may be reasonable.

ii) A description of the method used to devise particular mechanism modifications based on the simulation of chamber observations. For each development step that must be analyzed, documentation should list the particular chamber experiments simulated, the goodness-of-fit measures used to compare chamber data and simulation predictions, and a tabulated summary of the developer's and evaluator's simulation results and statistics. If necessary the evaluator may have to compare the rationale for choosing his and the developer's measurement criteria, explain why certain experiments should have been chosen and other available data were excluded, or point out experiments or sets of experiments that should have been used to test specific aspects of the mechanism.† As is the case with the developer, the evaluator is obligated to report (or reference) enough of his own input data for these simulations that another person could duplicate the newly reported simulations.

iii) A summary of the logical process used to minimize (aggregate or condense) the number of species for a particular portion of the organic reaction set. In such

† These special sets include side-by-side experiments, groups of related experiments over several days, and experiments that are similar except for the changing of one component such as NO_X and VOC levels, light and temperature, or composition of the VOC mixture.

a case, it will be necessary to trace the reasoning followed and decisions made in the progression from explicit to highly condensed organic reaction schemes, indicating the derivation of all important lumped reactions, stoichiometry, and kinetics expressions from the more explicit entities. As above, the evaluator may need to provide mechanism listings and report or reference particular chamber simulations and condensation tests. In addition, the measures of goodness of fit, the standards for defining an "acceptable" condensation, and the range of the tests used to evaluate the scope and accuracy of a condensation formulation, should be discussed. Any appearance of lost chemical resolution caused by the condensation should be listed, including mention of any explicit species that can no longer be represented in the condensed formulation. Finally, as above, enough information should be reported so that the evaluator's calculations can be duplicated.

iv) Significant assumptions, estimates, perceived or measurable limitations, and other areas of uncertainty should be indicated at any particular development and condensation step requiring scrutiny. Some have been listed above. In particular, significant uncertainties can arise due to limited knowledge about chamber conditions and measurements; and most significantly, uncertainties arise from the need to move on in the development and condensation process, even without good supporting information.

Reporting That Principal Mechanism Is Independent of Chambers

It is clear that the principal mechanism cannot be formulated and tested alone—chamber-dependent auxiliary reaction mechanisms must be added to the principal mechanism for each chamber used in adjusting principal mechanism parameters and used in testing the principal mechanism's accuracy. Given this situation, a major claim that must be established by the mechanism's developer is that, in the creation of the principal mechanism's inner workings, its formulations were not corrupted by the principal mechanism's union with the various chamber-dependent auxiliary mechanisms—that is, that the principal mechanism is *in every way independent of the chambers used in its formulation.* Thus, the formulator must establish that there are no compensating errors between the principle mechanism and the different chamber-dependent auxiliary mechanisms he used. A similar statement is needed for the principle mechanism and the chamber operations models, especially for those models that used inputs selected by the developer because they were not measured by the chamber operators.

Authors of existing mechanisms have not often made this claim about their principal mechanism and most have offered no evidence at all that their principal mechanism is free of such errors. In evaluating a mechanism, the evaluator must determine whether such efforts were made, and, in any case, he must certainly collect sufficient information to allow

himself to estimate whether the principal mechanism's formulation is independent of all the chambers used to formulate and test the principal mechanism.

If the developer failed to document adequately his chamber-dependent mechanisms, it might be quite difficult to conclude that the principal mechanism's formulation was free of compensating errors occurring between its internal parts and those of the chamber-dependent reactions. If the developer did document his auxiliary mechanisms, the evaluator must understand the details of each one that was used in the mechanism's formulation and testing. He must do this by explaining and justifying the origins of each reaction and parameter in each auxiliary mechanism that was used.

Uncertainty And Limitations

Significant assumptions, estimates, perceived or measurable limitations, and other areas of uncertainty that the evaluator finds should be reported. In particular, the report should address mechanism aspects that could not be well tested in the original formulation, either because the range of compounds and types of experiments were limited, or because experimental conditions differed from those of the intended application. The report should comment on the extent to which these items will be addressed in the current evaluation.

Fixing Mechanism Flaws

If it is determined while carrying out this assessment, that the mechanism (be it old or new) contains an inaccurate representation relative to current consensus kinetics knowledge, then, by the end of the mechanism documentation phase, the evaluator must reach a judgment as to the importance or relevance of this flaw to the purpose of the evaluation. That is, the evaluator must decide:

a) that it is reasonable to "act as if" the problem would *not* influence predictions of relevant species (the evaluator, however, should produce evidence that this assumption is valid during the comparisons of predictions and observations to follow); or

b) that the problem is bad enough that the mechanism should be "fixed" before any comparison of its predictions with observations because the mechanism could have internal compensating errors arising from formulating (or adjusting) parts of the mechanism when some other part was inaccurate, and thus any accuracy it exhibits in later comparisons of its predictions with observations would be compromised by the possibility that such results were due to compensating errors that might not compensate in the final ambient application, or

c) that the problem is bad enough that the mechanism probably cannot be fixed, and therefore it is useless to continue with comparisons of the mechanism's predictions

with observations, because any result (i.e., fitting or lack of fitting in any particular comparison) could be 'explained away' and therefore no meaningful conclusions could be reached.

When there is a *known flaw*, the question is, "If the mechanism is fixed, will the predictions of relevant species change?" If the answer is, "Yes" or "Probably" (the evaluator may have to conduct sensitivity tests to answer this question), then the choice of proceeding with the original mechanism must be seriously questioned: *a mechanism that achieves its accuracy via compensating errors is unacceptable.* This is because when a mechanism which is sensitive to a parameter and the original setting of the parameter is now known to be wrong, then any good fits it produces in chamber simulation tests could only have occurred because *some other part of the mechanism is also wrong.* This *unknown* wrong part is compensating the *known* wrong part to produce a prediction that agrees with the observations, but while it appears that the mechanism is accurate in this test, it may not be accurate for different conditions such as those in ambient air applications. What warrants the extension of a mechanism's accuracy in chamber predictions to the very different conditions in the ambient air is the belief that the inner workings of the mechanism are also as accurate as we know how to make them. Thus, when a mechanism has a known flaw in its inner workings, the ability of the evaluator to make a case that it would be reasonable to act as if the mechanism will perform accurately in ambient air applications is seriously weakened.

Thus, the reporting and documentation of the principal mechanism must locate and describe any such possibilities of such internal compensating errors. The evaluator must make a declaration of his understanding of the status of the principal mechanism relative to this issue. If he finds that it is reasonable to act as if the principal mechanism is formulated as well as we know how to make it, then he should continue with the simulation and comparison phases. If he finds that the principal mechanism has a known flaw, however, he must produce a complete argument why he is pursuing one of the following courses of action:

1) to continue with the evaluation because although the mechanism has some now known flaws, it is reasonable to act as if these problems would not affect predictions of the relevant species;
2) to halt temporarily the evaluation and repair the now known flaws in the principal mechanism's formulation, so that further evaluation may reveal where other significant flaws in the mechanism's formulation remain; or
3) to halt permanently the evaluation of the principal mechanism at the present stage because sufficiently large flaws have been found that would discredit any agreement of observations from chamber data with simulation predictions produced by the principal mechanism, and therefore it would be useless to continue to the next evaluation stage with this mechanism.

Chamber-Dependent Auxiliary Mechanisms

Chamber-dependent auxiliary mechanisms include additional chemical reactions and mass-transfer descriptions that need to be added to a principal mechanism to simulate chamber experiments. This is because the chamber surfaces introduce heterogeneous, or wall-related, processes into the chemistry that occurs in chambers. While the principal mechanism is expected to describe the homogeneous, gas-phase chemistry in all chambers, a unique set of chamber-dependent reactions is needed to describe the chemical transformation processes unique to a particular chamber. The only linkage of these two mechanisms is through chemical species common to both mechanisms. The processes that are responsible for the chamber-dependent transformations are highly uncertain, but it is now believed that these common species may include HONO which leads to $^{\cdot}$OH, a radical central to *all* principal mechanisms. In addition, the number of chamber experiments available for development and testing of chamber-dependent mechanisms is limited. Nevertheless, because the effects of these processes are easily detectable, a chamber-dependent mechanism must be used for each smog chamber that will be simulated in later stages of the evaluation.

In evaluating existing mechanisms, different developers have produced different auxiliary mechanisms depending upon their hypotheses about the underlying wall-mediated process. Because principal mechanisms are supposed to be independent of the chambers used to formulate them, the evaluator may be able to use: (1) the developer's auxiliary mechanism, (2) another modeler's auxiliary mechanism, or (3) one he himself formulates.† The ideal is to use the auxiliary mechanism that has the most predictive power in chamber simulations. Another option is to implement and test several existing auxiliary mechanisms.

To use an auxiliary mechanism, the evaluator must present evidence and arguments that his selection is an *acceptable* auxiliary mechanism. That is, it is as well-formed as we currently know how to make it, and it shows good predictive accuracy when evaluated with special chamber experiments intended for this application. Thus, because of its influence on the predictions of the combined mechanism, an evaluator must judge the quality of each chamber-dependent reaction mechanism he selects with the same enthusiasm he judges the principal mechanism.

The following sections describe the types of documentation and reports that must be supplied by the evaluator about the auxiliary mechanisms he selects for use.

Auxiliary Mechanism Operational and Formulation Reports

As with the principal mechanism, the evaluator must report the contents of the auxiliary

† The *development* of auxiliary mechanisms is not included in the guidance give here.

mechanisms he will use; he must report that he understands their formulation strategies and data bases. Items that must be included in the report are:

- Operational Issues:
  1) A complete description of the chamber-dependent auxiliary mechanism for each chamber, including reaction listings, kinetic expressions, and important entities and parameters so that each mechanism can be implemented in a computer chemistry solver.
  2) If auxiliary mechanism parameters or input conditions depend on conditions or concentrations that must be supplied external to the mechanism (e.g., the concentration of a wall-absorbed material, or a wall-reaction rate that varies from run-to-run), then a consistent, mechanism-independent methodology to derive these values must be documented, and the values chosen must be clearly stipulated in the simulation inputs.
- Formulation Issues:
  1) A description of identified and suspected chamber-related processes at each facility. Ideally, this would include discussion as to the physical and chemical reasons why specific processes are expected to occur. Where available, references to published sources of data should be included, but these alone may be inadequate to meet the reporting needs described here. The discussion should at least consider surface loss (sink) rates, heterogeneous reactions, surface-enhancement of heterogeneous reaction rates, and unique chamber source (contamination) reactions, especially for radicals and NO_X.
  2) The details of any empirical chamber-dependent mechanism parameters (not completely dependent on chemical or physical principles), including: the rationale for the formulation of these parameters in real-world terms, the reason the parameter is constant or variable, the method of variation if variable, and the range of possible values. These parameters must be described in enough detail so that they may be utilized in an operational mechanism.
  3) A complete description of any chamber-dependent mechanism formulation and refinement decisions made using chamber development evaluations. The evaluator must determine whether the developer made reasonable choices. As described in the above section, the supporting data and rationale behind specific decisions should be documented and demonstrated.
  4) Known or perceived limitations in the chamber-dependent mechanism formulation or potential application should be reported. Since the main application of the chamber-dependent mechanism is to simulate chamber-related processes in high concentration VOC/NO_X smog chamber experiments used to develop the principal mechanism, it is also the responsibility of the mechanism developer (or

without these actions by the developer, it is the responsibility of the evaluator) to provide not only a compilation of assumptions and uncertainties in each proposed chamber-dependent mechanism, but also to provide insight as to how such limitations might induce uncertainty during principal mechanism development.†

† The reason this is significant is as follows. Different chamber-dependent mechanism formulations will generate different wall-radical and wall-NO_X compounds from chamber walls at different rates. The rates of other processes, such as deposition or heterogeneous transformations, may also differ. These dissimilarities will cause different concentration-time characteristics for the significant wall-radical and wall-NO_X entities. While chamber characterization experiments may not be sensitive to all of these differences, high concentration VOC/NO_X experiments *could* show significant variations.

4

Reporting Chamber Data

Figure 3 in Chapter 1 presents a chain of arguments leading to the decision to use a given chemical reaction mechanism in EPA models. Three general claims that must be confirmed in this chain are:

1) each Chamber-dependent auxiliary mechanism is well-formulated;
2) the Principal Mechanism is well-formulated; and
3) the Principal Mechanism has predictive power.

The primary evidence that would support each claim is chamber data. For chamber data to be evidence in these matters, they must exist and they must be relevant to the claims being made. That is, the chamber data must be capable of causing the evaluator to change his mind about the validity of the claims. Thus, the first condition that must be reported by the evaluator is the particular chamber data *needed* to conduct a compelling evaluation; the second condition to be reported is the *existence* of the needed data; and the third condition is the *relevance* of any existing data to the case being made. The foundations of these arguments rest on the credibility of the chamber data used, and if the evaluator is to produce a successful evaluation, a reviewer of the evaluation must be convinced that the chamber data used in the comparisons are believable. Therefore, the fourth condition the evaluator must report is chamber data *credibility* and why he believes it has this quality. The evaluator then must report the extent to which the available, relevant, credible chamber data met the data needs. Where there is a significant shortfall in available data, the evaluator must report on how he will continue to conduct an evaluation when only part of the chamber data that is needed is available.

What Chamber Data Are Needed?

To build persuasive arguments to support the three general claims for a particular mechanism and for a particular evaluation purpose, the evaluator must first report the kinds of chamber data or specific chamber experiments that he believes would have to be accurately simulated to make a compelling case for the principal mechanism's acceptance. A compelling case is one in which the evidence (if it existed) would permit only one "reasonable" interpretation about the acceptability of the principal mechanism. This analysis should be independent of the availability of the chamber data needed to perform such tests. This reporting of data needed is very important to the process because it determines the standard against which the evaluator will measure his subsequent ability to test the mechanism fully.

In making this list of needed chamber data, there are several factors the evaluator must consider. First, the principal mechanism is composed of many mechanisms for different species and clearly it is possible that errors in one of these species' mechanisms could compensate other errors in another species' mechanism to produce seemingly accurate predictions in mixtures. Because of this possibility, evidently a good strategy for demonstrating well-formedness would be to examine the quality of each species' mechanism's predictions separately before examining the quality of several species' mechanisms operating together in a mixture where inter-species compensation could occur. A closely connected second testing-strategy would be to test all the product species separately before testing any parent species. Testing model or lumped species in the principal mechanism presents special challenges because it is not clear what observations can be compared with the lumped species predictions. A good strategy here might be to test the lumped species' predictions against several individual 'representative species' included in the lumped-species class and then to test the mechanism's predictions against mixtures of these species. Additional tests would be mixtures of single species drawn from each lumped species class and then mixtures of mixtures.†

Another factor that must be considered when making up this list is the need to demonstrate that the principal mechanism is independent of any particular chamber. The issue is that chamber-dependent auxiliary mechanisms have to be combined with the principal mechanism to fit each chamber, and inaccurate versions of these auxiliary mechanisms will influence the predictions from the principal mechanism. This is another source of compensating error (e.g., the auxiliary mechanism makes too many radicals while the principal mechanism makes too few radicals). Similar statements can be made about the effects of the

† For example, current condensed mechanisms seem to be having problems when mixtures of aromatics are added to mixtures of paraffin and olefin species. That is, ozone observations and predictions agree when mixtures of paraffin and olefins are tested, but when mixtures of aromatics are added to the same paraffin and olefin mixture the ozone predictions become too large.

chamber operations conditions on the predictions of the combined mechanism (e.g., too high an actinic flux). The necessary demonstration of chamber independence for the principal mechanism can be achieved by using the one principal mechanism with several different auxiliary mechanisms to simulate several chambers successfully, where each chamber has different wall and operating characteristics. The more chambers that are simulated successfully this way the less likely that the same compensating errors occur between the one principal mechanism and the different auxiliary mechanisms. To be minimally sufficient, demonstration of chamber independence requires data from at least four chambers, two indoors and two outdoors. Furthermore, the testing of the auxiliary mechanisms in each of the chambers requires special chamber characterization experiments that are used to evaluate and set the parameters in the chamber-dependent auxiliary reaction mechanisms.

In any case, the evaluator should report a list of evaluation comparison tests, ordered by explicit and model species in the principal mechanism, that he would conduct to evaluate the principal mechanism in an ideal data world. That is, the list should be independent of the availability of the chamber data needed to perform the test.

What Chamber Data Are Available?

Chamber data to perform some of the needed comparison tests do exist and these may also be in an acceptably documented state that would allow them to be used in the evaluation. The evaluator's next task, therefore, becomes one of identifying the specific existing chamber experiments that relate to his already listed needs, and to report their identities on the same list that was produced above.

For an evaluator to use chamber data to test a principal mechanism certain information about the chamber data must be known and must be reported in a reasonable format. Much of the needed information can only be obtained from each chamber operator. Four large data sets that will meet the standards required for use in testing by procedures described here are being prepared by the research centers that produced the chamber experiments. These are: the Dual Outdoor Smog Chamber data set from the University of North Carolina; the Indoor Teflon Chamber data set and the indoor Evacuable Chamber data set from the Statewide Air Pollution Research Center, University of California, Riverside; and the Dual Outdoor Smog Chamber data set from the Commonwealth Scientific and Industrial Research Organization, Sydney, Australia. Other chamber data may also be brought up to this standard soon.

Regardless of who is documenting the data or what data are supposedly valid, it *is* the evaluator's responsibility to demonstrate to the sponsor that the data he selects are of usable quality and that they meet the standards he needs for the purposes of his evaluation.

Of course, the mere existence of a particular chamber experiment is not sufficient to allow the evaluator to continue; he must establish that the particular experiment is relevant

to the specific and general claims he is trying to establish, and he must establish that the experiment has competence to challenge the simulation predictions. This may reduce further the number of available chamber experiments that can be used in the evaluation.

Are The Chamber Data Relevant?

In Chapter 1, we offered a warrant for why chamber data were relevant in a general way to the generic claims listed above. When the evaluator can articulate what good or bad agreements between *particular* chamber observations and their correlated simulation predictions have to do with his opinion about one of the generic claims, then these chamber data become relevant to establishing the validity of one of the three claims (and thus become evidence). Therefore, the evaluator must produce experiment-specific warrants that explain the relevance of each experiment he chooses to the claim it is being used to support. That is, what does the evaluator expect to gain from the use of a particular experiment to support one of the claims?

Once warrants are established that explain why certain chamber data are relevant to one of the three general claims, the evaluator must then go on to determine that these chamber data are both competent and credible. If it is concluded that this is so, then this chamber experiment's data can be used in meaningful comparisons with simulation predictions and reviewers and the sponsor will know why.

Credibility of Chamber Data

The foundations of the arguments discussed above rest on the credibility of the chamber data used, and if the evaluator is to achieve the goals of the program, the data's trustworthiness must be established before simulations are performed. Hence an important claim shown in the bottom of Figure 3—that credible chamber concentrations exist—must be defended by the evaluator. To support this claim, the evaluator must satisfy himself and others that it is reasonable to act as if this assertion about the chamber data is true, that is, he must determine that there is support for what is otherwise a chamber operator's testimony that the "numbers in the data file are accurate." Thus, ancillary evidence about the main evidence must be produced to establish the credibility of the main evidence.

An evaluator should no more use chamber data without judging and reporting on its credibility and quality of production than he would accept a principal mechanism without judging and reporting on its use of consensus kinetics data and the quality of the reasoning in its formulation. Yet, no past evaluation has ever established the quality of the chamber data before reaching a conclusion regarding the acceptability of the principal mechanism. Only recently have current chamber data been scrutinized sufficiently to reveal major flaws

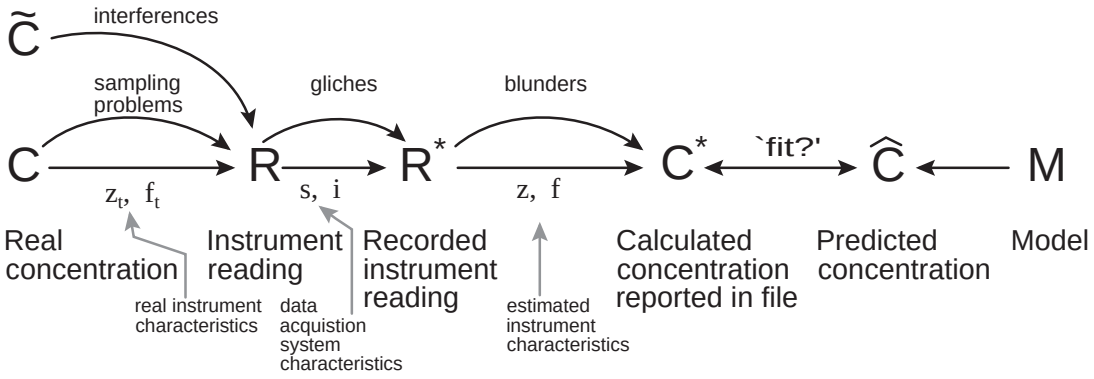

Figure 5. The Relationship Between Chamber Concentrations and Data Reported and Between Reported Concentrations and Simulation Predictions.

in some reported chamber concentrations. In some cases, these flawed data had direct influence on the formulation of parts of a principal mechanism.

What Makes Reported Concentrations Credible?

In dealing with the chamber data files, the evaluator must be careful to distinguish between the *report of the concentration in the data file* and the *concentration that was in the chamber.* Figure 5 shows one example of the sequence of transformations that are applied to continuous instruments such as the NO_X and O_3 monitors. Interferences, sampling problems, glitches, inaccurate analog to digital converters, inappropriate span and zero factors, and just plain blunders all stand between the in-chamber concentration and the number in the data file.

The usual ways an evaluator would come to believe that reported concentrations were accurate and precise representations of the chamber concentrations are that chamber operators would have presented evidence and an argument that establishes this claim for each data stream in each experiment and for the whole data set through use of recognized quality assurance principles and methods. Additionally, the quality assurance methods themselves must be backed by supporting evidence that includes calibration histories, histories of measurement performance and operating conditions, and sufficiently precise replicate results. To make their data acceptable for use in evaluations, the chamber operators must supply the statistics for these supporting data along with the concentration data streams themselves. The evaluator must assure that this information *is* available for the data he is using.

Besides the usual statistical measures, the evaluator should have general knowledge about how the data were produced and processed in each chamber data set, including characteristics of the instruments used, and the frequency of calibrations and quality assurance activities. The evaluator may also find it useful to have a general understanding of what

care and attention was given to all the stages of producing the data such as how the chamber operators track the items in Figure 5. The evaluator can obtain this information from supporting documentation supplied with the chamber data sets by the chamber operators.

The second way users come to believe chamber data to be reliable is that the reported results are shown to be repeatable.

Replicate chamber experiments most powerfully demonstrate the reproducibility and precision of chamber data and therefore can be a critical force in convincing evaluators and reviewers that a particular set of chamber data is credible. While the evaluator may not simulate all the multiple replicate experiments available in the chamber data sets, he should report the number and type of replicate experiments in each chamber data set and he should report some plots of replicate experiments from each chamber.

Categories of Data Credibility

There are some chamber data that are potentially useful but which did not have supporting data produced contemporaneous with the data collection. Some of these data have been "rehabilitated" by re-creating supporting data after the fact. While such actions allow the use of some chamber data that would otherwise be totally lost, experiments for which observational accuracy can only be inferred by totally indirect means, *must be assessed separately* in mechanism evaluations. This is because such 'evidence' is much weaker than fully supported chamber data and issues of fit or lack of fit to these weaker data should not be mixed with those from experiments with more precisely documented accuracy.

Because we would expect higher quality (i.e., more accurate and more precise) concentrations to provide more stringent tests of the accuracy of simulation predictions, the subsequent comparisons described in Chapter 5 should be done in categories of documented credibility for the relevant species concentrations. Thus, the evaluator must separate the selected chamber data he will use into quality categories according to his judgment of which experiments have better established concentrations than others and therefore should be given more weight. The evaluator must report these quality categories for each relevant species. Note that a particular experiment may be in different quality categories, depending upon which relevant species is being considered. Thus, the reported concentrations of ozone may be considered as high quality, while in the same experiment, the reported concentrations of formaldehyde may be considered of lowest quality (e.g., very imprecise).

In this regard, it would also be valuable for the evaluator to report when he believes that the chamber data available for use are too poor for the task at hand and should be placed in lower tiers of data competency in the data base.

Should Some Chamber Data Be Discounted?

There is another dimension for assessing chamber data: their fitness to challenge adequately the simulation predictions. While the individual data streams in an experimental data file might be very accurate and precise, other components of the data system might seriously weaken the competency of the experiment for testing mechanisms. This might happen in several ways:

i) some influential chamber operating condition could be too poorly characterized to be accurately specified to the simulation program (e.g., extremely humid conditions with possible condensed water on the chamber walls);

ii) the initial reactant conditions could be significantly different from those in the intended application of the principal mechanism (e.g., VOC composition very different from ambient and beyond the range of the condensation scheme)

iii) the meteorological or physical chamber conditions that prevailed during the experiment might be significantly different from those in the intended application of the principal mechanism (e.g., chamber temperatures at 45°C);

iv) the experiment might inexplicably differ significantly from its replicates or from its expected place in a series of variable conditions experiments; or

v) some other limiting factor

The reason such experiments have little competency to challenge simulation predictions is that any lack of fit between the chamber observations and the predictions could easily be "explained away" as due to an influential input not being correctly specified in the simulation data file. Such experiments only confuse the focus of the evaluation and should either not be included among the data simulated or if included should be categorized separately from the clearly competent experiments.

To classify an experiment as having low competency, the evaluator has to produce evidence and an argument supporting his classification. To reclassify an experiment because the evaluator believes a particular simulation input is too poorly specified requires a specific kind of argument. That is, while the explanation that a particular simulation input was the cause of the lack of fit between chamber observations and simulation predictions may be true, such "explaining away" will not be acceptable as an argument without supporting sensitivity analysis to show that the simulation system *does respond* in the manner hypothesized and with the expected magnitude.

To the list of experiments he will use, the evaluator must add comments about any chamber experiments that, as a result to his investigation of the data, he believes have competency problems. These experiments should receive special treatment in comparison with simulation predictions, and most likely their evidentiary weight should be significantly discounted.

What Can Be Accomplished With Available Data?

A second important claim made about chamber data in Figure 3 is that the data are complete, or that the data are representative and varied.

It is very clear that, at present, the existing chamber data set falls significantly short of what is needed for a totally compelling evaluation. Therefore, the evaluator must first document the magnitude of the difference between the ideal and the actual condition (i.e., how many experiments of which type does he have and what fractions are these of the number needed). Then he must explain how he will advance his case given the shortfall in data (potential evidence). This plan should certainly recognize that the accuracy of some species' mechanisms may still be in doubt when the evaluation is completed. Nevertheless, the plan is the basis of a successful evaluation and the evaluator and the sponsor must agree about what can be achieved given the state of available chamber data.

Selecting and Preparing the Summary Observations

Once the chamber data set is selected and agreed upon, the evaluator must produce for each experiment the *summary* observations for comparison with simulation predictions. These values are derived from the data streams in the chamber data files. *While these alone are insufficient to document the fit between observations and predictions,* they do provide one relatively easily assimilated view of the data.

Because our guidance is for oxidant mechanisms, the relevant chemical summary statistics include—but *are not limited to*:

1) the maximum change in ozone concentration over the course of the experiment (i.e., the ozone maximum concentration);
2) the maximum change in nitric oxide concentration over the course of the experiment;
3) the time for the maximum ozone concentration change;
4) the time for the maximum nitric oxide concentration change;
5) the slope of the ozone-time profile centered around the half-change in ozone concentration point (e.g., the difference between [O_3] at 60% of its maximum and [O_3] at 40% of its maximum divided by the time between the same two points; this is a good estimate of the maximum rate of O_3 production);
6) the slope of the nitric oxide-time profile centered on the half-change in nitric oxide concentration point (this is a good estimate of the maximum rate of NO consumption); and

7) the time from the start of the experiment to when nitric oxide and nitrogen dioxide are equal (this is called time to NO-to-NO_2 cross-over and is a good estimate of the initial radical production).

Each of these statistics is based upon a *single* reported concentration-time stream. The qualities of the reported data are therefore directly transferable to these measures and there is no possibility of compensation between two measures. The measurement techniques for NO and O_3 are interference-free; chemiluminsent NO_2 measurements are also interference-free till well after the NO-NO_2 cross over point. The measures describe the minimum qualities of interest in assessing "goodness of fit": maximum value, time to events, (from which it is easy to compute average rates) and maximum rates of production or loss. These measures, computed by the chamber operators, should already be in the documentation section of the chamber data files.

These single measures can be combined if desired, but the individual values above must be reported also.

Measures that combine two interconvertable species or two species perhaps measured on different instruments have been frequently used in past evaluations. While these statistics are useful, they *do* have some inherent weaknesses that can result in misleading impressions of goodness of fit. For example, some evaluations have measured oxidant production by used of $\Delta[O_3] - \Delta[NO]$ instead of just $\Delta[O_3]$ because this measure relates more directly to the underlying process of $RO_2^{\cdot}$+NO.† While this may be true, when comparing chamber observations and simulation predictions, it is preferable to use the directly observed variable as the *primary* measure of "fit" rather than a lumped or derived measure. The reason is that too low a value for $\Delta[NO]$ can be off-set by too large a value for $\Delta[O_3]$ in the simulation, yet the combined measure from the observations and the combined measure from the simulations would show good agreement. Thus, subtle flaws in the principal mechanism or in the chamber conditions may be hidden by such combined measures.

Another combined measure that has been used is a measure of integrated $^{\cdot}OH$ radial concentration. This value can be computed in experiments with known dilution and measured values of a sufficiently rapidly reacting compound that reacts only with $^{\cdot}OH$, or in experiments with unknown dilution and measured values of two such compounds that react with $^{\cdot}OH$ at different rates (this includes most experiments with aromatics present). Again,

† One argument frequently advanced for the use of combined measures is that the relative errors, E, between the observations (O) and the predictions (P) which is $E = (P - O)/O$ are more meaningful for a measure like $\Delta[O_3] - \Delta[NO]$ than for say just $\Delta[O_3]$ when the observations are very low and the absolute difference $(P - O)$ is larger than O, but is still smaller than a 5–10 ppb. This is a reasonable concern, but this problem can be readily solved by just testing for a low threshold value of $\Delta[O_3]$ and switching the calculation of relative ozone error to $\Delta[NO]$ instead.

because it involves calculations with concentrations from different sources (e.g., a tracer concentration on one gas chromatograph and a HC concentration on another gas chromatograph), this measure may have a large uncertainty and should be used with caution.

There are other chemical statistics that have been used in previous evaluations. Any of these may be *added* to the list of measures given above, but they can not replace the required measures.

Similar summary statistics for other relevant species in the evaluation (e.g., HCHO or PAN) should also be described. The evaluator should operate on the chamber data files to extract these measures and produce the appropriate values.

To the list of experiments the evaluator has been building throughout this chapter, he should add the observed summary measures (and any flags relevant to their interpretation or competency). In the next chapter, the evaluator will add the predicted values of these measures and will compare these with the observed values.

5

Reporting Simulations

The first item that must be reported about simulations is which experiments from which chambers will be simulated. This can be accomplished by the evaluator producing a refinement of the list generated in the previous chapter. For the needs here, the list should contain at least the following items:

a) experiment identification (or data base identification number);

b) the general class of the experiment (i.e., characterization, single species, simple mixture, complex mixture)

c) initial total concentrations of VOCs and NO_X;

d) initial and average or maximum chamber conditions;

e) the judged competency of the whole experiment to challenge simulation predictions (as produced in the last chapter);

f) whether the experiment was used by the developer to set parameters in the principal mechanism; and

g) for each relevant species in the evaluation, an indication of whether a data stream for this species is included in the file, its quality category as judged in the last chapter, and its summary statistics.

In this chapter recommendations for adding items to this list will be made. We will refer to this list as the "experimental conditions data base."

After building the experimental conditions data base the evaluator must prepare a simulation data input file for each experiment in the list. These provide the simulation solver program with:

1) identity of the combined mechanism to be used by the solver;

2) the initial concentrations of all relevant species present in the chamber at the beginning of the experiment (in the form that the mechanism represents them);

3) the in-chamber pressure;
4) and, given as a function of time for the duration of the simulation:
    a) either the in-chamber actinic fluxes or the species photolysis rates pre-computed from the in-chamber actinic fluxes;
    b) the in-chamber air temperature and water vapor concentration;
    c) the chamber's dilution rate;
    d) the variation of any chamber-wall mediated reaction rate or parameters.

The file usually also contains simulation control statements such as error control, maximum step size, and often a list of which species concentrations to print.

The first part of this chapter describes the reporting requirements for the production of the simulation inputs. The rest of the chapter describes the initial statistical analysis of the simulation outputs.

Creating Simulation Input Files

In Figure 4, there were a number of operations and models that were shown in-between the chamber data files and the simulation input files. These were described as the *chamber conditions models* and they are usefully classified as those that are *chemical species related*, and those that are *physical environment related.*

In the former category are the methods for specifying the initial concentrations of the necessary input species, and the species photolysis rates. These methods will differ with the choice of both principal mechanism and the chamber. This is because each principal mechanism has a different set of explicit and model species and their concentrations have to be produced from the concentrations of the actual species injected into the chamber. The principal mechanism must have rules for carrying out this step.†

Each chamber also has a different set of actinic fluxes that must be combined with the mechanism-species' cross-sections and quantum yields to produce a chamber and mechanism specific set of species photolysis rates.⋆

† Some solver programs may perform the transformation of explicit species to model species internally just before starting the simulation. These solvers would then need the explicit species from the chamber data file given in the simulation input file (along with the rules to be applied), rather than the model species.

⋆ Some solver programs may perform the calculation of species photolysis rates during the simulation rather than in a separate step before the execution of the solver program. These solvers then need the specific chamber actinic flux file supplied to the solver for each experiment simulated.

In the second category of chamber conditions models are those simulation data file inputs that specify the other chamber physical conditions during the experiment. These may be specified as a function of time and include the chamber air temperature, and the chamber water vapor concentration. These are mechanism independent, i.e., they would be the same no matter which chemical mechanism was being tested.

It is unfortunate, but even the best reported chamber data are incomplete in all the terms needed to create simulation input files and the evaluator must report his strategy for filling in the needed items. Therefore, it is useful to distinguish the origins of the various simulation data file values by considering whether they were:

a) directly based on observations taken during the chamber experiment,

b) based on observations taken at various times before and/or after the experiment and therefore would have to be estimated during the experiment,

c) based on values typical of other chambers where they may have been measured previously, or

d) based upon the evaluator adjusting inputs after seeing how they effected the simulation predictions.

Initial Concentrations

The operator-injected species concentrations are always specified in the chamber data file (or, in some experiments, in supporting ancillary data files).† Species that are part of the chamber filling air are often not measured and therefore are not included in the chamber data file, but they are sometimes estimated by the chamber operators. Similarly, some species are created by reactions that occur during injection, for example, nitrous acid. Nitrous acid is very infrequently measured and usually must be estimated.

Therefore, there are usually three distinct techniques that must be developed and justified by the evaluator for use in specifying initial conditions for the simulations (these are collectively called the Initial Concentration Algorithms, see Figure 4):

1) Algorithms for use when the input values are observed values reported in the chamber data file, for example, the reported values for the injected compounds. This type of input cannot be changed by the evaluator unless he can show that there are errors in the data

† For example, for constant composition mixtures such as those injected from VOC mixtures in high pressure, high concentration gas tanks, the chamber data file may specify the total injection of the mixture and the mixture composition may be specified in a separate data file. Where the in-chamber mixture is made up of individual species injections, however, the initial concentration of each species is usually given in the chamber data file because the mixture composition may vary significantly from experiment to experiment.

base and its documentation. Note that for initial concentrations of model species in the simulation data files, the reported values in the chamber data files have to be transformed by whatever method the principal mechanism uses to establish correspondence between the model and the real world, e.g., conversion of paraffin to PARs in the Carbon Bond Four mechanism.

2) Algorithms for estimating initial concentrations of species that were not measured, but which the chamber operators are sure are present in the chamber, for example, H_2 or CH_4 or background air VOCs. Such initial values may be estimated by the chamber operators without knowledge of which mechanism will be used to simulate the data, and may be in the form of ranges of values, without a specific value recommended. In this case, the evaluator must develop and report a procedure for selecting a value within the range.

3) Algorithms for estimating initial values of simulation inputs which can be, or are, modified after an initial simulation and based at least in part on how the inputs affect the simulation results, for example $[HONO]_o$ values, or levels of wall-loadings for emitted species. While such initial values or ranges of values may be recommended by chamber operators, setting their values and reporting the methods used to estimate them are entirely the responsibility of the evaluator.

These algorithms are crucial elements in an evaluation. They must be fully explained and justified in the evaluator's reports. Where input values are estimated by the evaluator, the final adjusted values must be added to the experimental conditions data base and summary statistics including the geometric mean,† the geometric standard deviation, and the maximum and minimum values of all estimated initial conditions must be computed, reported and discussed by the evaluator.

Chamber Actinic Flux

One would ordinarily think that measuring the physical operating conditions of the chamber and reporting these values would be entirely the responsibility of the chamber operators. Unfortunately, there is no method in current use for measuring actinic flux or spherical spectral radiant flux density (a three-dimensional spectral measurement) during an experiment. Therefore for the chamber data to be useful to evaluate reaction mechanisms, the operators must supply sufficient auxiliary data that the actinic flux can be *computed* after the experiment, either by the operators themselves or by the evaluator. If the chamber operators computed the actinic flux, the evaluator has a responsibility to determine that their calculated values are both documented and acceptable for his needs. Chamber data lacking the supporting data for calculating actinic flux cannot be included in the evaluation. The required data varies depending upon the light source.

† These particular statistics are requested because the expected distribution of the variables is log-normal, which is characterized by a geometric mean and geometric standard deviation.

For indoor chambers, where the light source is constant throughout an experiment, it is sufficient to provide data for actinic flux calculations that were taken before and/or after the experiment and thus are not reported as data streams in the data files. In this case, histories of the values and trend analysis are necessary to support the stability assumption. For outdoor chambers, the primary auxiliary data must be taken during the experiment and typically consists of broad band radiometer (i.e., TSR and UV) readings that are included as time-dependent data streams in the data files. In this case, reports of quality assurance activities and histories of radiometer calibration are necessary to support use of reported values in chamber data files to derive actinic flux. In either case, additional auxiliary data necessary to complete the calculations must be included in the documentation sections of each chamber data file, but some auxiliary data, particularly chamber irradiance spectra, may be supplied as separate computer readable files associated with the chamber data.

Actinic Flux in Indoor Chambers

The best measurement that can be done in chambers now is to measure spectral radiant flux density on a horizontal surface. Because it is only over the upper hemisphere and has a cosine-law response weighting, however, this measurement represents much less than the total actinic flux in the chambers. In most indoor chambers with their uniform banks of lights, this spectral irradiance measurement can, however, represent an accurate *relative* spectral radiant flux density. The in-chamber spectral measurement must:

i) cover the spectral region from 300 to 850 nm with a resolution of at least 2 nm;

ii) be performed at a minimum of five locations in the chamber (e.g., center and four corners of one side facing the lamp bank);

iii) be taken throughout the typical duration of a chamber experiment;

iv) be appropriately averaged for the chamber and be presented in graphical and computer readable forms;

v) include ancillary data supporting the accuracy of the measurements.

The evaluator must assure that the relative chamber irradiance spectrum conforms with these requirements. Indoor chambers lacking such spectra cannot be included in the data base and cannot be used in the evaluation.

To convert this relative spectral measurement into an absolute actinic flux in the chamber, the spectrum must be calibrated using a measured species photolysis rate inside the chamber. As with other reported chamber values, the evaluator should be careful to distinguish the reported value of the species photolysis rate from the actual value of the species photolysis rate in the chamber. The reported photolysis rate is actually the results of fitting a reaction mechanism model to measured concentrations of the photolyzing species in a special exposure volume and inferring from the model's parameters the species' photolysis

rate that would be consistent with the measurements. Requirements for reporting these values are:

i) the physical system must be described, including drawings illustrating where the measurements occur relative to the lamps and the reaction vessel;

ii) the reported values of the measurements are subject to the same criteria for accuracy and precision as is the chamber data; these must be established by ancillary data;

iii) the exposure apparatus has certain parameters which must be accurately known; this must be supported by ancillary data; and

iv) the reaction mechanism model has reactions and kinetic rate constants that are subject to the same criteria as the principal mechanism and these must be supported by documentation.

Because rate constants for some reactions used in these calculations have changed during the last 20 years, rather than merely reporting the calculated species photolysis rate, a data base reporting the raw species measurements should be maintained by the chamber operators and available to the evaluator. This allows the evaluator to use newer kinetics parameters to re-compute the older photolysis rates if needed, and it provides an indication of the variation in measured values and hence the uncertainty in the computed rates.

The evaluator is responsible for establishing the credibility of the relative chamber irradiance spectrum and the reported species' photolysis rate used to compute the in-chamber actinic fluxes for indoor chambers.

Actinic Flux in Outdoor Chambers

Outdoor chamber data must include data streams from at least a broad band or total solar radiometer, and normally would include data streams from a broad band ultraviolet radiometer (these standard meteorological instruments measure the solar irradiance averaged over different wavelength regions). The sensitivities of the TSR and UV radiometers are adequate for detecting the variations that occur in atmospheric transmissions.

Because the extraterrestrial solar flux is quite constant, and because variation of the factors affecting solar transmission through the atmosphere is not large, the surface solar irradiance spectrum tends to exhibits only a small variation. Radiation transfer models use the extraterrestrial solar flux and meteorological and aerosol conditions that occur during the radiation's passage through the atmosphere to compute the surface spectral and broad band irradiance as a function of time. If a radiation transfer model can accurately simulate the surface irradiance measured by the two broad band radiometers operating at both short and middle wavelength spectral regions, it is *inferred* that the model will accurately predict the spectral irradiance and actinic flux at the surface.†

† In tests of the radiation transfer model in use at UNC, the model showed excellent predictive power when

Most necessary input parameters for radiation transfer models can be obtained from standard meteorological observations (e.g., air pressure, total column O_3, total column H_2O, and relative humidity). If spectral irradiance is available on the day of the experiment, it can be used to determine the values of the aerosol optical properties, although without this measurement, the aerosol optical properties can be adequately estimated using the less sensitive broad band radiometer responses and the model. The broad band radiometer observations are also used to determine any "cloud transmission" functions needed for a particular day.

In addition to the TSR and UV data streams, outdoor chamber data must have the necessary parameters for the operation of radiation transfer models included in the documentation sections of the file. The user of these radiation models must provide ancillary data to support the accuracy of the models for predicting surface irradiance and actinic flux values. If the chamber operators ran the models and supplied the actinic flux values, the evaluator must review the calculations and report that he understands how they were done and that he knows the sources of the auxiliary input data used. If the evaluator is computing actinic flux himself, he must report the methods and procedures used, as well as where he obtained the necessary input data.

While the computation of the surface actinic flux is relatively straightforward given the TSR and UV radiometer readings and auxiliary data, computing the actinic flux *inside* Teflon chambers or bags is not. The reason for this condition is the light transmission properties of Teflon film. First, the transmission is a function of wavelength, with lower transmissions at shorter wavelengths. Second, and this is the real problem, beyond an incident angle of about 45°, Teflon film has a large reflectance and a much lower transmissions, with the transmissions decreasing to zero at 90° incident angle. If there is not a predictable relationship between the film surface and the sun's direct beam as a function of time, it becomes very difficult to estimate the film transmission (this would be true for example in a collapsing bag used as a chamber). Another necessary factor is to estimate the volume of the chamber receiving radiation through each Teflon area that is transmitting radiation into the chamber. A final necessary factor is to estimate the reflected radiation from any receiving surface inside the chamber, for example the reflective floor in the UNC Outdoor Chamber. Thus, a fully functional in-chamber Actinic Flux Model must have procedures for:

a) computing the surface actinic flux as a function of location and time of day;

compared with a large series of measured surface radiant flux densities. Further, in inter-comparisons of the UNC spectroradiometer with other high-quality solar spectroradiometers, some from the National Institute of Standards and Testing, the UNC instrument showed excellent agreement above 295 nm. This one instrument has been used to measure the spectral irradiance in all the chambers included in the EPA Standard Chamber Data Base.

b) computing the incident angle of the direct solar beam on the chamber light transmitting surfaces;

c) computing the transmission of the chamber's surfaces as a function of incident angle and wavelength;

d) computing the chamber volume receiving radiation transmitted by each receiving film surface.

Because this calculation is so crucial to the predictive power of the mechanism the evaluator must report his understanding of the model and its input data, and must assure himself and others that the computed actinic fluxes are computed as satisfactorily as we presently know how to make them.

Photolysis Rates

While the actinic flux ($\mathcal{A}_\lambda$ with units of photons-cm^{-2}-sec^{-1}-nm^{-1}) is a property of the chamber, the species absorption cross sections ($\sigma_{\lambda i}$ with units of cm^2-molecule^{-1}) and quantum yields ($\Phi_{\lambda i}$ unitless) are properties of the photolyzing species i in the principal mechanism. To provide the photolysis rates that are appropriate for each chamber experiment to be simulated the integral

$$j_i = \int_{\lambda_m}^{\lambda_n} \mathcal{A}_\lambda \sigma_{\lambda i} \Phi_{\lambda i} \, \mathrm{d}\lambda$$

must be evaluated (λ_m is the lowest and λ_n is the highest wavelength at which $\sigma_\lambda \Phi_\lambda > 0$). This integral is usually approximated by numerical methods, for example applying Simpson's Rule to the interval products. Such methods have to account for the difference in wavelength resolution between the actinic flux spectrum and the species cross section and quantum yield spectra, as well as the different forms of reporting the wavelength intervals for these data. Great care to minimize numerical errors must be taken when carrying out the approximate integration over wavelength (for example the use of a double trapezoid rule instead of the simple Simpson's rule when computing the area product of the actinic flux and the reaction potential spectra).

The evaluator must report the method he elects to use to compute the species photolysis rates and he must support the accuracy of the method.

Temperature, Water Vapor, and Dilution

These chamber physical conditions are almost always monitored by the chamber operators and data streams for these appear in the chamber data files. Therefore, in preparing simulation data input files for cases with these data, the evaluator only has to process the data from the form they are reported as to the form needed by the simulation solver program

(e.g., dew point temperature to water vapor concentration). For some experiments and some measurement systems, this step may involve the application of "smoothing functions" to average out some noise in the reported data (e.g., this is true for some dew point data in the early UNC chamber data files). Dilution of the chamber contents occurs in some chambers but not in others (i.e., the SAPRC ITC has a collapsing side that compensates for air samples withdrawn and thus there is no dilution in this chamber). In chambers that do have dilution, it is commonly measured by observing the loss of essentially inert tracer species such as carbon tetrachloride which were injected into the chamber at the start of the experiment. The chamber data files report the relative concentration of the tracer. The evaluator must convert these relative concentrations into whatever form of dilution his solver needs.

It is the chamber operator's responsibility to establish and report the accuracy and precision of these physical characteristics data. It is the evaluator's responsibility to assure that supporting data are available for the chamber data he uses. He must report his level of comfort with the measurements and the quality assurance support offered by the chamber operators. The evaluator must also report (in appendices) the algorithms used to convert the physical condition data from the chamber data file to the forms used.

Sometimes these data streams are missing in the chamber data file. If the evaluator still wants to include these experiments, it is his responsibility to justify a procedure for utilizing "surrogate" data to fill in the missing values. For example, data from another experiment with very similar conditions could be substituted, or averaged data from the same chamber data base could be used. The evaluator must show that the substituted data are reasonable values. In case the values may influence the quality of the simulation predictions, the evaluator must record flags in the experimental conditions data base that indicate that substituted data were used in the experiments. If a problem of agreement between the observations and the predictions is found in the later evaluation phase, these flags will alert the evaluator to consider whether the substituted data might be the cause.

Performing the Simulations

Once the simulation data input files are complete, they can be passed to the solver and the simulation output files will be created.

Reporting Simulation Outputs

Compute and Report Summary Statistics

A first step in reporting the simulation results is for the evaluator to compute the corresponding summary measures for the predictions as was done for the observations and add these to the list of experiments that was built in Chapter 4.

Calculate Relative Errors

Next, the evaluator computes for each experiment, absolute and relative errors for measures of amplitude changes by use of

$$\text{absolute error for amplitudes} \equiv E_i = P_i - O_i$$
$$\text{relative error for amplitudes} \equiv R_i = \frac{E_i}{O_i}$$

where

O_i is a summary observation measure from the i^{th} experiment, and

P_i is a summary prediction measure for the i^{th} simulation.

A positive error means the simulation *over-predicted* the observations. Absolute and relative errors for measures of time, however, are computed by use of

$$\text{absolute error for times} \equiv \hat{E}_i = O_i - P_i$$
$$\text{relative error for times} \equiv \hat{R}_i = \frac{\hat{E}_i}{O_i}$$

The reason the time errors are reversed from the amplitude errors is so that a positive time error means that the simulation is *fast* relative to the observations. Often poor simulations are both "over and fast" or "under and slow" and we wish to produce a correlation in sign for the standard measures.

Display Relative Errors

To provide ready access to the large number of error values just computed, they must be shown individually on various graphs. The selection and organization of the graphs are important. Warrants that explain the relevancy of each graph to the evaluation should be produced by the evaluator to support his use of such tools. Graphs that are not relevant to the argument should be excluded from presentation as they merely serve to confuse the observer.

For example, a scatter diagram of a measure from all experiments is poor choice of a graph because any impressions it may give of relative accuracy are potentially quite misleading due to various compensating factors producing the appearance of a central tendency that the human eye automatically seeks outs, but which may not exist in the data. A situation that could arise, for example, is that all the positive error values were from simulations of one chamber and all the negative error values were from simulations of another chamber. This is yet another form of compensating error leading to seemingly

good results. The possibilities of jumping to such conclusions should be avoided by not producing such graphs in the first place.

Another situation to avoid is a graph that mixes up data that have been judged of highest competence with data that has very low competence to challenge chamber data. Good or bad agreements with low competency, low quality data tell the evaluator little about the quality of the mechanism, but poor agreements with high competency, high quality data provides the evaluator with powerful evidence that the mechanism may have flaws. The evaluator would have already classified each experiment as to its quality and competency in Chapter 4, and this makes it easy to segregate the data by quality.

Thus, the first requirement for examining errors between observations and predictions is to sort the observed data set by the data quality and competency judgments already made. The *errors must be presented to emphasize data quality and competency first.*

The second requirement is that *the errors must be presented on a chamber by chamber basis.* This helps detect compensating errors between the principal mechanism and the particular chamber auxiliary mechanism, or between the principal mechanism and the various chamber conditions models, e.g., actinic flux in error. The evaluator has an obligation to show that the principal mechanism works accurately in each chamber. When this has been shown for four chambers (two indoors and two outdoors), then it is reasonable to act as if the principal mechanism was not contaminated by any chamber-dependent reactions.

Within each chamber's data, the third requirement is that *the errors must be presented by progressing from simple to complex individual species and from simple to complex mixtures to aid recognition of between-species and between-class compensating errors.* This helps establish that the mechanism is free from inter-species compensating errors.

Within each species or experimental class, *the errors must be presented by sorting from smallest to largest.* This helps find patterns of fit.

Also, if they are known, experiments that were used by the developer to set internal mechanism parameters should be indicated on the graphs.

One type of graph that can readily be prepared by spreadsheet programs and that we have found useful is shown in Figures 6 and 7.* To make this type of plot, the evaluator would:

* The data plotted in the figures is from Carter and Lurmann, "Evaluation of a detailed gas-phase atmospheric reaction mechanism using environmental chamber data,", *Atmospheric Environment*, **25A**:12, 2771, 1991.

i. plot on one graph all of one data quality of one kind of error† from one chamber for experimental classes of simple VOC species, and on another graph the same for VOC mixtures (an example of the latter is not shown);

ii. layout each graph so that the ordinate represents the increasing complexity of the VOC simple species or the VOC mixture and the abscissas represents the relative error of the measure (in the example plots, the ordinate is for runs with single compounds as labeled on each cluster of runs and the abscissas is the relative ozone error);

iii. mark lines across the ordinate at the abscissas for the estimated experimental uncertainty of the summary measure for the selected data quality level in the particular chamber (in the example plots these are at $\pm 20\%$ relative error but are not based on the chamber operator's estimates of data uncertainty);

iv. plot all relative errors for each experimental class at one value on the ordinate (these are the vertical clusters on the example plots).

This results in a type of "distribution" diagram for all the experiments of the same experimental class and all experimental classes in one chamber are lined up beside each other. A similar diagram for each chamber allows for a meaningful comparison of error results in each chamber and as a function of organic reactant complexity. It is easy to see potential compensating errors between different VOC classes on such a diagram.

These graphs allow the evaluator to identify all large deviation experiments. These poorest fits must be individually explained. We highly recommend that these experiments and simulations be graphed, that the experimental data streams and documentation be examined, and that the simulation inputs be carefully inspected. Hypothesis about explanations having to do with simulation inputs must be tested by conducting sensitivity analysis and presenting the results. There will be more on this process in the next chapter.

These graphs also quickly show the extent and coverage of the chamber data bases in the different chambers and show useful overlap in conditions between different chambers that allows additional comparisons to be done.

Another type of statistical analysis that should be done in the case where simulation inputs were adjusted to improve agreement between the predictions and observations is to investigate if the values chosen for the inputs are correlated with measures of reactant concentration or run conditions (e.g., higher $[HONO]_o$ with higher initial NO_x, or lower wall

† When producing this graph, the evaluator should replace a measure of fit yielding relative errors that appear large because the observed value is very small with a more reasonable measure of fit. For example, the relative error based on $\Delta[NO]$ should be substituted for the relative error based on $\Delta[O_3]$ when $\Delta[O_3]$ is less than $\simeq$ 40 ppb. This has not been done for the example plots shown here because the needed $\Delta[NO]$ data were not available at the time the plots were made.

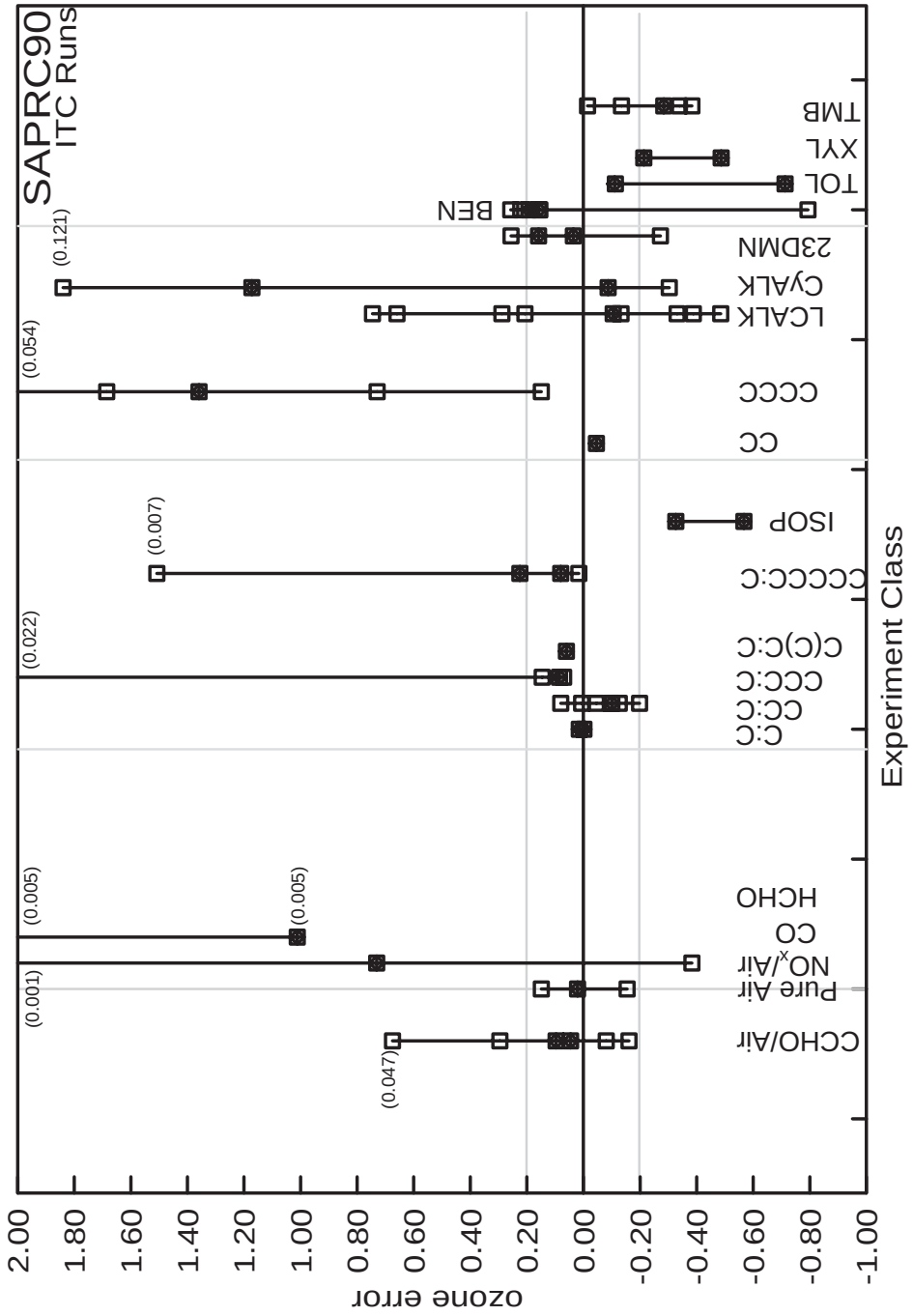

Figure 6. Ozone errors for simulation of single VOC compounds with the SAPRC90 mechanism in the UCR Indoor Teflon Chamber. Filled boxes are the median simulation(s). The numbers in ( ) are the observed O_3 maximum.

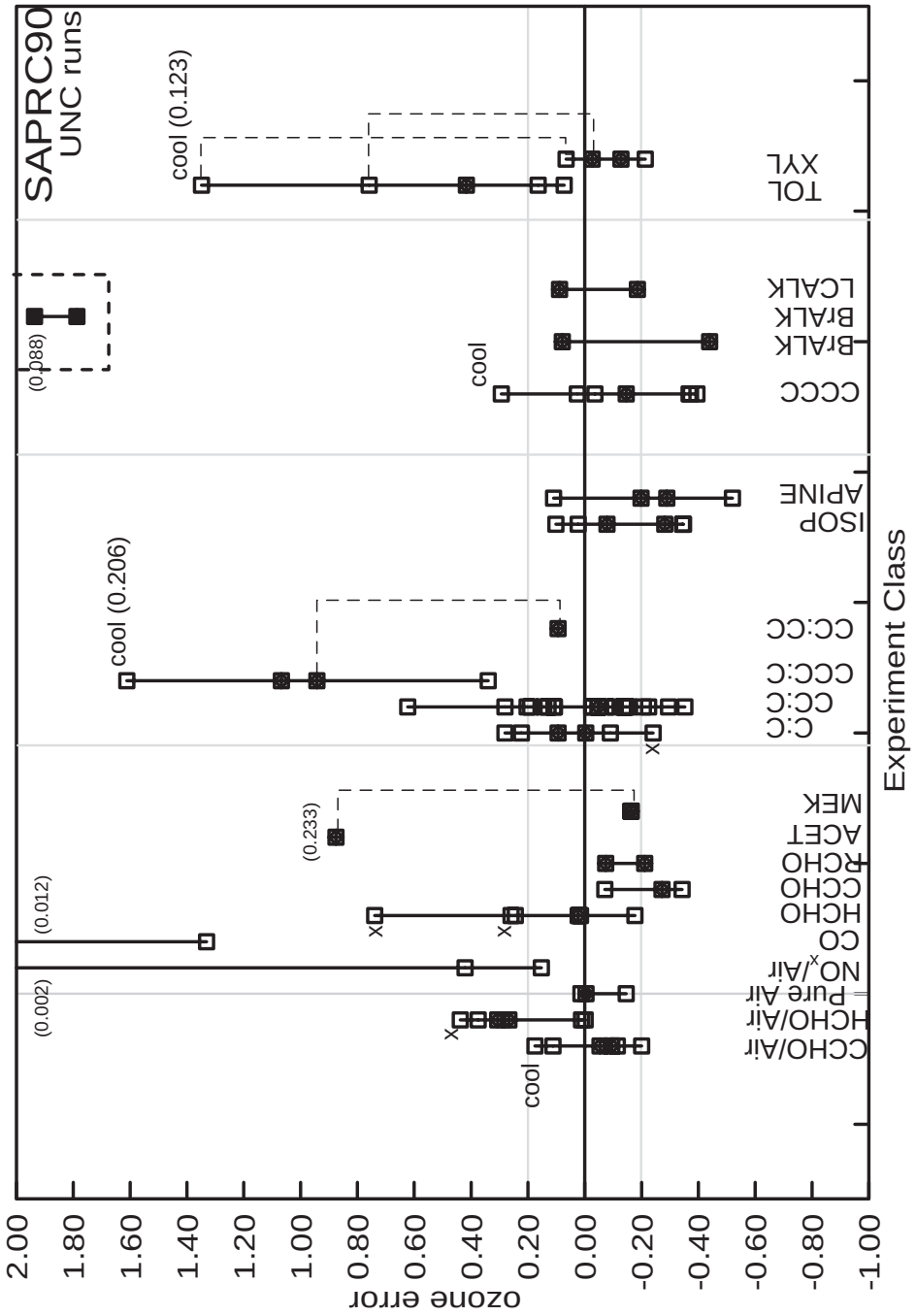

Figure 6. Ozone errors for simulation of single VOC compounds with the SAPRC90 mechanism in the UNC Dual Outdoor Chamber. Filled boxes are the median simulation(s); the numbers in ( ) are the observed O_3 maximum; x mark experiments no longer included in the data base because of low competency. Dashed lines connect side-by-side, same day experiments.

radical rate with lower chamber temperature). Results of sensitivity analyses of the predicted summary measures to the adjusted inputs should be used to translate the variability of the selected inputs to a variability in simulation predictions. This will demonstrate the extent to which the observed data are capable of detecting the combined mechanism's predictive capabilities, that is, the extent to which the demonstrated goodness-of-fit is dependent upon unmeasured inputs.

Make Observation and Prediction Profile Plots

Some simulation systems can automatically produce graphs that plot the chamber concentration-time profiles and the matching simulation output profiles for each experiment simulated. While this may be a nearly overwhelming amount of visual information, these are the only means of non-judgmentally conveying 'fit' to another person. It is highly recommended that each experiment and its simulation be graphed and a hard copy made. These will be very helpful in the evaluation exercise described in the next chapter.

Some experiments in each chamber are so critical to demonstrating the correct performance of the principal mechanism combined with the chamber auxiliary mechanism that the fits for these experiments must be demonstrated by plots of predicted and observed concentrations-time profiles. If they exist in the chamber data set, these experiments should always be plotted and reported:

i) formaldehyde
ii) acetaldehyde
iii) carbon monoxide and NO_X
iv) formaldehyde and NO_X
v) methane and NO_X
vi) acetaldehyde and NO_X

Successful simulations of these experiments would demonstrate that the core reaction set is functioning correctly and that the chamber conditions are reasonably represented. If there are more than three experiments in a class, only the middle, highest and lowest error cases should be plotted.

There are requirements on the presentation style for these graphs, and in particular, specifications as to what combination of species must be plotted together on these graphs. The species NO, NO_2 and O_3 should all appear on a *single* set of x and y axis's, and all three should be presented using the same x and y scales. If available, PAN should be plotted on the same scale as NO_2 on the same graph. This presentation style aids in judging the timing of the NO_X transformation process and its relationship to the production of O_3. This timing gives many clues as to whether the right numbers of radicals are being produced at the right

time. A second graph should be used to show the loss of the principal VOC species and the appearance of the oxygenated organic products. If it is available, the inferred integrated $^{\cdot}$OH can be plotted on the VOC graph.

6

Reporting The Evaluation

Looking back to Figure 3, we see that the evaluator now has all of the raw evidence he needs to defend the claims

> *"It is reasonable to act as if the chamber-dependent auxiliary reaction mechanisms are well formulated."*

and

> *"It is reasonable to act as if the principal reaction mechanism is well formulated."*

and

> *"It is reasonable to act as if the principal reaction mechanism has predictive capability."*

If these three are supported then the evaluator can conclude

> *"It is reasonable to act as if the principal reaction mechanism describes, with a stated degree of uncertainty, the chemical transformations of relevant species on urban and regional scales."*

which is the evaluator's overall purpose.

If the first of the three claims above is shown to be true, then it establishes a necessary condition of being able to describe the effect the chamber itself will have on the predictions made by the principal mechanism. In Chapter 3, we required the evaluator to report on the operational aspects and the theoretical formulation of the auxiliary mechanisms he selected for use, but we have not yet discussed the evaluator's assessment of their accuracy and this must be done in this chapter.†

† Most likely the evaluator would have performed assessments of the auxiliary mechanisms *before* he carried out all the simulations to test the principal mechanism. But for the purposes of describing the process in a logical and straightforward way here, we have not shown this doubling back. That is, in an actual case, the evaluator would have carried the chamber characterization experiments and simulations all the way through the process and would have decided their accuracy and suitability before performing the analysis with all the rest of the chamber data. Here we describe it as if it were all done in one pass.

Organizing the Evaluation

Building the Case To Accept A Mechanism

In Figure 3, we showed that to conclude that the principal mechanism is "well formulated," the evaluator must establish that it has the necessary characteristics of consistency with theory, simplicity in formulation, and accuracy in formulation. The first two characteristics were examined in Chapter 3. Before any simulations would have been performed, the evaluator would have had to conclude that the mechanism had these first two characteristics (or else it was decided to act as if it did whether or not the evaluator could establish this point). This judgment that the mechanism was consistent with accepted theory allowed the evaluator to continue with investigating the other characteristics of the principal mechanism.

In a compelling evaluation, the evaluator would demonstrate that the auxiliary mechanism is soundly based on kinetics measurements and when coupled with a core reaction set and used to simulate chamber characterization experiments, it is always accurate.

In a compelling evaluation, the evaluator would demonstrate that the principal mechanism has the characteristic of being accurate in formulation, by showing the extent of agreement between predictions and corresponding observations for each explicit and model species in the mechanism. In these tests, the predictions would be from simulations using the principal mechanism combined with appropriate auxiliary mechanisms and the observations would be from chamber experiments conducted with single organic species typical of urban air and from series of experiments conducted with simple mixtures of organic species. In the series, the number of species included in the mixture would increase over the series.

The other two characteristics that acceptable principal mechanisms are expected to have are scope and fertility. Recall that these express the ability of the mechanism's predictions to extend beyond the particular observations it was initially designed to explain and the mechanism is expected to predict novel phenomena that were not part of the original set to be explained.

In a compelling evaluation, the evaluator would demonstrate these characteristics by showing that simulations using the principal mechanism combined with appropriate auxiliary mechanisms would accurately predict the relevant species in complex VOC-mixture chamber experiments. In the best demonstration tests, the experiments would be both typical of urban-like conditions and cover the range of expected variations of conditions in the application domain. Experiments designed to stress the mechanism would be included, for example, side-by-side large variations in VOC composition, or side-by-side large differences in total VOC that must be simulated with the same set of other inputs. Many of these experiments would not have been used to formulate or to adjust principal mechanism reactions

and parameters and so accurate predictions for these experiments would contribute to the belief that the principal mechanism had predictive capability.

The difference between the accuracy tests and the predictive capability tests is that, when the predictions failed to agree with the observations, in the accuracy tests it is essentially one part of the mechanism that failed, whereas in the predictive capability tests all the parts of the mechanism are being exercised at the same time and a failure diagnosis would be much harder and probably impossible to make. Thus, the accuracy tests are conducted in an environment that allows the failure to be pinpointed and subsequently explained, but in predictive accuracy tests, the evaluator only knows that the mechanism performed poorly, but can not readily explain why.

Difficulties in Making the Case

Establishing that the principal mechanism has the characteristic of accuracy in formulation and has predictive capability are somewhat difficult undertakings for three reasons.

The first reason is that chamber data are incomplete in their coverage of single species and simple mixtures. Furthermore, the experiments we do have are frequently missing important product measurements or these product data are often too unreliable to be helpful in demonstrating predictive accuracy for these products.† The number of the urban mixture experiments is larger than the single compound experiments, but these also are frequently lacking reliable product data. The range of even urban conditions is not covered well in the existing data. Thus, there is the possibility that formulation accuracy for some species mechanisms in the principal mechanism cannot be tested at all. Other species might have only one or two experiments. Thus, there probably is insufficient experimental evidence to support a conclusion that certain parts of the principal mechanism are well-formulated because the accuracy of their predictions cannot be assessed.

The second difficulty in supporting the claim that the principal mechanism is accurate is that to simulate chamber data the principal mechanism must be combined with an auxiliary mechanism. Because the theory to support the auxiliary mechanism is much weaker than that for the principal mechanism, and because there are too few chamber characterization experiments, the evaluator may have found it necessary to use an auxiliary mechanism that was more "tuned" than it was formulated. Thus, there is the real possibility that the demonstrated accuracy of the combined mechanism is caused by compensation between inaccurate components in both the auxiliary mechanism and the principal mechanism. Since the principal mechanism will be separated from the auxiliary mechanism in its ambient

† Such chamber experiments might still be useful in evaluating the mechanism for O_3 or NO_x, however.

applications, any inaccurate components it has can directly effect the principal mechanism's ambient air predictions.

The third reason it is difficult to establish that the principal mechanism is accurate is that most chamber data are incomplete in the necessary inputs to simulate accurately an experiment and these inputs must be estimated by the evaluator. If there is a bias in the choice of unmeasured but necessary simulation inputs, then again there is the real possibility that the demonstrated accuracy of the combined mechanism in simulating chamber observations is caused by compensation between the inaccurate inputs and inaccurate components in the principal mechanism, the auxiliary mechanism, or both.

Limitations to any evaluation

Regarding completeness of testing, the evaluator will have to vindicate the acceptance of the whole mechanism as having been evaluated the best we currently can do, given the available data. Essentially, the evaluator would claim that no one can do any better.

For those species mechanisms within the principal mechanism for which there is chamber data to evaluate them, establishing the agreement of the mechanism's predictions with observations is a necessary step in the evaluation. It is also necessary to believe that when the combined mechanism does exhibit accuracy in simulating chamber data, the simulation is also free of compensating errors between the principal mechanism and the auxiliary mechanism or between the principal mechanism and the simulation inputs. Until a complete kinetic explanation is available for the wall-mediated reactions that occur in chambers, and until all influential simulations inputs are reliably measured and reported by chamber operators, however, it will not be possible to make a compelling case that the accurate predictions by the combined mechanism are free of these compensating errors.

Thus, the evaluator must settle for being able to decide based on evidence available that "it is reasonable to act as if there were no compensating errors between the principal mechanism and other parts of the simulation system." This means that he would then act as if the principal mechanism's internal workings were independent of any chamber characteristics.

Because of the incompleteness problem, the evidence the evaluator will be able to produce to support such claims, however, may not allow even a presumptive standard to be used, that is, one in which the preponderance of evidence would suggest a "most likely" interpretation, but for which "less probable" hypothesis would still merit consideration. The next lowest standard would be a permissive standard, one that would permit or not rule out the claim, but would also support other claims as equally likely. The last category of evidence in these matters is missing evidence—evidence that is just not there, but would be

very helpful if it was. The identification of such evidence calls attention to the extent of the shortfall in making a compelling case.

Strategies for Treating the Difficulties

The compensating error problem

Among the authors there are two different opinions about how to treat these problems.

Carter believes that:

- an auxiliary mechanism should use the minimum number of empirical parameters required to be consistent with the available data;
- empirical parameter values in an auxiliary mechanism should be set by adjusting them in simulations of specially identified chamber "characterization" experiments (these are experiments designed to detect and display the effects of various wall-mediated processes in chambers); these simulations use the principal mechanism coupled with the auxiliary mechanism being calibrated;
- no chamber experiment where any auxiliary mechanism parameter was adjusted to improve fits can be used to make any conclusions about the principal mechanism being evaluated.

Carter believes that if the evaluator does not have a method to predict a mechanism parameter *a-priori* for each chamber experiment, then if he adjusts it after seeing its effect on the simulation results, there is a possibility that poor performance of a flawed mechanism could be hidden in the variation of an unpredictable parameter, i.e., that there might be a compensating error between the variable input and the internal workings of the principal mechanism. Therefore, Carter believes that he can treat the compensating error problem by using only one value of each parameter in all experiments in one chamber. That is, he only adjusts the parameter globally for a set of all experiments in a particular chamber and the single value to use is determined by simulating all experiments of all types in a single chamber, adjusting the parameter and again simulating all experiments until he is satisfied he has selected a reasonable value. Carter is willing to accept that any particular experiment may not be simulated accurately, but believes that the possibility for compensating error between the auxiliary mechanism and the principal mechanism has been minimized.

Similarly, Carter's strategy for eliminating any potential bias in selecting the necessary but unmeasured simulation inputs is also to select them only on a global basis, that is, only one value for an unmeasured input for the whole data set or one relationship between a measured input and the unmeasured input† would be allowed in the simulation input files.

† For example, the 7% rule used at UCR for $[HONO]_o$ in the EC, in which the input is always estimated to be 7% of the initial $[NO_X]$.

Jeffries has observed that if the simulation predictions are sensitive to the difference between what was the actual chamber condition and the single value chosen as the simulation input for all experiments, the use of the Carter method results in almost all simulation predictions exhibiting an inaccuracy that is not indicative of the performance of the principal mechanism, but instead could be caused by both an incorrect choice of simulation input and any inherent mechanism inaccuracy.

Carter argues in return that separating these two sources of inaccuracy by using sensitivity analyses for the these inputs is quite difficult because of the danger of introducing compensating error except in the case where the agreement of predictions and observations are quite poor and they cannot be improved by any reasonable adjustment of the unmeasured simulation inputs. This would suggest that either the principal or auxiliary mechanism is the source of the inaccuracy and not the inputs. Jeffries responds by pointing out that without adjusting the inputs to see if the simulation predictions can be accurate, the evaluator will never know which process is the cause of the inaccuracy: a variable input that was potentially poorly represented to the simulation solver, or inaccuracies inherent in the principal mechanism.

Under Carter's approach, it is easy to dismiss the mechanism's failures as caused by unpredictable variations in sensitive inputs that were not well represented in the simulations.⋆ Such action carries a hidden assumption: that the simulation predictions *could be made accurate* if the inputs were adjusted. This possibility is never established under this approach.

Because the adjustment is done once for all experiments in a particular chamber, this strategy emphasizes a minimization of errors over *all* experiments, but gives up the possibility of making very meaningful statements about the accuracy of any one simulation. The power of individual experiments to discriminate among principal mechanisms is totally removed under this strategy—only the whole set is meaningful. This is therefore *a very conservative strategy* but also one that will not be able to make strong claims about the principal mechanism's performance.

Jeffries and Gery believe that:

- an auxiliary mechanism should be formulated the same way that principal mechanisms are formulated and they should be as process-oriented as possible given the kinetics literature;
- characterization experiments should be used to evaluate the formulation of an auxiliary mechanism and to determine the range and variability of needed input parameters to the auxiliary mechanism (e.g., extent of nitrate-loading on chamber walls from previous experiments);

⋆ See for example Carter and Lurmann, *op. cit.*, p 2793.

- it is reasonable to adjust the necessary auxiliary mechanism inputs for each experiment simulated in a chamber to account accurately for the particular chamber history and character.
- and a similar approach applies to the likely present but un-measured species concentrations that may have an impact on the accuracy of the simulations.

Jeffries and Gery believe that where it is known that simulation predictions are sensitive to input values and these were not measured, but estimated and are known to vary with chamber condition, it is reasonable to make limited adjustments in these simulations inputs to improve the accuracy of the simulation predictions in an experiment. While the inputs may be unpredictable, they are not unbounded.

To Jeffries and Gery, it is important to know that the simulation's predictions can be made to agree with the observations within the uncertainty of the simulation inputs because this at least permits the possibility that the principal mechanism is accurate.

To guard against the compensation of errors arising from a flaw in the principal mechanism and the choice of these varied un-measured inputs, an analysis of the distribution of inputs and their correlations with any other variables must be done and the results must be reported. It is also necessary to explain in mechanistic terms how these values influence the mechanism's predictions. This allows unusual cases of simulation predictions to be recognized and to be investigated further for the possibilities of compensating error.

Importantly, however, the evaluation will be performed in at least four chambers, each with a somewhat different auxiliary mechanism and perhaps different ranges of un-measured inputs necessary for accurate simulations. That the principal mechanism could produce accurate simulations in all these circumstances would be a good reason to believe that it is free of compensating errors.

This is a case where we can make no clear recommendation because the views are essentially in conflict. Our advice is for the evaluator to choose the method he will use to treat these two problems of necessary but un-measured inputs and of the basic formulation of the auxiliary mechanism and to make the best case he can with it.

It is essential, however, that the evaluator conclude whether his simulations suffer from compensating errors between the simulation inputs or the auxiliary mechanism and the principal mechanism. Further the conclusion must be supported with evidence and a fully developed argument that must be presented.

Lack of Data Problem

If we assume that there are no compensating errors, we can more easily address the lack of data problem.

The characteristic accuracy in formulation for a principal mechanism must be examined by comparing the mechanism's predictions with chamber observations for four kinds of experimental conditions:

i) species represented in the principal mechanism explicitly; simulation predictions for the organic species can be directly compared with chamber observations;

ii) single species that are to be represented by a generalized model-species in the principal mechanism; when the simulation inputs include as the model-species concentration only the single species transformed by the mechanism VOC transformation rules, the simulation predictions for the transformed VOC species can be directly compared with the chamber observations;

iii) simple mixtures of species which are to be represented by a single generalized model-species in the principal mechanism; when the simulation inputs include only one model-species concentration which represents all the single species in the simple mixture transformed by the mechanism VOC transformation rules, the simulation predictions of VOC-species still can be directly compared with the chamber observations; and

iv) simple mixtures of species that have species that are represented by more than one model-species class in the simulation; when used after tests of each of the organic species have been completed satisfactorily, these multi-class mixture experiments test the mechanism's representation of trans-class communication through common species between the classes.

To vindicate the claim that the principal mechanism has been evaluated the best we currently can do, given the available data, the evaluator should proceed as follows:

i) For species mechanisms within the principal mechanism for which no chamber experiments exist, the evaluator should report that it is reasonable to accept the species mechanism as being accurate only because its theoretical formulation was judged to be acceptable by the guidelines in Chapter 3 and other *evidence* which would permit a different interpretation *is missing.* These species should be assigned a very large uncertainty that would in part depend upon the uncertainty in the consensus kinetics data for the species.

ii) For species mechanisms within the principal mechanism for which there are only a small number of applicable chamber experiments (1–3 experiments in any chambers),

   a) the evaluator should report that it is reasonable to accept the species mechanism as being accurate because its theoretical formulation was judged to be acceptable by the guidelines in Chapter 3 and the extent of agreement of the simulation predictions with the chamber observations *permits the possibility* that the mechanism is accurate. Or,

   b) if the extent of agreement with the chamber observations is sufficiently poor, the evaluator should report that it is reasonable to conclude that, even though its

theoretical formulation was judged to be acceptable by the guidelines in Chapter 3, the extent of disagreement of the simulation predictions with the chamber observations *permits the possibility* that the mechanism is inaccurate (this would likely require large errors in the few simulations available).

In either case, the species should be assigned a large uncertainty.

iii) For species mechanisms for which a large number of applicable chamber experiments exists,

a) the evaluator should report that it is reasonable to accept the species mechanism as being accurate because its theoretical formulation was judged to be acceptable by the guidelines in Chapter 3 and the extent of agreement of the simulation predictions with the chamber observations was within the repeatability of the observations in most (or many) of the test cases, therefore the preponderance of the evidence supports the mechanism being accurate. Or,

b) if the extent of agreement is poor in many of the tests, the evaluator should report that it is reasonable to conclude that, even though its theoretical formulation was judged to be acceptable by the guidelines in Chapter 3, the extent of disagreement of the simulation predictions with the chamber observations in many cases was sufficiently larger than the observational repeatability and therefore the preponderance of the evidence supports the mechanism being inaccurate.

In the case of agreement, the uncertainty for the mechanism should dependent upon the spread of the agreement.

These suggestions would be applicable to all four kinds of experimental conditions described just above.

To report on the accuracy of formulation for the principal mechanism, the evaluator must create a table that lists the explicit and model species in the principal mechanism in rows down the side of the table. The entries should include explicit species, single species, pairs and triples of species, and simple mixtures of species. Next there should be a column in which the evaluator enters the level of acceptance of the theoretical formulation of the species mechanism and a second column to indicate the level of uncertainty associated with this judgment. There should also be entries across the top of the table for each chamber. Within a chamber's set of columns should be columns for:†

i. the number of applicable experiments that exist in the chamber

ii. a column labeled "evidence missing" which would be checked if there is no evidence about the accuracy of the species or species mixture

† An evaluator may modify the table contents to suit his needs as long as he reports his decision about each row entry in the table.

iii. a column labeled “number of accurate permissive evidence experiments”

iv. a column labeled “number of inaccurate permissive evidence experiments”

v. a column labeled “number of accurate preponderant evidence experiments”

vi. a column labeled “number of inaccurate preponderant evidence experiments”

vii. a column labeled “report of accuracy” (for ‘yes/no’ or 1-10, or some other way)

viii. a column labeled “species mechanism uncertainty”

ix. a column labeled “reason for conclusions” (contains a number reference to an associated list of reasons that could be similar to the roman numeral list above).

The evaluator should use the simulation results produced in Chapter 5, the summary graphs and the time profile graphs for each experiment that can be used to assess accuracy and fill out the table described here. When the table is finished, the evaluator should make a summary conclusion about the overall accuracy of formulation for the principal mechanism and should provide a discussion of the uncertainty in his conclusions.

The characteristics of scope and fertility in prediction for a principal mechanism must be examined by comparing the mechanism’s predictions with chamber observations from experiments with all the model species classes represent, but at different total concentrations and compositions. Thus, typical experiments include:

i) propene/butane/toluene;

ii) paraffin mixtures/olefin mixtures/aromatic mixtures of various complexities;

iii) synthetic urban VOC mixtures;

iv) synthetic automobile emissions mixtures;

v) for dual chambers, side-to-side experiments with different compositions on each side; with different total carbon on each side; and with variations in both composition and concentration.

To vindicate the claim that the principal mechanism has been evaluated the best we currently can do, given the available data, the evaluator should expand the table created above to include these classes of experiments, and to make similar judgments as was done above.

When the table is finished, the evaluator should make a summary conclusion about the overall characteristics of scope and fertility for the principal mechanism and should provide a discussion of the uncertainty in his conclusions.

Presenting the Case to Accept a Mechanism

The evaluator has now converted the raw evidence into organized and more easily digestible forms supporting the three claims given at the beginning of this chapter. The evaluator now needs to make one final consolidation of his processed evidence and support or refute the claim

> *"It is reasonable to act as if the principal reaction mechanism describes, with a stated degree of uncertainty, the chemical transformations of relevant species on urban and regional scales."*

In supporting this claim the evaluator should briefly review his evidence, restate the warrants, supply backing evidence, qualify the strength of the conclusions, review any rebuttals, and conclude.

7

Conclusion

Looking back to Figure 3,

- we have now completely described all the elements the evaluator must consider in performing an evaluation,
- we have made clear those issues that are relevant,
- we have been explicit about what the evaluator must report as the basis of the decisions he makes.

From here, the evaluator and the sponsor must produce a specific set of agreements—a protocol—for a particular evaluation of a particular mechanism using a particular set of chamber data. As an aid in this process, the Appendix contains a summary of all the necessary reports. The two parties need to agree on how these reports will be produced.

The evaluator will need to assemble and report on a set of credible chamber data relevant to his protocol (i.e., chamber data capable of causing the evaluator to change his mind about the validity of the claims he wishes to make). Three sets of chamber data that meet the criteria given in Chapter 4 will be available by the end of 1992. Two of these, the UNC Dual Outdoor Smog Chamber data set and the UCR Indoor Teflon Chamber and Indoor Evacuable Chamber data sets were produced as part of this total study and will be described in future reports from UNC and UCR. The third chamber data set, the Dual Outdoor Smog Chamber data set produced by the Commonwealth Scientific and Industrial Research Organization (CSIRO) in Sydney, Australia is being documented and quality assured now and should be available by the end of 1993. It will be described in a future report from UNC and CSIRO. Another future report from UNC will describe necessary documentation and formats for any additional chamber data that will be added to the "standard data base."

In Chapter 6, we explained that because of the data incompleteness problem, it is presently **not possible** to make a *compelling* case to accept a chemical reaction mechanism as accurately describing urban chemical transformations. In fact, the chamber evidence

may not allow even a presumptive standard to be used. For many species of interest, there is only missing evidence.† Therefore, present evaluations must be satisfied with vindicating the use of a mechanism, that is, to claim that the mechanism has been evaluated the best we can currently do, given the available data.

Clearly, the "standard chamber data base" for use in evaluating chemical reaction mechanisms needs expansion, both by the creation of new experiments and by the documentation and addition of existing experiments from a number of facilities. The former requires that we wait two years for each small additional set of experiments. The latter, however, requires significant efforts from a researcher who was involved in the data production and who has the time to adequately document and support the credibility of the chamber data so it can be used by evaluators. Both require continued support.

† This is evidence that would be persuasive if it were available, but it does not exist. The identification of such evidence calls attention to the extent of the shortfall in making a compelling case.

A

Reports Summary

I. Report on The Purpose and Scope of Evaluation

1. Select principal reaction mechanism.

2. Select candidate chamber-wall reaction mechanisms.

3. Decide on relevant species.

4. Identify the potentially relevant chamber data bases.

II. Report On the Mechanisms

1. Produce an operational description for the principal mechanism that includes:
    a) a list of species in the mechanism including their names, symbols, types (explicit or model), and that describes what the species represent in the real world;
    b) a list of reactions where each reaction includes:
        i) reactants, products, and product stoichiometry;
        ii) for thermal reactions, the specific reaction rate in molecules, cm^3, and seconds units in as complex a form as needed to express the value as a function of temperature, pressure, and other third-body interactions;
        iii) for photolytic reactions, a set of absorption cross-sections and quantum yields in as complex a form as needed to express the values as a function of wavelength, and possibly, temperature and pressure;

2. Produce instructions for using the mechanism, such as:

a) how to evaluate the reaction rate constants as given;

b) how to evaluate any other parameters which appear in the mechanism; and

c) how to convert real-world species into the mechanism's explicit and model species including tables for ambient and inventory compounds;

3. Provide other technical documentation including:
   a) general formulae used in the kinetic and photolytic rate calculations;
   b) any special methodology for encoding mechanism reaction properties that do not fit into the general procedures; and
   c) for variable stoichiometry, complete instructions, formulae, or tables needed to determine product yields;
   d) the date on which the mechanism's kinetics data and structure were "frozen".

4. Describe results of mechanism implementation tests.

5. Describe a quality assurance of the mechanism's formulation including:
   a) judgments about the development protocol;
   b) knowledge of and judgments about the chamber data used in the formulation, e.g., a list of all chamber data that were used, the revision dates of these data, a listing of the chamber-dependent auxiliary mechanisms used, and copies of other chamber operations models used by the developer;
   c) judgments about the formulation and condensation rationale employed by the developer;
   d) where the evaluator feels there may be problems in the formulation, he should report detail data, simulations, and plots as needed;
   e) judgments about the evidence offered by the developer as to why the principal mechanism should be accepted as being independent of chambers used in its formulation;
   f) judgments about significant assumptions, limitations, and uncertainties;
   g) if it is concluded that the principal mechanism has significant flaws in its formulation, the evaluator must report how he will proceed, e.g., ignore it, fix it, or stop the process and document the case;

6. Describe the operation and formulation of chamber-dependent auxiliary mechanisms for each chamber;
   a) origins of auxiliary mechanisms;

b) list of all suspected processes and species in each chamber;

c) description of formulation process;

d) complete listing of chamber mechanism(s) adopted including reactions, rates, and other parameters;

e) instructions for estimating chamber-specific values for parameters used in the chamber-dependent mechanisms.

III. Report On Chamber Data

1. List chamber data needed to conduct a compelling evaluation of a particular principal mechanism;

2. List existing chamber data that relate to the listed needs above; for each experiment listed, the evaluator must:
   a) review and comment on each to establish that existing data are of usable quality and meet the standards the evaluator needs for the purposes of his evaluation;
   b) provide experiment-specific warrants that explain the relevance of each experiment to the claim it is being used to support;
   c) describe the evidence assembled to support claims that the selected chamber data are credible; this must be done for each data stream to be used for comparison with simulation predictions;
   d) assign each data stream a quality category according to which experiments have better established species concentrations than others and therefore should be given more weight;
   e) assign each experiment a total competency to challenge simulation predictions and provide evidence to support the assignments;

3. Describe the difference between the data needed to conduct a compelling evaluation and the data available and present a plan for how to advance the study given any shortfall in data;

4. List summary statistics for each data stream for each experiment selected.

IV. Report On Simulations

Simulation Inputs

1. List the experiments in each chamber that will be simulated and include:

a) the general class of experiment, i.e., characterization, single species, simple mixture, complex mixture, in sufficient detail for later use in plotting groups of experiments;

b) the initial total concentrations of VOCs and NO_X;

c) the initial and average or maximum chamber conditions, e.g., starting and maximum temperature;

d) the judged competency of the whole experiment to challenge the simulation predictions;

e) whether the experiment was used by the developer to set parameters in the principal mechanism;

f) for each relevant species in the evaluation, an indication of whether a data stream is included in each experiment file, its quality category and its summary statistics;

2. Provide descriptions of:

a) algorithms for producing simulation file concentration inputs when:

i) species is measured and directly reported in the data file,

ii) species was not measured, but was estimated by chamber operators,

iii) species was not measured and is estimated by evaluator,

b) algorithms and data used to produce in-chamber actinic flux;

c) algorithms used to compute species photolysis rates;

d) algorithms used to specify in-chamber temperature, dilution, and water vapor;

3. Include detailed comments in each simulation input file documenting the source of all the specified values. All data files should be suitable and available for distribution.

Simulation Outputs

1. Report that for each simulation the predictions of the relevant species have been plotted against the corresponding observations and the resulting figures have been examined by the evaluator; hard copy is not required for all experiments, but if they exist in the chamber data sets, the following types of experiments should be plotted and reported:

a) NO_X

b) formaldehyde

c) acetaldehyde

d) carbon monoxide and NO_X

e) formaldehyde and NO_X
f) methane and NO_X
g) acetaldehyde and NO_X

2. List the corresponding summary measures for the simulation results;

3. List the calculated absolute and relative prediction errors;

4. Display plots of relative prediction errors in which:
   a) the errors are presented on a chamber by chamber basis;
   b) within each chamber, the errors are presented by progressing from simple to complex individual species and from simple to complex mixtures;
   c) within each species or experimental class, the errors are sorted from smallest to largest;

5. Present plots of all large deviation experiments and offer hypothesis to explain each one.

V. Report On Evaluation

1. Present a complete logical argument (including citing data, i.e., evidence, presenting warrants, and producing backing) that concludes whether or not the simulations suffer from compensating errors between the simulation inputs or the auxiliary mechanism and the principal mechanism (see the text for guidance about issues that must be addressed);

2. Assuming that lack of compensating errors is demonstrated, assess the characteristic of accuracy in formulation and provide vindication of claims by producing a list of explicit and model species in the principal mechanism which has the following columns for each chamber in which the species were tested:
   a) the number of applicable experiments that exist in the chamber
   b) a column labeled “evidence missing” which would be checked if there is no evidence about the accuracy of the species or species mixture
   c) a column labeled “number of accurate permissive evidence experiments”
   d) a column labeled “number of inaccurate permissive evidence experiments”
   e) a column labeled “number of accurate preponderant evidence experiments”
   f) a column labeled “number of inaccurate preponderant evidence experiments”

g) a column labeled “report of accuracy” (for ‘yes/no’ or 1-10, or some other way)

h) a column labeled “species mechanism uncertainty”

i) a column labeled “reason for conclusions” (contains a number reference to an associated list of reasons that could be similar to the lists given in Chapter 6).

3. Report and justify a summary conclusion about the overall accuracy of formulation for the principal mechanism and provide discussion of uncertainty in conclusions;

4. In a similar manner, report and justify a summary conclusion about the characteristics of scope and fertility in prediction for a principal mechanism using simple and complex mixture experimental classes and provide discussion of uncertainty in conclusions;

5. Support or refute the final claim that the mechanism does describe, with a stated degree of uncertainty, the chemical transformations of relevant species on urban and regional scales; review the evidence, cite the warrants, supply any needed backing evidence, qualify the strength of the conclusions, review any rebuttals, and make the conclusion.