DRAFT REM GUIDANCE REVIEW

<u>Charge Question 1:</u> *Has EPA sufficiently and appropriately identified the best practices, such that decisions based on models developed and used in accordance with these practices may be said to be based on the best available, practicable science?*

1.0 Interpretation of Best Available and Practicable Science

In developing and applying a model for supporting a regulatory action or decision, it is important to meet the criterion stated in question 1 – "based on the *best available, practicable science*." To the Panel, this means that the model uses the best current science that is consistent with the model's intended use, whether that use is regulatory, management or scientific. The term "practicable" refers to consideration of problem specification and programmatic constraints (data quality and availability, and limitations of time and resources) in selection of model complexity (i.e., spatial, temporal, and process resolution). Thus in the context of Figure 2 of the draft guidance document, the Panel suggests that the location of the minimum (both in the x- and y-directions) in the uncertainty versus model complexity curve will depend on the problem specification and programmatic constraints. The Panel believes that when a model complexity is most appropriate for the problem and available data and resources, it is obtaining the minimum possible uncertainty and, hence, using the *best available, practicable science*. The Panel interprets this question as asking whether the guidance provided aids the modeler in finding that level of model complexity.

1.1 General Comments

In general, the Panel finds the REM initiative provides a common and much needed vision for modeling across all of the offices within the Agency. The draft document in particular provides a comprehensive overview of modeling principles and best practices, in a concise manner. The Panel also finds that the Agency has been responsive to previous SAB advice on modeling practices, and commends the REM participants for their leadership. The Panel looks forward to working together with the Agency to make this an excellent guidance for modeling to improve decision making in the future. In particular the Panel applauds the emphasis in the document on using the peer review process to insure that a Regulatory Environmental Model is using the best available, practicable science. The Panel encourages the document to urge that *any* regulatory modeling project include a peer review plan in its QAPP. Furthermore, the Panel suggests that the peer review plan implement *ongoing* peer review through all stages of the modeling process, not just after the model application. Such a proactive practice will assist in avoiding crucial technical errors or omissions that are difficult or impossible to rectify after the project is over. Also, the Panel favors an open modeling process for Regulatory Environmental Models, in which modeling decisions and results are shared with stakeholders through model development and application. This practice avoids a situation where the model fails to address the regulatory questions as conceived by the various stakeholders in the process.

1.2 Problem Specification

The Panel appreciates the distinctions made in the guidance document between model framework development and model application. Nevertheless, the Panel finds that this

distinction is not consistently maintained throughout the document. For example, the terms "application tool" in section 2 means problem-specific model implementation whereas "model application" in section 4 means model based decision making. **The Panel recommends that the term application tool be replaced with "problem-specific implementation".**

The Panel believes that *Problem Specification* is a critical element of *any* modeling project. It guides the development of the conceptual model and it governs the model complexity. It must, therefore, include a clear and complete statement of policy, management, and/or scientific objectives, model spatial and temporal domain and resolution characteristics, as well as program constraints (e.g., legal, institutional, data, time and economics). This process must involve interactions among all stakeholders. **The Panel recommends that *Problem Specification* be given greater emphasis in the guidance document by elevating it to a separate, initial step in the modeling process. In this context the Panel offers an alternative Figure 1 for the guidance document.** The Panel believes that the alternative figure better reflects the central role of stakeholders in the public policy process and their interaction points in the modeling process. It also represents a better delineation of the modeling process itself and the review and iterative nature of that process.

In accord with this observation the Panel offers the following suggestions that should be included for completeness and clarity in the problem specification portion of the document for each of the above aspects of problem specification:

- **Regulatory or research objectives** are statements of what questions a model has to answer. The statement of modeling objectives should include the state variables of concern, the stressors (model inputs) driving those state variables and their control options, appropriate time and space scales, model user acceptance, and, very importantly, the degree of accuracy and precision of the model. The paragraph on Data Quality Objectives (DQOs) in the document is good, but the relation to desired accuracy and precision of the model is not made clear.
- Under scope of guidance, the Panel suggests an alternative way to describe model types covered by the guidance is to compare and contrast: empirical vs. mechanistic, static vs. dynamic, simulation vs. optimization, deterministic vs. stochastic, lumped vs. distributed.
- Specifying the **model domain characteristics** includes: identification of the environmental domain being modeled, specification of transport and transformation processes within that domain that are relevant to the policy/management/research objectives, specification of important time and space scales inherent in transport and transformation processes within that domain in comparison with the time and space scales of the problem objectives, and any peculiar conditions of the domain that will affect model selection or new model construction.
- Problem specification should include a discussion of the potential **programmatic constraints**. These address time and budget, available data or resources to acquire more data, legal and institutional considerations, computer resource constraints, and experience and expertise of the modeling staff.
- These factors, collectively, allow the modeler to determine the "complexity" of a model that is necessary and sufficient for the application under consideration.

Comment: The draft guidance definition of "complexity" makes it, in my opinion more difficult to quantify/understand than simply saying that model complexity is determined by spatial, temporal, and process resolution.

Deleted: define

Deleted: , and should be distinguished from the definition of "complexity" given in the glossary (Note: look up glossary definition).

1.3 Model Calibration and Sensitivity Analysis

The panel applauds the overall treatment of model quality assurance and evaluation in Appendices B and C of the guidance document. **However, the panel recommends that the process of "model calibration" receive increased attention regarding guiding principles and best practices, both in the main text of the document and in the appendices.** While calibration of air models may not be desirable or important, it is an integral part of water quality modeling and one of the more poorly understood steps in the modeling process (for example, the document could discuss how sensitivity analysis can be used during the calibration process). Most process-oriented environmental models are underdetermined; that is, they contain more uncertain parameters than state variables that can be used to perform a calibration. Therefore, good model calibration practice uses sensitivity analysis to determine key processes for a given problem-specific implementation and then recommends empirical determination of the rate of those key processes as part of the calibration process in addition to measuring the time and space profile of relevant state variables. This practice can help further constrain a model for which parameterization by calibration is difficult. An example of this practice would be to measure the rate of photosynthesis (process) in a lake in addition to the biomass of phytoplankton (state variable).

1.4 Model Post-Audit

The practice of model post-auditing is defined as the ongoing observation of the response of the system to the actual implementation of a policy or management action relative to the model's forecast of how that system would respond, and is crucial to the ongoing improvement of environmental models. **The Panel recommends that the guidance document acknowledge the value of post-auditing of models and associated data collection. This practice deserves a section of its own in the model application area. That section might also discuss the role of regulatory modeling in adaptive management of environmental systems.**

1.5 Document Organization

The Panel believes that there are best practices for the development of a generic model framework (such as, for example, WASP, QUAL2E, and AQUATOX) however most users of the guidance document will *not* be model developers. Therefore, the document should contain additional best practices that should be followed for a site-specific or problem-specific implementation of a model framework. **In order to clarify the guiding principles that should be considered for each type of project, the Panel recommends that the Agency consider organizing the guidance document according to the steps involved in carrying out a modeling project from inception to completion.** The Panel identifies these steps to be: Problem Specification; Existing Model Framework Selection or New Model Framework Construction (the document should recognize that a site-specific modeling project may be conducted by either new model construction or by selection of an existing model framework); Problem- and Site-specific Configuration; Model Calibration and Sensitivity Analysis; Model Code Verification; Model Evaluation through Confirmation/Corroboration, Sensitivity Analysis and Uncertainty Analysis; Model Problem-specific Application (use of models to address specified questions); Model Post-Audit; and Overall Documentation. These activities should be covered in a QAPP for any given modeling project.

Comment: I am not sure if this comment goes in CQ1 or not, but I will make it here and see if the panel agrees and, if so, where they think it belongs. My comment is that I do not believe that the guidance document places enough emphasis on or presents any examples of the value of diagnostic use of models for both evaluation and problem-specific application. For example, I find that plotting the cumulative distribution of observations of a state variable on the same plot as the cumulative distribution of model computation of that state variable on the same spatial and temporal scale is very valuable. I also find that development of a model mass balance diagram of a given state variable over appropriately chosen space and time scales (e.g., whole lake water column over course of a year) to be very valuable for identifying significant mass flow pathways, for addressing management questions, and for helping to guide monitoring programs.

<u>Charge Question 2:</u> *Has EPA sufficiently and appropriately described the goals and methods, and in adequate detail, such that the guidance serves as a practical, relevant, and useful tool for model developers and users? If not, what else would you recommend to achieve these ends?*

2.0 Introduction

The general goals of the document are clearly stated (page 6), i.e., to provide guidance on how to assess the quality of regulatory environmental modeling. The assessment is to be made on the basis of a number of "performance criteria" or "specifications" (page 3) that characterize the three major components of regulatory environmental modeling; namely (1) model development. (2) model evaluation, and (3) model application. The document provides specific (and alternative) methods by which the performance criteria for each of these three components may be assessed.

The Panel agrees that the document is an excellent start to defining the process of and providing the measurement tools for quality assurance in regulatory environmental modeling. Furthermore, the Panel makes particular note of the critical importance of problem specification at the beginning of any modeling project. Problem specification supplies the modeling objectives that tie together the modeling components described in the document (see Charge Question 1).

2.1 Intended Audience and Scope of Use

Upon first reading, the document appears to identify the intended audience as being composed of two general categories: model developers and model users. The three components of regulatory environmental modeling have varying relevance to each of these audiences. The model development component is targeted at model developers, the model evaluation component is relevant to a broad range of modeling constituencies, and the model application component is focused primarily at managers and decision makers

After closer reading however, other important modeling constituencies are identified. For example, three groups are explicitly identified in the "communication" criterion under "model application" (page 3): modelers (i.e. developers), analysts (i.e. users who setup and generate model output), and decision makers (i.e. managers who use model output). It would be useful to elaborate on the distinction between the model users who generate model output (those who setup, parameterize, run, calibrate, etc, particularly with model framework software like WASP or QUAL2E), and those who are managers and are principally users of model output. They are both users, but play different roles in regulatory environmental modeling, and as such are likely to use this guidance to assess different quality criteria. It would also help to clarify the intent of the guidance and its relationship to its different regulatory audiences (at least 2 groups): regulatory decision makers, and regional and state "assessors"/advisors for permit applicants. Panel discussions also suggested including other stakeholders in this audience, e.g., those to whom the results will apply or affect. For less experienced audiences, the document may be insufficiently explanatory. **The panel recommends that the Agency clarify the use of this guidance for the variety of intended audiences and suggests that the Agency specifically describe or suggest how the different constituencies in a modeling project might beneficially use this guidance.**

A general concern about the overall document is its scope of use. The Panel finds that the guidance document provides a valuable resource to modelers in a wide range of disciplines, but unlike typical EPA guidance documents, does not lay out a step-by-step course of action. Instead, it identifies a set of key "best practices" which should be adhered to, along with supporting materials. **Because this document differs in scope and content from other "guidance", and because the term "guidance" has specific connotations in certain areas of model application, the Panel suggests that EPA consider using a term such as "guiding principles" instead of "guidance", both in the body of the document and in the document's title.** A second general issue related to the scope of the document is that much of the introductory parts of the document refer exclusively to regulatory applications of models, yet it is clear that the intent of the CREM process is to bring consistency to all environmental applications of models, (e.g., regulatory support, research, resource assessment, evaluating alternative management actions, economic evaluations, etc.). **Therefore, the Panel recommends that the guidance document, including its stated purpose, be revised to reflect these additional uses.**

2.2 Glossary

One of the keys to a workable guidance document for quality assurance in environmental modeling is that the various modeling constituencies share a common language and definition of key ideas and terms. The Panel believes the Agency has made a commendable effort in attempting to establish a common vocabulary for the purpose of environmental modeling. The glossary is an excellent component of this document for providing the basis of that shared understanding.

However, there is room for improvement and consistency, not only in the glossary, but also in the text. For example, some of the terminology and definitions are subject to multiple interpretations, which is to be expected for a document that combines vocabularies from a variety of fields. The Panel notes that the document's use of certain terms, e.g. "guidance" is at times at variance with past definitions, including some of the Agency's own previous modeling documents many of which are cited in the references. Thus, the Panel recommends that the Agency clarify the document's use of terminology and definitions that may not always agree with past Agency usage.

The current terminology used to describe graded approach needs to be clarified. For example, "managerial controls" should be replaced with a more generic terms such as "level of effort" or "allocation of resources". Another problematic area is the potentially misleading or overly generalized use of common statistical terms such as "reliability" and "sampling errors". Where the Agency's use of terms is intentionally different from prior or accepted use, they should be noted as such, and a brief, appropriate rationale should be provided. Appendix A gives specific examples. [Note: Need these examples].

The panel suggests the Glossary be expanded to include more terms to make it as comprehensive as possible. Some key terms that should be added are: "validation" (add a note: see Model Validation), "documentation", "user manual", "proprietary models", "secondary applications", "flow chart (code), etc. Some panel members questioned whether the glossary definitions are the consensus of those in the Agency, or in the modeling community, or both? For example, "corroboration" is an interesting and appealing substitute for "validation", but one that is not yet widely used in practice. Inclusion of definitions of terms used in the Data

Dictionary of the Models Knowledge Base may also improve the utility and consistency of the document. [See also the text in Charge Question 6].

In summary, the Panel recommends that the Agency

- **clarify the document's use of terminology and definitions that may not always agree with past Agency usage,**
- **revise the definitions of certain statistical terms that may be at odds with accepted practice,**
- **expand the glossary to include more terms to make it as comprehensive as possible,**
- **include in the glossary, terminology used in the Models Knowledge Database - Data Dictionary.**

2.3 Model documentation, project documentation, and user manual

The only model documentation referred to in the guidance is in the model application component, i.e. a comprehensive project documentation to address "transparency" issues. However there is a need for model documentation during development, especially for complex modeling frameworks. In addition no mention is made of the need for an adequate user manual (or user guide) for the "analyst" group of model users. It is unclear if this is assumed to be part of the documentation. Some panel members think it is separate and distinct from model documentation, and is essential. A model user manual should contain example applications of the model (or model framework).

Charge Question 3: *Has EPA sufficiently and appropriately proposed a graded approach, such that users of the guidance can determine the appropriate level of evaluation for a particular model use? If there are deficiencies in the proposed approach, what would you recommend to correct it, and why?*

3.0 Definition of "Graded Approach"

The concept of a "graded approach" is implicit throughout the Draft Guidance document, as it should be. Usually "graded" is expressed implicitly through the use of the descriptor "appropriate." The term "graded approach" does not appear until page 18 under "Model Evaluation." However, the sentence in which the term is introduced applies to all phases of modeling—development, evaluation and application—not just evaluation. The Panel recommends that the sentence on page 18, along with the concept of a graded approach, be introduced earlier in the document before the discussion of model development, as part of overarching concepts that are part of all of the modeling stages. **More explicitly, the Panel recommends that the sentence now on page 18 be modified to read: "Model development, evaluation and use should always be conducted using a graded approach that is adequate and appropriate to the decision at hand as determined by the Problem Specification process described in the panel discussion of Charge Question #1 [6, 7]."(TLT query—what are 6,7?)** This introduction should then be followed by a brief discussion of how "graded" applies throughout the modeling process. For example, in the context of model development, "graded" refers to the extent to which existing models are modified to fit the problem specification or that screening models are used instead of more complex models.

The concept of a graded approach, as discussed in this report and in the guidance document, is not clearly reflected in the definition of graded approach provided on page 31. Graded approach is defined on page 31 as the "process of basing the level of application of managerial controls applied to an item or work according to the intended use of results and degree of confidence needed in the results [7]." The Panel recognizes that formal modification of the EPA's definition of "graded approach" may not be practicable for the REM Guidance document. This term's current definition is pervasively cited throughout numerous high-level Agency guidance documents and seems consistently to limit its scope to the implementation of quality control systems. To the best of our knowledge, the concept of applying a graded approach to implementing technical approaches to activities does not appear to be explicitly discussed in most (if any) Agency guidance. (TLT query—what is the point of this paragraph? Is there a recommendation? Does this feed into something else?)

3.1 Modeling Complexity

The scope of models that can be used for a particular application can range from the simplest models to the very complex. In addition to providing some additional comment on where the grading starts (i.e., what is the simplest model to be considered as a model in the REM guidance document or in the Models Knowledge data base), the guidance document needs to comment in more detail on the level of evaluation or "grade" of evaluation that might be appropriate for different models and their applications. Currently, the discussion on page 18 applying to the application of a graded approach to evaluation is quite brief and is not addressed anywhere else in the guidance document. For instance the example of a "screening test" is discussed in the guidance document as a case where less rigorous model evaluation is required.

Comment: I think this discussion should recognize that the draft guidance document does discuss model complexity in Section 2.3.1 (and Figure 2) of the document. I think that this is a good figure, but perhaps needs some more discussion regarding the factors that determine the location of the "point of minimum uncertainty" in the uncertainty-complexity space depicted in the figure. My philosophy in thinking about choosing model complexity is that this point is determined by **problem specification** and **available data** for model calibration and evaluation. In other words, if a difficult problem is specified it would require a more complex model; and if additional data are available to support that level of complexity, the point of minimum uncertainty would not only move to the right but would move down on the y-axis (lower uncertainty relative to that problem definition). *(I can discuss this more or elaborate if desired by the group.)*

More complex situations should also be addressed in order to clarify the extended scope of evaluation that may be needed in such cases.

The draft guidance document does not alert the reader that external circumstances can affect the rigor required in model evaluation. For example, in cases where the likely result of the modeling will be costly control strategies, court actions, or alienation of some sectors of the population, detailed model evaluation may be necessary. In those cases, all aspects of the modeling will come under close scrutiny, and it is incumbent upon the modeler to probe deeply into the model's inner workings (sometimes called "process analysis") to support subsequent regulatory decisions. This level of deeper model evaluation also would be appropriate when modeling unique or extreme situations not previously encountered.

The draft document should also note that gradation in evaluation can apply within complex model applications. For example, in modeling urban air quality, most areas use a regional modeling domain nested to provide higher resolution over the region of primary interest (e.g., Amar et al., 2004). Clearly the most intensive performance evaluation should be directed towards the object of the modeling (the "fine grid"), but at least some level of evaluation should be applied to more distant areas (the "coarse grid").

The Panel finds that the draft guidance document acknowledges the scope and complexity of the models being used, but recommends that it provide more examples of appropriate evaluation steps for different models and model systems (i.e., combinations of models linked to address a particular issue) and for their particular applications. The Panel recommends that the guidance document broaden the discussion of the graded evaluation approach to discuss evaluation requirements for additional circumstances such as using models in potentially litigious applications or in unfamiliar or unique situations.

Model evaluation in most every situation basically involves expert judgment, examination of model output, sensitivity and uncertainty analysis and comparison with observational data. The guidance document needs to discuss the appropriateness of using the more qualitative evaluation steps such as expert judgment to "screen" the model performance and application appropriateness (i.e., how well does the numerical model agree with the conceptual model) before launching into more formal and complex, or higher grade, intercomparisons with observations or sensitivity analyses. (TLT query: Is this redundant with CQ1?)

Comment: I don't think it is redundant, but perhaps this paragraph should go into CQ1 after the model calibration section and before the model post-audit section. It should also contain some appropriate references such as those of Beck and perhaps Small.

3.2 Use of Multiple and Linked Models

Many environmental problems require use of multiple models, with the models often linking together and interacting to varying degrees. For example, air quality modeling often links meteorological, emissions, and air chemistry/transport models. Integrated assessments that attempt to evaluate multiple, interdependent benefits and costs of a problem such as the overall value of the Clean Air Act as is done in EPA's studies on Section 812 of that act (U.S. EPA, 1997, 1999) and the work of the Grand Canyon Visibility Transport Commission (GCVTC, 1996) require linkage of a wide variety of atmospheric, environmental, economic and social models.

In cases in which multiple models are linked together to address a particularly complex issue (the 3MRA modeling framework is a good example of this), each model needs to be evaluated individually to assure that the model is being used within its proper domain and that it is performing properly in the context of the integrated assessment. In addition, evaluation of the full modeling system needs to take place to make sure that the overall analysis is adequate and appropriate for the application. Just because individual modeling components are behaving

properly does not necessarily mean that the full system will provide authentic overall analyses. When using a system of linked models, it is essential to beware of compensating errors, which can lead to "getting the right answer for the wrong reason." In air quality modeling, for example, it is possible to achieve reasonable ground-level pollutant concentrations even though the modeled emission rates are too low, if the meteorological model generates insufficient atmospheric mixing (e.g., the Houston/Galveston Air Quality Science Evaluation [this citation need more detail] http://www.tnrcc.state.tx.us/air/aqp/airquality_contracts.html)

The Panel recommends that the guidance document acknowledge that many applications require the linkage of multiple models and that this linkage has implications for assessing uncertainty and applying the team of models. Each component model as well as the full system of integrated models needs to be evaluated for a given application.

3.3 Uncertainty (TLT note: Not sure if this is repetitive with what is in CQ4—in red)

The Panel commends the document authors for recognizing that the definition of data includes data sets generated from modeling exercises as well as from the literature and existing databases. However, the guidance also needs to clearly discuss treatment of uncertainty associated with the application of these diverse model-generated data as well as data sets derived directly from observations.

Data derived from modeling analysis that are then used for another modeling application also must be evaluated for uncertainties, caveats, and limitations in applicability. The evaluation then must be carried with the data throughout their future uses.

One example of this need for propagation of data uncertainties and limitations is the use of emission inventories in regional air quality modeling. The emission inventories often are the result of complex data collection, analysis and emissions modeling. The inherent uncertainties in the emissions data and the emissions modeling need to be somehow quantified. Use of the data as input for the next phase of modeling then needs to recognize the uncertainties and their impacts on the next modeling steps.

Sometimes, the uncertainties can be treated explicitly and quantitatively and other times, the uncertainties can only be acknowledged qualitatively. Regardless, the uncertainties need to be noted and considered throughout the modeling system. This complex relationship between data and models needs to be discussed in the guidance document.

3.4 References

Amar, P.; R. Bornstein; H. Feldman; H. Jeffries; D. Steyn; R. Ramartino; and Y. Zhang (2004). Review of CMAQ Model, December 17-18, 2003. Submitted March 1, 2004. www.epa.gov/c**air**/pdfs/PeerReview_of_CMAQ.pdf

Grand Canyon Visibility Transport Commission (GCVTC) (1996) Recommendations for Improving Western Vistas: Report of the Grand Canyon Visibility Transport Commission to the United States Environmental Protection Agency. Dated June 10, 1996. *http://wrapair.org/WRAP/Reports/GCVTCFinal.PDF*

Houston/Galveston Air Quality Science Evaluation [this citation need more detail] http://www.tnrcc.state.tx.us/air/aqp/airquality_contracts.html

U.S. EPA (1997) Final Report to Congress on Benefits and Costs of the Clean Air Act, 1970 to 1990. Report EPA 410-R-97-002. http://www.epa.gov/air/sect812/

U.S. EPA (1999) Final Report to Congress on Benefits and Costs of the Clean Air Act, 1990 to 2010. Report EPA 410-R-99-001. http://www.epa.gov/air/sect812/

Charge Question 4: *Has EPA sufficiently and appropriately provided practicable advice for decision-makers who must deal with the uncertainty inherent in environmental models and their application? What additional advice should EPA consider in dealing with uncertainty, and why? A number of researchers recommend a Bayesian approach to help decision-makers incorporate uncertainty into their decisions and to do so in a transparent fashion. Is the use of methods such as Bayesian networks an effective and practicable way for EPA decision-makers to incorporate uncertainty within their decisions and to communicate this uncertainty to stakeholders? If so, how? Are there alternative methods available?*

4.0 General Comments on Uncertainty

Experience suggests that shifts toward new, more informative, but potentially more complex, quantitative uncertainty assessment (QUA) methods inevitably present decision makers with challenges. A greater knowledge of uncertainty, absent an equally sophisticated framework for decision-making and communication, may only increase management challenges. More sophisticated QUA techniques do not automatically create more sophisticated regulatory decision-making. Thus the effective incorporation of uncertainty in decisions by decision makers, and the acceptance of these decisions by stakeholders, will not be accomplished with different or ever more elaborate QUA tools alone.

Deleted: t

Specific methods for performing sensitivity and uncertainty analysis are discussed in Section C.5 and Section C.6, respectively, of the guidance document. The guidance appropriately recommends a sequential approach to evaluating the sensitivity of the model to its components and boundary values to be followed by more in-depth investigation of components and potential interactions that prove to exert the greatest influence on the variability of model outcomes. This is a sound recommendation for developing an understanding of sensitivity in complex models with many factors and many possible interaction effects among those factors. In addition to the work by Saltelli *et al*. cited in the report, other authors have proposed experimental test frameworks (Kleijnen, 2005) for formally examining sensitivity to individual effects and interactions in multi-parameter models. The matrix of statistical methods in Section C.5.7 provides a convenient comparison of the strengths and weaknesses of a progressively more complex set of approaches to sensitivity analysis.

While the merits of various methods for QUA have been discussed, debated, enthused over, and at times derided, including everything from simple bounding analyses through 1-D and 2-D Monte Carlo analyses, to Bayesian techniques, the presumption implied by charge question 4 is that incorporation of uncertainty into decisions is somehow only a function of finding the right mathematical or modeling QUA "tool". Because scientists and researchers are often more comfortable focusing on the "hard science" of models/tools than on the "soft science" that governs the decision making process, often too little attention is given to problem formulation (in its fullest meaning), risk communication, or the perspective of decision makers. The panel cautions that searching for the "right" modeling tool (or uncertainty analysis) may miss the point; namely that models for regulatory purposes are a service to decision makers. Before deciding on a QUA tool, it is incumbent on the modeler to seek input from decision makers and stakeholders as to how and to what extent they may accommodate uncertainty in their regulatory decisions.

To a scientist, expressing and quantifying uncertainty is a good thing. But the single value has a long history of use in regulatory decision-making. Asking decision makers and stakeholders how they view scientific uncertainty, how they would like to see it expressed, and how they see it being used in the decision making process is the necessary precursor to effective and transparent use of any QUA method. In short:

- How much discretion does the decision maker have in addressing uncertainty? During policy development or for an action not directly governed by statute or rule, they may have considerable leeway to do so. Once a statute or rule is in place, they may have much less or no such leeway. Procedural regulations seem particularly resistant to incorporation of uncertainty. Many regulations work with reference to a fixed point (a "brightline" standard) and, despite an awareness that uncertainty exists in where this "fixed" point is actually located, decisions are simply based on whether or not the outcome is above or below that value.
- How will stakeholders react to knowledge of uncertainty and how will this reaction shape the decision making process? To a stakeholder, expressions of uncertainty can be taken as "you don't know", which undercuts support for regulatory decisions. Knowledge of uncertainty also allows opposing interests in a regulatory decision to focus on the highest or lowest value, regardless of its probability. Because there are often significant costs associated with choosing one specific value over another, arguments can erupt over differences in values that are, because of "uncertainty", statistically indistinguishable.

The definition of the term "uncertainty" has been a source of considerable confusion in EPA documents and discussions of models used in environmental risk assessment. The REM Guidance document attempts to clarify the use of the term by: 1) identifying types of uncertainty (model, data, application niche) in Section 3.1.3.1; 2) distinguishing uncertainty from natural variability in model inputs and parameters for different modeling applications; and 3) defining uncertainty analysis (parameters) as distinct from sensitivity analysis (model form and importance of model factors). The Guidance provides some useful but too brief advice (Guidance §4.1.2) on how this uncertainty might be effectively communicated to decision makers and stakeholders. Much more emphasis must be placed on performing a robust and iterative problem formulation with modelers, decision makers, and stakeholders and on correctly conveying model results using non-technical, non-quantitative, and non-condescending communication techniques. Any transparency of QUA methods is only possible if decision makers and stakeholders are engaged early on by inclusive, effective communication and outreach strategies. **The Panel recommends that the REM Guidance should strongly advise modelers to begin model development or use only after they have obtained an awareness of how a decision maker plans to use the information on uncertainty they will be providing.**

Comment: In making this recommendation, we should refer to our discussion of "Problem Specification" in CQ1. And we should say that if the Problem Specification step is done correctly, then the information needed to make this assessment will be available.

4.1 Sensitivity Analysis vis-à-vis Uncertainty Analysis

In Section C.5.1, the REM guidance obscures the distinction between the goals of sensitivity analysis and uncertainty analysis, where it states "…the distinction between these two related disciplines may be irrelevant" (p. 50). While the Panel agrees that the two are interrelated and sometimes confused, the distinction should be clarified in the guidance. Sensitivity analysis is an examination of the overall model response to a perturbation of model

Comment: I totally agree with this assessment.

inputs. The analysis thus can be used to inform model users, decision-makers and stakeholders of where to focus the most resources in terms of developing a better understanding and characterization of the uncertainties for particular components of the model identified as "most sensitive" to perturbations of underlying model parameters. Rather than perpetuating any possible confusion between the focus or goal of these two analyses, the REM guidance should be more transparent in describing the purpose of each, their interrelationship, and the distinction between them. For example, the discussion in Section C.5.5 relating to Monte Carlo analysis currently reads more like a discussion of uncertainty analysis, rather than sensitivity analysis.

Section C.5 would benefit from improved clarity in the distinction between sensitivity and uncertainty analysis. As noted in Cullen and Small (2004), sensitivity analysis is an important adjunct of uncertainty analysis, determining the impact of particular model inputs and assumptions on the estimated risk. Sensitivity analysis is often conducted as a precursor to uncertainty analysis, helping to identify those model assumptions or inputs that are important. If the model outcome is not sensitive to a particular input or set of inputs, there is no need to examine these inputs as part of a more sophisticated uncertainty analysis. Sensitivity analysis is revisited in the subsequent phases of an uncertainty analysis to identify those inputs and assumptions that are significant contributors to the overall variance of the output and/or critical to pending decisions (for an example of the latter, see Merz *et al*., 1992), thereby identifying the uncertainties that matter. In this manner, priorities can be established for further research and data collection efforts. **Therefore the panel recommends that the guidelines articulate a more tangible set of alternatives for addressing model sensitivity/uncertainty.**

Comment: I think that this recommendation is a little vague. I would like us to state something like "additional resources for model improvement should be focused on those processes to which the model state variables are most sensitive **and** are less certain in terms of their formulation and/or parameterization."

4.2 Value of Information – Identifying "Uncertainties that Matter"

After identifying model inputs and assumptions that contribute significantly to variance in the output, it is necessary to consider how to use this knowledge (Cullen and Small, 2005). VOI techniques seek to identify situations in which the cost of reducing uncertainty is outweighed by the expected benefit of the reduction. In short, VOI is helpful in identifying model inputs that are significant because: i) they contribute significantly to variance in the output, and ii) they change the relative desirability of the available alternatives in the decision under consideration. **The Panel recommends that value of information (VOI) techniques be used for assessing the importance of the variability and uncertainty contributed by individual inputs to the expected value (or conversely, the "loss") associated with a decision under uncertainty** (Raiffa and Schlaifer, 1961; Raiffa, 1968; March and Simon, 1958; DeGroot, 1969; Henrion, 1982, Evans, 1985; Finkel and Evans, 1987; Taylor et al., 1993; Dakins et al., 1996; Thompson and Evans, 1997; Costello et al., 1998; Solow et al., 1998). TLT query: Do we need all these references?

4.3 Uncertainty Analysis Practices/Methods (REM guidance Section C.6)

Section C.6 on uncertainty analysis is incomplete in relation to the coverage given to sensitivity analysis in Section C.5. Returning to the discussion of types of uncertainty in Section 3.1.3.1, this section tries to address the "niche uncertainty" under the label of model suitability and "data uncertainty" through a weakly defined discussion of frequentist and Bayesian interpretations of probability. Unlike the rather detailed discussion of methods for corroboration and model sensitivity analysis, there is little true guidance on how to evaluate uncertainty in model parameters and the effect of this uncertainty in decision-making based on model outcomes. The current Draft Guidance touches on the notion of a Bayesian framework and the

use of prior knowledge, expert advice to reflect uncertainty in the model inputs (including parameter values). It also does not distinguish carefully between Bayesian estimation of posterior distribution and associated inferences and decision theoretic approaches which incorporate explicit loss functions for certain errors in inferences. It would be very useful to have a "Box" example of an uncertainty analysis in which there are an established prior for an "uncertain" model parameter, a likelihood for the input data and an updated posterior distribution for model parameters or predictions of interest. **Thus the Panel recommends that the REM guidance (and/or Knowledgebase) provide more practicable information through inclusion of "case study" examples of where and how EPA is currently incorporating uncertainty analysis in environmental models as an integral component of decision-making. In addition, the Panel recommends that section C.6 be enriched to a level comparable to section C.5 on sensitivity.**

The Panel agrees that Baysian approaches are one of several candidate methods suitable for quantifying data uncertainty in appropriate situations. Bayesian methods are certainly appropriate to treating uncertainty in environmental modeling and may be particularly effective in modeling applications where empirical data on the distribution of model parameters in real applications are sparse and expert judgment may provide the most realistic assessment of the prior distributions. A Bayesian treatment of a simple model application or a more complex model with a network of dependencies (conditional relationships) is a theoretically appealing approach to incorporate prior uncertainty into posterior distributions of model outcomes (*e.g.* exposures, concentrations, expenditures, morbidity, mortality, *etc*.). Current software and iterative estimation algorithms have removed many of the computational barriers that once stood in the way of Bayesian treatment of a model application. Yet the removal of computational barriers does not eliminate the need for a solid understanding of the scientific basis for the model and in fact may require a heightened understanding (subjective, expert knowledge) of the prior distributions of parameters. Furthermore, adoption of Bayesian uncertainty analysis methods does not reduce the importance of sensitivity analysis to establish the importance of the model components and their interactions. The effectiveness of the Bayesian approach will be greatest when information on the prior distributions is accurate and new data to support the model application are plentiful. If the prior information is weak or uninformative or the amount of new data available for model parameter estimation is large, the model results will be dominated by the new data. If the new data inputs to the model are weak, the posterior distributions for outputs will be dominated by the prior distribution assumptions.

The panel endorses the recognition that QUA should be an inherent consideration when using models to support regulatory decisions. Yet, given the enormous breadth of modeling paradigms (spatial and temporal scope and degree of complexity), the panel remains cautious in its recommendations regarding specific methods of QUA (*e.g.*, "frequentist" vs. Baysian as suggested in the charge question). The nature and complexity of any particular model, its application within a particular regulatory program, availability of data and resources, *etc*. will all influence the choice of QUA that is appropriate. In some applications, simple sensitivity analyses may be all that is required. Regulatory decisions with far-reaching impacts should endeavor to use QUA tools to provide the public and stakeholder community with greater appreciation for the uncertainty range in the model output decision variables that ultimately define regulatory decision points.

While the panel understands that the REM guidance is not intended to be proscriptive, in its effort to provide an overview of QUA methods, it does not provide sufficient context currently for an end user (e.g., modeler within the regulatory community) to be able to determine the level of QUA that would be appropriate within a particular context or application. Without being proscriptive, the REM guidance could consider providing a more concrete decision framework to help guide the choice of appropriate/available QUA methods. Perhaps as a starting point, the REM guidance could include in the Knowledgebase examples of the nature and degree of QUA currently being implemented or adopted within various EPA programs. For example the Panel is aware of the extensive uncertainty analysis that is an integral component of the 3MRA model. While it is clear that this one example should not be taken to endorse a particular QUA, the Knowledgebase would provide one means of assembling a "library" of such examples with the nature of the QUA, the data requirements, limitations, *etc*. would provide at least some options by way of example that model users and decision-makers could turn to as a resource beyond the cited statistical reference methods.

The appeal of QUA is that it can be used to provide quantitative estimates of the "degree of confidence" when using model results as a component of regulatory decisions. Nevertheless the results should be presented with some caution. It might be tempting to assign a high degree in confidence in the uncertainty analysis based on the adoption of a highly elaborate or complex analysis. Yet, the validity of the QUA is of course dependant on the quantity and quality of the information available for the analysis. The choice of appropriate QUA (frequentist, 1-D, 2-D Monte Carlo, Baysian, *etc*.) can only be made if the intended audience of the REM guidance understands the data requirements (and associated level of effort to conduct the analysis) of the various types of QUA. As compared to the REM guidelines describing best practices for model development/evaluation, the guidelines do not contain a similar set of "best practices" for evaluating, presenting, and incorporating model uncertainty in decision-making. **While references cited in the guidelines provide an array of applicable methods to address model uncertainty, the draft guidelines do not provide sufficient discussion, context, and recommendations necessary to provide a model user/decision-maker with "practicable" information relating to appropriate uncertainty analysis methods and how to convey the results of such analyses.**

4.4 Communicating Uncertainty

Independent of the choice of particular QUA tools, the panel recommends that the REM guidance provide more discussion on the importance of the manner in which results of QUA are communicated to the decision-maker (and public/stakeholders). Graphical methods often serve to convey complex statistical/probabilistic results in a more understandable manner, and the REM guidance should consider including a range of examples in the document. Again, the Knowledge-base with examples would be useful in this regard.

As the analyst/modeler and decision maker are usually not the same individual, it is important to accompany results with the full set of assumptions and caveats encompassed in the analysis. How can uncertainty or probabilistic results be interpreted to help identify the uncertainties that matter most, and to point the analyst to further study or data collection activities that can be most beneficial in reducing these critical uncertainties? As noted earlier, most often only a relatively small subset of inputs is responsible for a majority of the variance in a model output. Morgan and Henrion (1990), Cullen and Frey (1999) and others describe the use

of summary statistics, visual methods, regression approaches and other sensitivity analysis tools to help find the most important input uncertainties. Broader approaches for risk communication and methods for testing the effectiveness of alternative presentations are discussed in Finkel (1990), Bostrom et al. (1992), Morgan *et al*. (1992), Fischhoff *et al*. (1998), Cullen and Small (2005).

The REM Guidance should be clear on the types of model uncertainty that most QUA tools address. That is, the preponderance of QUA methods focus on what the REM Guidance defines as "data uncertainty." Quantitative "model uncertainty" and "application niche uncertainty" present significant challenges that are rarely feasible to address. In addition, empirical or observational data are themselves subject to uncertainty depending on the quantity and quality of empirical data, and it is important to recognize these uncertainties in the context of evaluating the importance of model uncertainties. In the case of directly observed data, there are uncertainties associated with the measurement techniques and with the data analysis processes themselves. In the case of data that are generated by modeling, uncertainties arise as a result of modeling analyses that produced the data. A common example is the difficulty of comparing environmental data collected at a particular point in time and space, to a model prediction based on averaged conditions for a grid cell with spatial parameters and time steps necessarily much different from the conditions under which the measurement was made.

These data uncertainties mean that using data to evaluate models is very much an imperfect process. As a result, the discrepancy between observed data and model simulations does not mean that the model is wrong or not useful. It is particularly important to communicate this concept to decision makers who may favor discounting modeling results if the comparisons between observations and models are less than perfect. In addition, when analysis of data is used in lieu of modeling results because the modeling results do not completely agree with observations, the potential errors and/or uncertainties in the data used for the analysis must be acknowledged. In some cases these uncertainties actually may be more significant than the uncertainties determined for the modeling itself.

The complex nature of data uncertainties and modeling uncertainties needs to be carefully communicated to decision makers. **To promote this discourse as part of the general practice of modeling, the Panel recommends that the guidance document be revised to reflect this important aspect of how model evaluation is to be interpreted and used.**

4.5 References

Bostrom, A., B. Fischhoff and M.G. Morgan. 1992. "Characterizing Mental Models of Hazardous Processes: A Methodology With an Application to Radon," *J. Social Issues*, 48(4), 85-100.

Costello, C.J., Adams, R.M., and Polasky, S. (1998). The value of El Nino forecasts in the management of salmon: A stochastic dynamic assessment. *American Journal of Agricultural Economics*, 80: 765-777.

Cullen, A.C., and Frey, H.C., 1999, *Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing With Variability and Uncertainty in Models and Inputs*, ISBN 0-306-45956-6, Plenum Press, NY.

Cullen, A.C. and M.S. Small. 2005. Uncertain Risk: The Role and Limits of Quantitative Assessment. In *Risk Analysis and Society: An Interdisciplinary Characterization of the Field*. Edited by T. McDaniels and M. Small, Cambridge University Press, Cambridge, UK.

DeGroot, M.H., 1969, *Optimal Statistical Decisions*, McGraw-Hill Book Company, New York.

Evans, J.S., 1985, The value of improved exposure estimates: A decision analytic approach, *Proc. Of the 78th Ann. Mtg. of the Air Pollution Control Assoc*., (June 16-21, 1985), Detroit, Michigan.

Finkel, A., 1990, *Confronting Uncertainty in Risk Management: A Guide for Decision Makers*, Center for Risk Management, Resources for the Future.

Finkel, A.M. and Evans, J.S., 1987, Evaluating the benefits of uncertainty reduction in environmental health risk management, *J Air Pollut Control Assoc*, 37:1164-1171.

Fischhoff, B., D. Riley, D.C. Kovacs and M. Small. 1998. What information belongs in a warning? *Psychology and Marketing*, 15(7): 663-686.

Henrion, M., 1982, The value of knowing how little you know: The advantages of probabilistic treatment in policy analysis, Ph.D. dissertation, Carnegie-Mellon University.

Kleijnen, J.C. (2005). "An overview of the design and analysis of simulation experiments for sensitivity analysis". European Journal of Operational Research, Vol. 164, No. 2, pp.287-300.

March, J.G. and Simon, H.A., 1958, *Organizations*, John Wiley and Sons, New York.

Merz, J., M.J. Small and P. Fischbeck. 1992. Measuring decision sensitivity: A combined Monte Carlo-logistic regression approach. *Medical Decision Making*, 12: 189-196.

Morgan, M.G., Henrion, M., and Morris, S.C., 1980, Expert Judgment for Policy Analysis, Brookhaven National Laboratory, BNL 51358.

Morgan, M.G. and Henrion, M., 1990, *Uncertainty: A Guide for Dealing With Uncertainty in Quantitative Risk and Policy Analysis*, Cambridge University Press, Cambridge, UK.

Morgan, M.G., B. Fischhoff, A. Bostrom, L. Lave, C. Atman. 1992. "Communicating Risk to the Public," *ES&T*, 26(11), 2048-2056.

Raiffa, H., 1968, Decision Analysis: Introductory Lectures on Choices Under Uncertainty, Reading, MA, Addison-Wesley Publishing.

Raiffa, H., and Schlaifer, R.O., 1961, Applied Statistical Decision Theory, Cambridge, MA, Harvard University Press.

Solow, A.R., Adams, R.F., Bryant, K.J., Legler, D.M., O'Brien, J.J., McCarl, B.A., Nayda, W., and Weiher, R. (1998). The value of improved ENSO prediction to U.S. agriculture. *Climatic Change*, 39: 47-60.

Thompson, K.M., Evans, J.S. 1997. The Value of Improved Exposure Information for Perchlorethylene (Perc): A Case Study for Dry Cleaners, *Risk Analysis*, 17(2): 253-271.

Charge Question 5: *The Panel should consider that environmental models will be used by people whose technical sophistication will vary widely. EPA has therefore attempted to cull information about models that broadly serve the needs of all users, using a data template to collect this information (see Table _.) Has EPA identified, structured and developed the optimal set of information to request from model developers and users, i.e., the amount of information that best minimizes the burden on information providers while maximizing the utility derived from the information?*

5.0 General Comments and Suggestions

As indicated in Table __ and Attachment D, the major categories of information collected for the models in the REM Models Knowledge Base (MKB) include:

A. General Information, regarding the model name, contact information, overview, and web link;
B. User Information, concerning technical requirements and basic guides for obtaining and using the model;
C. Model Science, including the conceptual basis for the model and discussion of evaluation steps that have been undertaken and documented for the model (code verification, corroboration with observed data, sensitivity and uncertainty analysis); and
D. Model Criteria, summarizing applicable regulations and the problem domain(s) addressed by the model, including types of pollutants, sources, environmental media, and key fate and transport and exposure and effects processes.

The information targeted in the current data entry sheet addresses most of the critical elements needed by potential users to assess the overall relevance and utility of a model in the MKB, and does so in an effective and efficient manner. However, some additional general categories of information should be added to this list. The Panel believes that this can be achieved without a significant increase in the overall information burden, given the possible consolidations of the current information elicited that are suggested below (TLT query-is this still true?)

A. General Information

The general information entries for the MKB data sheet include:

1. Model Name
2. Model Overview/Abstract
3. Contact Information
4. Model's Home Page

This information is appropriately informative and concise, and the examples we considered in the current MKB provide useful introductions to the models.

B. User Information

The user information entries include:

1. Technical Requirements
    1. Computer Hardware
    2. Operating Systems
    3. Programming Languages
    4. Other Requirements and Features
2. Download Info (with URL)
3. Using the Model
    1. Basic Model Inputs
    2. Basic Model Outputs
    3. User's Guide
    4. Other User Documents

The information requested is useful and appropriate. Most users will not need to know the programming language used by the model, since they will access, download, and use an executable version of the model. Nonetheless, this information could be useful for some users and provides a useful context for system requirements. The MKB should indicate whether the underlying programming language(s) must be obtained or licensed for use of the model.

Under the "Using the Model" section of the data entry, the Panel believes that it would be useful to indicate the level of expertise, both environmental and computer, needed to understand and use the model, and the level of user support provided for the model by its developers, the Agency, or other sources. This information is provided for a number of the models currently in the MKB as part of the User's Guide or Other User Documents fields. Still, it would be useful to explicitly ask for this information as part of the data entry sheet.

C. Model Science

The model science categories include:

1. Conceptual Basis of the Model
2. Scientific Detail for the Model
3. Model Framework (equations and/or algorithms)
4. Model Evaluation (verification (code), corroboration (model), sensitivity analysis, uncertainty analysis)

The requested information addresses many of the key elements needed to document and assess the scientific basis for a model. **However, the Panel does recommend some modifications and additions to the list above. First, defining the Model Framework as the 'equations and/or algorithms' for the model (as is also done in the model glossary) appears counter to the usual use of the word "framework". This term is usually associated with the broader conceptual basis for the model or (by some, see the EPA's Modeling QAPP Guidance Document, page 54) as "the model and its supporting hardware and operating system". A clearer request for the underlying model equations and/or algorithms would be provided using the descriptor "Model Structure and Calculation Methods". Second, the mention of corroboration (model) under**

Model Evaluation should explicitly mention the model's ability to predict observed monitoring data.

The Model Evaluation section of the Model Science entry considers many of the key issues needed to evaluate the scientific rigor behind the underlying model development and previous applications, and addresses many of the elements of good modeling practice that are emphasized in the Draft Guidance document. Indeed, the Panel views an important purpose of the MKB as providing an incentive for model developers and purveyors to conduct and openly communicate their efforts in model evaluation. **From this perspective, the Panel recommends some additional pieces of information that should be elicited and reported, including:**

- **Documented examples of peer review for the model, including reviews conducted by the EPA, other agencies or panels, and papers presented in the peer reviewed literature. Key limitations and needs for improvement that were identified in these evaluations should be reported.**
- **Benchmarking studies in which the model's predictions and/or accuracy were compared with other models.**

The Panel also recommends the inclusion of a section, following Model Evaluation, for the model developer to summarize key limitations of the model and plans or needs for modifications and improvements. This type of self-critique would be both informative to users and motivating to the ongoing improvement of the models in the MKB.

D. Model Criteria

The model criteria elicited and reported include the major categories of:

- Regulations
- Releases to the Environment
- Ambient Conditions
- Exposure or Uptake
- Changes in Human Health or Ecology

The Panel notes that the criteria elicited are highly focused on models for pollutant fate, transport, exposure, and effects. Much of this information is not appropriate for models that address economic activity, behavior, and emissions. These models are differentiated by other key criteria, including whether they predict at the level of the individual, household, firm, sector, region, or national or global economy; whether they are normative (predicting how people *should* behave under various assumptions of rationality and information) or descriptive (reporting how people actually *do* behave); and whether they address the costs or benefits of environmental regulations. As such, the Criteria should first note the genre of the model, whether economic/behavioral vs. physical or engineering science models (though some models, e.g., for predicting emissions, could combine elements of both), and include different subset of information for these.

Specific suggestions by the Panel:

- **Under Regulations, those populating the MKB should be given the opportunity to identify "Other Regulatory or Decision Support**

Applications." These could include US regulations, such as NEPA or Natural Resource Damage Assessments (what law is this under?? Oil Pollution Act/CERCLA), or international agreements or treaties, such as those for ocean disposal or controls on persistent organic pollutants (POPs). It could also include non-regulatory decision support applications, such as for risk communication efforts by state environmental or public health agencies, or life-cycle assessment in support of green design decisions by firms.

- **Under the Releases to the Environment section, a differentiation should be made between models for natural systems (emphasized in the current list) and engineered environments, such as buildings, treatment plants, and water distribution systems. (Models for the latter, such as EPANET, have received increased attention in recent years due to concerns regarding drinking water quality at the tap from accidental contamination and homeland security, and should be sought for inclusion in the MKB.) Also, under Source Type, area source models should be explicitly noted to include larger scale sources, e.g. for nonpoint source runoff in watersheds, biogenic emissions in regional air quality models, or distributed natural or anthropogenic sources to groundwater.**
- **Under Ambient Conditions, the Panel feels that the terms included under Processes (transport, transformation, accumulation, and biogeochemical), while useful information for many fate-and-transport models, is specific enough that it need not be included in these general model criteria. The Panel recommends that this information be replaced with the following, more-general criteria:**
    - **Time scales addressed in the model and whether the model predicts for dynamic or static conditions**
    - **Spatial scales or economic units addressed in the model and whether it provides a primarily distributed vs. lumped representation of the modeled system**
    - **Whether the model is deterministic, predicting single values for model outputs, or statistical/stochastic, predicting a range or distribution of values to characterize variability and/or uncertainty**
- **Under Changes in Human Health or Ecology, the options should be expanded to include natural resource or materials damage, to consider effects, e.g., on visibility, historic buildings, or property value.**

In addition, the Panel recommends that an additional major category of information be elicited and reported (in addition to the major items A-D). The additional category would be list as, E. Model Applications, and point site users to specific examples of regulatory or non-regulatory applications of the model (distinguishing between the two) in the public record and the peer-reviewed scientific literature.

5.1 Track Versions of Models

The Panel recommends that revision tracking be incorporated into the MKB. Such a feature would have several benefits. First, it better reflects the realities of modeling than the current framework in which models are implicitly treated as unchanging. Second, it facilitates a tighter connection between policy analysis and modeling: the documentation for an analysis would specify a particular model version whose characteristics could be retrieved from the database. Third, it would provide valuable insight into the evolution of models over time. It would be possible to observe the extent to which changes in a model are driven by: developments in the underlying science; the availability of new data; the availability of new software or algorithms; the demand for new features; and the correction of programming bugs.

Revision tracking could be implemented as follows:

- A version field and a date field would be added to the data entry form. The contents of the version field would be a character string supplied by the model developer. The string should contain enough information that the developer (or a subsequent maintainer) could reconstruct and rerun that version of the model at a later time. The date field would be the date at which that version of the model was released or placed in service.

- Each time a new version of the model is added to the database, there should be one or more fields describing the significant changes in the model from its previous version. In addition, all other fields associated with the model should default to their settings from the previous version. However, it should be possible to provide an updated version of any field without losing the corresponding field from the previous version of the model.

The documentation burden imposed on model developers would be small. In particular, models whose development has been sponsored, at least in part, by EPA will already have significant changes spelled out in grant proposals or cooperative agreements. Ideally, the MKB would also include information on bugs fixed between versions. With revision tracking in place, the main page for each model would have a link to “Previous Versions”, which would take users to a page showing the dates and revision numbers of all previous vintages of the model in the MKB. Each previous version should be a clickable link showing the list of changes embodied in that version (from above) and include links to other information specific to that version of the model.

5.2 Listing of Key Publications and Applications of Models

The Panel believes that it would be useful to include a list of key references for each model: publications and reports where the model is described or documented, and important applications. Model developers will be able to provide this information easily and it will allow potential users to: (1) find out more about a model; and (2) avoid duplicating previous research; and (3) see example applications. This information would also help answer charge question 7c by showing how widely used and thoroughly peer-reviewed each model is.

5.3 Clarify Questions C1-C3 (TLT query—have the main points below been addressed above?)

The distinctions among questions C1, C2 and C3 in the REM guidance should be made clearer. Question C1 and C3 are intended to match section 2.2 and 2.3 of the guidance document but most model builders and users will probably regard those sections as overlapping considerably. Section 2.2, for example, requests a clear statement and description of each element of the conceptual model, plus documentation of the science behind the model, including: its mathematical form, key assumptions, the model's scale, feedback mechanisms, etc. It seems, in short, to be asking for essentially complete documentation for the model. However, section 2.3 begins with a request for some of the same information: a formal mathematical specification of the concepts and procedures of the model. It is not clear how that differs from the mathematical description requested in 2.2.

It seems as though the intent of C1-C3 is the following. The answer to C1 would be a broad conceptual overview of the model that would be relatively free of technical detail (no equations) and would be accessible to readers from a wide range of backgrounds. It would usually include a diagram showing the relationship between major components of the model. The answer to C2 would provide the technical detail missing from C1 (namely, the model's key equations) and would have specialists as its intended audience. It would provide the theoretical basis for the model. The answer to C3 would describe the model's numerical implementation (data, algorithms, computer programming). This approach would be useful but needs to be spelled out more clearly in instructions accompanying the form. It would also integrate well with version tracking: the answer to C3 will usually change with each revision of the model; the answer to C2 will change periodically; and the answer to C1 – which defines the essence of the model – will generally be fixed.

5.4 References (Shouldn't there be references for this section?)

Charge Question 6: *EPA has developed a data dictionary and database structure to organize the information it has collected on environmental models (**see Attachments E and F**). Has EPA provided the appropriate nomenclature needed to elicit specific information from model developers that will allow broad inter-comparisons of model performance and application without bias toward a particular field or discipline?*

6.0 General Comments

This charge is one of the most specific, yet it overlaps that for CQ5, which provides much of its input, and CQ 7, which provides an insight into the effectively available output. The discussion of the elements of this question is based primarily upon relatively terse, but sometimes vague, information provided by the REM Data Dictionary and the REM Entity Relationship Diagram. The Panel's review of the Data Entry Sheet (CQ5) and related documentation of several individual models appearing in the REM Models Knowledge Base (MKB) were also considered in this question. **This has led the Panel to recommend that the technical issues concerning the specific design of the MKB be addressed by either (1) a separate knowledge base topical report, or (2) an additional appendix to the current guidance document, to allow the main report to concentrate on the Agency's overall plan for the use of this important tool, without ignoring the details of its functional design.**

The Panel's expectation is that the developers of the MKB database structure would also perform the necessary QA review of their Data Dictionary and entity relationships to assure that they are properly drawn and functioning. This aspect is virtually impossible for the Panel to evaluate thoroughly on the basis of the limited details provided on the database structure in the two documents provided. It is similarly difficult for panel members (who are not information technology specialists) to provide much useful advise without a better understanding of the strategy and implementation of the design. Perhaps the separate topic report or MKB Appendix could include all of this definition information and outline of the database design strategy. Panel members were not sure this would be helpful. As noted below, review of the individual model documentation in the MKB provided the Panel with the most insight on the effective results of the application of these tools within its system.

Although the Glossary presented in Appendix A of the report is an undisputed "plus" for the model guidance documentation effort, there are very few of the terms in the Data Dictionary repeated there, as may be expected and appropriate, given the specialized nature of database terminology that is usually unique to the particular database software program for which it was specified. For a database, its functional terminology use has to be clear and internally consistent, regardless of its conformance to the "outside world." It has been noted elsewhere that several of the Glossary terms have varying definitions, as used in different sections of the Guidance report and MKB references—even though they are intended to conform to the Guidance definitions put forth in the Glossary. Although it initially appeared that ongoing efforts may have to include variant definitions (with footnotes to indicate model association); the use of "special guidance-specific" definitions for some terms may be satisfactory if the authors of the guidance carefully

review their use of terminology for consistency of use, and alter the text accordingly. As suggested above, however, the MKB Data Dictionary can function independently and quite satisfactorily, as long as the translation of Data Entry Sheet terminology to database definitions is precisely specified. **The Panel therefore recommends that the Agency follow it own standard QA/QC program procedures for ensuring quality of the all of the underlying information in the MKB system.** From evidence presented to the Panel, it appears that this has already been substantially completed for the functions currently defined. As new functions are added to support new features, including those recommended elsewhere in this report, it will of course be necessary to expand and update this Data Dictionary and repeat many of the QC checks to verify functionality.

The Panel has varying opinions on whether the overall Glossary should include all of the Data Dictionary terminology to assure that referencing is clear to all users. For the reasons outlined above, it appears as though this would potentially add more opportunity for confusion than enlightenment. **Therefore, the recommended approach that would isolate the Data Dictionary in its own self-standing report would seem most advantageous at the current time.** Regardless of the location of this documentation, the panel re-iterates its encouragement to extend the QA/QC procedures followed to establish the initial quality of the MKB into the larger QA program needed to maintain the information, as well as the hardware and software systems needed to implement it.

6.1 Model Performance Information

This charge asks about including database information that is "unbiased". However, as indicated by the presentations made by Region 5 and 10 representatives on February 7, there is also a need for a place in the database for additional "classification" information, which may go beyond that requested from the developer, and which may appear "biased", if it includes "recommendation" information. This would be a subsection of the database specifically devoted to information that helps agency regulatory-model application staff and "outside applicants" to identify the "most appropriate" candidate models. (A new "model selection program" that is under development by ORD was demonstrated at the panel's review meeting. It appeared to be a potentially valuable tool, but several panel members cautioned that it should produce an output file that includes a matrix of candidate models, rather than a singe "recommendation", so that the user of the tool can more fully consider which of several candidates best fits the problem application at hand). Much final model-selection decision making is presently achieved by regional or state agency discussions that come to agreement on the most appropriate site-specific model choice for major projects at a particular decision-point. However, as noted further below, the MKB would be more valuable, if cumulative EPA problem application experience could be more consistently represented in the database, along with the present basic model description information.

The Panel is in concurrence on the importance of eliciting and including information on historical model performance and particular application experience from various model users (both other modelers and decision makers), as well as model developers. This was not especially motivated by any desire to minimize "biases" in reporting. There was some concern that developers of a model may not be in a position to fully (or objectively) judge its behavior in various contexts. Avoiding or minimizing bias would seem to require gathering reviews from as broad a user base as possible. It now appears that the current approach, which utilizes only information volunteered by the model developers, would tend to ensure that individual "biases"

are included, without any real opportunity to neutralize them. This situation may be the unintentional result of using a more open narrative format for developers to explain features of the model. It may be noted that the panel review of the current Data Entry Sheet, the Data Dictionary, and the Entity Relationship Diagrams did not suggest that there were any particular features that would "bias" the selection or representation of models. Instead, as noted both above and below, the reviewers were interested in seeing more information, as this could include application experience with "competing" models.

In fact, the inclusion of additional information on the history of performance suggested by several panel members would be more likely to include "opinions" as to the quality of performance, hopefully supported by comparison with appropriate measurement data sets. This extra information was viewed as important to prospective model users, even though it would be likely to also include some "biased" information. As long as instances of "preconceptual bias" can be identified and flagged or filtered, the availability of previous application experience (especially successes) would be a valuable component of the MKB information set. (Given the wide variety of models included, this "openness" may be helpful to both agency and "outside" users; but perhaps some form of warning of the risk of potential bias should be included with any new "performance history " element, so that the new users are fully aware of this limitation). **The Panel recommends that the Agency clarify the intended roles of the "inside" and "outside" users of the MKB system and how that affects the priorities for the user interacting with the system (including supplemental, even if "biased," application history information).**

6.2 Additional Recommendations

To address details issues of CQ 6 more specifically, the panel reviewers observed that the dictionary and database do capture much of the information necessary to assess model performance; but there were some noted exceptions:

- *CONCEPT*: This results from problem formulation, but may or may not convey to the user useful information about the problem or set of problems (Guidance §2.1) for which the model was developed. Another field should be added *("PROBLEM")* to concisely capture descriptive information about the original application problem.
- *DECISIONDOCS*: As written, this field seems to focus on how to use (run?) the model, how to produce output, and what experience there has been with running the model. This (or a new) field should include information or links to examples of when, how, and where the model was used to support an actual decision or decisions. Qualitative opinion on how the model performed would be acceptable/desirable. What benefits and problems did decision makers and stakeholders experience when using the model? This element should include a date entry so potential users can better judge the currency of the model.
- *DOWNLOADINFO:* This should include information on the size of the model (zipped and unzipped), whether it is one file or a collection of files, and whether its setup will require changes in system files.
- *DIR ENTRY STATUS* and *REVISION_DATE:* It is not clear what is meant by "last reviewed"—whether the date given would be for when the model itself was reviewed or when its entry into the dictionary was last updated? There should be information on when the model itself was last reviewed by its developer, as well as documentation (or links to such) of any and all changes, including errata and enhancements. It would also be useful to

have documentation of problems encountered (or improvements suggested by) actual users of the model. All of this may be considered in MODELCONTACTINFO but the database appears to be placing any "institutional memory" of the model's behavior in a person, who may or may not be available. The reviewers thought that there should also be a fields consistently indicating whether model documentation is available online, who is responsible for preparing and maintaining this documentation, and the date it was last reviewed and/or updated.

- *EVALUATION* includes four questions, but without performance information, the first three seemed less useful (recognizing that they might represent the only information available for newer models).
- *MODEL_CATALOG* Table information given in Data Dictionary is too cryptic to tell whether any model performance information would fall into the descriptions provided there.
- *PROG_LANGUAGE:* This should also indicate whether any other software (particularly proprietary, e.g., ArcINFO) is required to operate the model.

Panel reviewers considered their observations in reviewing the **Aquatox, CalPuff, IPM,** and **TRIM_FATE** models in reaching their conclusions about the performance of the identified database elements. Overall, the construction of the system appeared to be generally well-designed, but with opportunity remaining for expanding its focus to include more consistent information on model use experience and performance in a format that would make it more uniformly easy for users to compare models of interest for a particular candidate application. There are several key features that the Panel would like to see improved or expanded so that the MKB can be most effectively used by the EPA and its stakeholders. The existing Data Dictionary and Database Structure appear to be adequate to address existing features of the current MKD. However, as this tool is expanded to include new features recommended by either this panel or the agency's developers, it will be necessary to add new structural elements and data elements; and this will require an ongoing additional QA/QC effort. **Therefore, the Panel recommends that the following issues should receive further consideration and attention:**

- **A consistent QA review of the current content of the information contained in the MKB [some model feature/description errors (at the user interface level) were noted by panel members];**
- **Follow-up requests to developers who supplied original information to supply missing data for the minimum set of descriptors that the agency decides are essential to proper model selection;**
- **Entries into the data dictionary be clearly defined and made as consistent as reasonably possible, with the text in the guidance document and data entry forms.**
- **Provision of a mechanism that actively solicits feedback from the user community regarding application experience and model performance, both inside and outside the agency, beyond voluntary e-mails to designated contacts for individual models.**

Interest in seeing continuing improvement in what appears to be an extraordinarily valuable model information system led the Panel to express concerns for a near-term commitment by the agency to the possible appointment of a Knowledge Base "System Librarian". This might be someone within EPA, or an appropriately qualified contractor (e.g. a national laboratory technical library). This position would emphasize those aspects that affected input of new information and system QA to improve information consistency and reliability with

time, making the MKB a national resource for quality comparative information on both new and established models used in the regulation of the environment.

Charge Question 7: *To facilitate review for this particular charge question, the panel should focus on three models that represent the diversity of model information housed within the Models Knowledge Base. These models are: (1)* ***Aquatox,*** *a water quality model; (2)* ***Integrated Planning Model (IPM)****, a model to estimate air emissions from electric utilities; and* ***NWPCAM****, an economic model.*

Using these three models as examples and emphasizing that EPA is not seeking a review of the individual models, but rather the quality of the information provided about the models, EPA poses the following questions to the Panel. Through the development of this knowledge base, has EPA succeeded in providing:

(7a) easily accessible resource material for new model developers that will help to eliminate duplication in efforts among the offices/regions where there is overlap in the modeling efforts and sometimes communication is limited?

(7b) details of the temporal and spatial scales of data used to construct each model as well as endogenous assumptions made during model formulation such that users may evaluate their utility in combination with other models and so that propagation of error due to differences in data resolution can be addressed?

(7c) examples of "successful" models (e.g., widely applied, have been tested, peer reviewed etc.)?

(7d) a forum for feedback on model uses outside Agency applications and external suggestion for updating/improving model structure?

7.0 General Comments

The Panel commends the Agency for developing the Models Knowledge Base and strongly supports its continued improvement. This type of resource has been needed for some time and even in its draft form, the Knowledge Base provides an easily accessible resource for the modeling community that, if maintained and used, will significantly improve the development and application of models both internal and external to the Agency.

In answering questions 7b-7d, the panel focused primarily on the two suggested models (i.e., AQUATOX and IPM) along with a third model selected by the Panel (CalPuff). However, it was necessary to go beyond these models to address question 7a. The Panel interprets question 7a as being asked in the context of a model developer who might use the MKB to screen existing Agency models for use in a specific application or for model technology to include in a new model to support a specific decision. In this case the Panel found it necessary to identify a number of similar models (i.e., atmospheric dispersion models or water quality models) and assess first the number of models available to choose from and, second, the consistency, transparency and comparability of the data for these similar models.

In answering charge question 7a, the Panel finds that the MKB has the potential to provide readily accessible information about models; however the amount and quality of information can

be improved. For charge question 7b, the Panel recognizes that the information provided in the MKB is not highly detailed. As a result, sufficient level of detail about scales of data used and assumptions made during the formulation of any specific model in the cannot be obtained from this tool alone. However, the MKB does allow for the initial identification of candidate models with links and references for obtaining further information. For question 7c, the Panel agreed that the three models considered in this review were all good examples of successful models both in their regulatory role and in the way they are presented in the Knowledge Base. For the final question, the Panel was not satisfied with the current form of feedback mechanism for the Knowledge Base. More detailed observations, suggestions and recommendations follow.

7.1 Vision for the Knowledge Base

The issues surrounding which models to include in the MKB are not trivial; the Panel recognizes that this choice can have significant implications for the application of this tool in support of decision-makers. The Panel is concerned that without a clear vision, the MKB may increase the burden on Regional and State offices by implying that a particular model is "endorsed" by the Agency. **The disclaimer on the main page of the MKB makes it clear that models in the Knowledge Base are not endorsed by the Agency but the Panel suggests that this disclaimer be clearly presented at the top of each "Model Report" page as well.**

Part of the Vision for the MKB should specify the role of this resource in the development or life cycle of models. More specifically, there needs to be a clear statement about what models are included in the Knowledge Base and what models or types of models (if any) are excluded. This will require that the Agency provide a clear definition of what a "Regulatory Model" is or move away from this terminology towards a more inclusive title. The Panel recognizes that in addition to providing a repository or library of mature models that are actively used by the Agency; the Knowledge Base can play an important role in the development of new models and the improvement of existing models. **For this reason, the Panel recommends that the Agency include models at all stages of their life cycle with a process for identifying to users those models that are new, actively being develop, currently used for decision making and nearing retirement.**

An important aspect of any model repository from the perspective of a model developer or new model user is that it be as comprehensive as is feasible. In other words, users must be confident that when they use the MKB to identify an appropriate model for a task, it is likely that all relevant models have been considered. The draft MKB provides a good start but needs to continue to incorporate additional models used by the Agency. Many of the Agency's Offices, Programs, and Regions have developed their own clearing house for models; the Agency should make an effort to bring these existing data bases under the umbrella of the Knowledge Base. **The Panel recommends that the Agency identify these parallel Agency supported databases (e.g., the Support Center for Regulatory Air Models (SCRAM), the Center for Exposure Assessment Modeling (CEAM), etc.) and develop a plan to incorporate them into the MKB. If it is not feasible to incorporate these existing databases at this time, then the Panel suggests providing a current list of – and links to – these additional databases on the main page and the search page of the MKB.**

The process of identifying and including existing models is clearly an important step to insure that the Knowledge Base is comprehensive. It is also important to continue to populate this MKB with new models as they emerge. **To accomplish this, the Panel recommends that**

the Agency incorporate new models into the Knowledge Base as part of their initial application within the Agency. The information in the MKB for a given model is, or should be, part of the model development process so submitting this information as part of a model's initial application should not be an added burden to the model developers. Nevertheless, the Panel recognizes that it may be necessary for the Agency to provide additional incentive (positive or negative) as part of their plan to encourage what is currently a voluntary effort by modelers to put their model in the MKB.

7.2 Quality Assurance and Quality Control

In addition to its role as an institutional memory, the MKB, in its current form, is clearly a tool designed and developed to support regulatory decisions by delivering useful information about prospective models for specific applications. The database itself is not unlike other "models" developed to support regulatory decisions. As noted in CQ6, the development of the MKB and the information provided in it should be subject to the same level of quality control and quality assurance that any Agency modeling effort is expected to include. **Therefore, in addition to the Vision Statement discussed earlier, the Panel recommends that the Agency provide a link on the main page of the Knowledge Base that takes the user to the Agency's plan for insuring the quality (integrity, utility and objectivity) of information provided.** At a minimum, this should contain the following elements:

- Problem specification that identifies the drivers for setting up the MKB (i.e. reduce duplication of effort, improve networking, facilitate model development, satisfy training needs, …)
- Clear identification of the user community or "clients" for the MKB. There was some ambiguity among the Regional representatives at the face-to-face meeting about whether the Knowledge Base satisfied their specific modeling needs and as a result there appeared to be a lack of "buy in" from the Regions.
- Identify specific performance criteria for the MKB information along with selection criteria for models in the database and identify who will be responsible for insuring that these criteria are met.
- If non-Agency models are eventually included in the MKB (see previous bullet on selection criteria) then the QA/QC plan should identify how these models will be treated or presented and who will absorb the burden of oversight for these models.

The level of detail provided by each model should also be balanced. In the draft MKB, the details provided for models differ widely. An example of a model where information is very sparse is TRACI. Scientific detail is often just a statement of units used in the model (e.g., the SWIMODEL includes only the following statement under Scientific Detail "The model uses fixed units (S.I.)." and is missing Conceptual Basis all together). In other cases, it is not apparent that the sections include comparable information. For example, it is often difficult to distinguish between the Conceptual Basis, Scientific Detail and the Model Framework sections. **The Panel recommends that improved guidance be provided as part of the data entry sheet to insure that the correct type of in formation is input into each field.** This will also facilitate search functions by making sure those submitting the information realize what fields are searched. It may be necessary to request a keyword list from the model developer. As an example of this last point, the Panel found that the CalPUFF was not identified in the key word search using the phrase "air dispersion". Although "air" and "dispersion" are in the title or abstract, the phrase

"air dispersion" is missing and as a result the model is not identified when the search is based on this common phrase. In another case, a search for "vapor intrusion" models (currently a timely topic), there were no matches in the MKB. A search for "indoor air" models produced three matches, but none that appeared usable for the vapor-intrusion set of problems. This illustrates that there is still some significant work ahead to verify that the priority regulatory problems being addressed in Regional offices of EPA today are adequately considered in selecting candidate models to be included in the Models Knowledge Base.

7.3 Layout and Navigation of Knowledge Base

The Panel reviewed the information provided in the MKB in Question 5 and, in addition to information that is currently provided, identified several additional pieces of information that should be elicited when a model is introduced into the Knowledge Base. In this section, the Panel provides observations about the current layout of the MKB and provides suggestions for where new information should be presented.

The current layout of the MKB is logical and generally easy to maneuver (with some exceptions noted later). The Panel found that much of the summary level material was readily accessible on the three main Report pages. The more detailed information is generally available through appropriate links. However, the Panel notes that in several cases, including the CalPuff model, information is not provided for specific fields and rather than leave these fields blank, they are apparently removed from the Report. For example, the "Model Framework" and the "Model Evaluation" fields are often missing. The Panel recognizes that the Agency attempted to "cull information about models that broadly serve the needs of all users…" but once this minimum information is identified, it should be provided for all models. **The Panel recommends that if information is not provided for specific fields, those fields should be left blank rather than be removed from the Report. A blank field provides clear information about a model while a missing field is ambiguous.**

Overall, it was possible to use the MKB to obtain general information about the existence and availability of frequently used models and more detailed information about a specific model. But, really understanding how a given model works and what its specific strengths and weaknesses are would appear to require either going into the detailed documentation or contacting an actual user. Navigating the knowledge base was somewhat cumbersome, in that apparently different links go the same destination, links to critical information (e.g., model change bulletins) are obscure, return links (i.e., return from exit disclaimer) when to the key word search page. In addition, several different pages (10 in the case of CALPUFF) needed to be accessed to gain a sense of model operation and capabilities. Perhaps accommodating the somewhat bewildering array of models and their varying characteristics is what's causing these navigational inefficiencies but, regardless, it would be helpful if access to model information could be more streamlined.

7.4 Updating the Knowledge Base

The Panel recognizes that the MKB is a "living demonstration of the recommendations from the Guidance for Environmental Models". This suggests that the Knowledge Base will evolve and adapt to the specific needs of the user community. The comments above also support the premise that this will be an ongoing process of optimization. Optimizing the MKB will

ultimately require an understanding of the user community and an active and transparent feedback mechanism. To facilitate this, the panel recommends that voluntary user profile and registration information be requested so that use profiles can be developed. This information can also provide a mechanism for announcements to be distributed when necessary.

Improving the MKB and the models contained in it will ultimately depend on the quality of feedback from "external users" and the ability of new users to access this information. The Knowledge Base is currently limited to a single contact and does not provide any suggested format for comments nor does it provide for open dialogue and discussion of modeling experience. This seriously limits the Agency's ability to adapt the MKB and improve its utility. This lack of an open forum also limits the model developers from gaining experience from model users and it limits the ability of new modelers to learn about specific experience and application of a particular model. **The Panel recognizes the challenges associated with hosting an open forum on an Agency web site but recommends that the Agency reconsider including a transparent user feedback mechanism that will facilitate an open dialogue for the models in the MKB.**

7.5 The Role of the Knowledge Base as a "Model Selection Tool"

The panel is not entirely convinced about the utility of a model selection tool or expert system that accesses the MKB to facilitate model selection. However, the Panel suggests that if such a tool is developed for application at the regions, labs and states, then the effort should be considered "model development" and as such should clearly follow the guiding principles in the Guidance on Environmental Models (TLT query—is this the same as the REM guidance?).

If such a model selection tool is developed, it will likely be used early in the life of a project so identifying specific needs and valuing these specific needs in a way that would facilitate a model ranking would be difficult to achieve. **Therefore the Panel recommends that any tool developed by the Agency to facilitate model selection based on the Knowledge Base should simply present the models in a comparative matrix in the form of a side-by-side comparison table like one would see in the car sales industry.**

Appendix B provides more detailed information about Panel members' experiences in accessing and using specific models.

Appendix A

Definition of Modeling Terms

Appendix B

Panel Members' Experiences using the MKB (TLT note: requires further editing)

1.0 CALPUFF

This charge question is addressed first in general and then specifically for the CALPUFF air dispersion model, which was selected as the example for evaluation. The CALPUFF example evaluation starts from the "Models Knowledge Base" page, and then goes to the listing of available models, and from that to the CALPUFF model report. With respect to Question **7(a),** if we weren't going to a specific model, it would be hard to decide, using this list alone, how to choose from among the several seemingly air-related models listed (however, the keyword search capability is helpful for this). A model overview on the "general information" page provides information that addresses, in part, Question **7(b).** Going to the "user information" page gives us information on downloading and the availability of user's guides. Here the heading "Using the Model" is slightly misleading in that it implies information on how the model is used to make decisions but is actually about how a modeler would run the model. This section also provides no citations or links as to application of model results in actual decision making. Moving to the "model science" page, we find much of the information relevant to Question **7(b).** Although the "Recommendations for Regulatory Use" section is informative, it also provides no citations or links as to how model results have faired in actual decision making. The "Model Evaluation" section is clearly about evaluation of the model as a model and not as a decision support tool.

One of the download links from the "user information" page takes us to EPA's SCRAM website, as does a similar link for "model homepage" on the "general information" page. The SCRAM website is apparently the only point at which it is possible to access the critical "Model Change Bulletin" and "Model Status" records, which are obscurely included only as "Notes" in smaller font. There appears to be considerable overlap in these two sets of information and the question arises why they couldn't be combined in one obviously accessible location (e.g., on the "user information" page). The link to the NTIS site is probably necessary but models without online documentation would appear to be at a disadvantage. Getting to CALPUFF on the SCRAM website from either the "general information" or "user information" pages provides you with a link to the model developer's website, who is a contractor and not EPA. A link directly to this website is also on the "user information" page. Thus you have three apparently different links on two different pages all leading to the same destination, a non-EPA website. This seems unnecessarily convoluted. It is not entirely clear until this point that the real substance on the model resides with a contractor and not with EPA. Re: Question **7(d),** is the role of the EPA "model contact" that of an Agency internal and external interface for the model? If so, such a role is not clear at this point. It also seems that a more direct link to the actual developer and maintainer of the model would be helpful. The top page of developer's website provides little information about the science of the model but does nicely summarize model updates, provides links to its regulatory status, a download, and training opportunities. The "regulatory status" page provides information similar to that found on the EPA "model science" page but goes further by offering links to notices and reports on regulatory use.

2.0 IPM (AK)

Before going through the charge, I focus on the IPM write-up as it appears on the CREM site. This write-up is very thorough on what is asked. It is clear, concise and helpful as a first description of what this model contains and what it is used for. It turns out that almost all of the write-up is a verbatim cut and past from the IPM Model Documentation. I don't see anything wrong with that, as long as the appropriate items are covered in the appropriate depth. However, in examining the IPM Model Documentation, I note that pg. 2-5 begins a section on Key Methodological Features (e.g., details of how the load duration curve is specified and information on how the dispatch order is determined) that could be simplified and incorporated into CREM to give the reader one level further down in detail.

7a) It is beyond the MKB to provide adequate information for new model developers because these users will want detailed information about potentially competing models, information that can only be obtained, if at all, from model documentation. The IPM site does contain links to such documentation. So, in this sense, new modelers may benefit. But, to be frank, an internet search or a search of the EPA site would immediately bring up such documentation without the need for the KB. New developers would be particularly keen on knowing the IPM's limitations, questionable assumptions, and the like, none of which is in the CREM. This is information that could be asked of models for the database, which is currently not asked.

As for IPM, in particular, this model is extremely well entrenched in the Air Office and would be, therefore, unlikely to attract "new model developers."

7b) Assumptions are not endogenous. By definition they are exogenous (but this is only semantics). Evaluating the utility in combination with other models should be in contrast to other models so that errors in model selection/duplication can be minimized. I can understand that a high spatially resolved model would be more accurate than one of lower resolution, but choices about resolution always involve tradeoffs, such as in model complexity, data availability, and model flexibility, and the types of questions a model is designed to answer. The charge a question does not encourage this kind of thinking (although earlier questions may) and the database is silent on providing information to aid in this type of thinking as well.

For IPM, spatial resolution is clearly given – all 48 states plus DC are covered along with the number of coal producing regions are identified. Temporal resolution is less clear. The time step for the model is not explicit but the forecasting horizon of the model is clear. Exogenous assumption is not given, but model documentation provided in the link would surely provide this information for this model. There is a list of key assumptions (e.g., perfect foresight, pure competition) in the IPM Model Documentation document and this information should be provided in the CREM. Again, as noted for question 7a, modelers should be asked to provide a write-up for CREM of significant limitations of their models in terms of simplifications, strong assumptions, factors ignored and outside the scope so that could be argued should be in the model, and the like.

7c) For the IPM model, yes.

7d) the feedback mechanism is not readily apparent.

- **CalPUF (BH)**

7a) The KB does provide sufficient information to accomplish goal 7a, in that it allows users of the data base to locate candidate models which might serve their purpose. It would probably be impractical for the KB to provide enough information so that users can determine which models are suitable for every application, but the KB can certainly help eliminate duplication by providing a limited number of candidates that must be evaluated.

7b) The answer is no, but the data base could not reasonably be expected to provide sufficient detail to fully address the question. The KB can and should answer basic questions such as "at what temporal and spatial scales has the model been shown to operate successfully?" and (for air models in the GAQM) at what scales are these models considered to be "preferred" or acceptable alternatives to preferred models. This information should be sufficient to allow users of the KB to ask the right questions, but probably cannot provide complete answers, since understanding the "endogenous assumptions made during model formulation" will require detailed understanding of the model algorithms beyond the scope of the KB.

7c) The KB obviously includes many highly successful models (including CalPUFF), but it is not clear how users of the KB will be able to determine for themselves which ones are "successful". Clearly models "preferred" in the GAQM qualify, but a similar gold standard may not exist for other media. Other GAQM models may be assumed to have achieved some measure of "success". A list of the successful applications of a model could be useful in providing a measure of its success.

7d) the answer is no. The KB appears to have no formal feedback mechanism other than contacting Mr. Pasky Pascual. Feedback from model users could be extremely valuable to others who have specific modeling needs. The information would help users answer the questions posed in 7a-c. The KB could solicit comments from users of the models, and post these comments on a bulletin board linked to the KB. Postings should allow for anonymity, as some model users might not want to be identified personally as users of the models – it's not unusual for busy modelers to get ph0one calls from graduate students wanting help running complex environmental models for thesis projects.

General comments on the KB: The KB should attempt to provide at least a minimum set of key information for all models. Under User Information, essentially all that is provided for CalPUFF is links to the SCRAM and to the developer's web site, but for some other vendor-supplied models, summary information is provided n the KB itself (plus appropriate links). Because vendors may provide information on models as they see fit, it would be beneficial to have at least a summary of basic information about each model in the KB. This information should include computational requirements (including operating systems supported and requirements for other software), descriptions of input data requirements, and descriptions of model output. Additional useful information could include some examples where the model was successfully applied, along with references and contact information to facilitate further research into the suitability of models for specific applications.

Something is wrong in the keyword search feature on the KB primary panel, since entering "air dispersion" produced only three results, all related to the RAIMI. This search should produce several hits including CalPUFF. (the search is only performed on the title and abstract so if the word is missing from this field it will not be found. In CalPUFF, the abstract does not include the word "air")

The "browse for models by selecting for environmental indicators" seems to have no search criteria which locate CalPUFF.

After inadvertently selecting "Exit Disclaimer" on the CalPUFF User Information page, I tried to return using the "Return to Previous Page", but was instead taken to the "Browse to Knowledge Base" pate.

On the CalPUFF model developer's website, a reference is made to the GAQM, while in the KB, there is a reference to Appendix W. In fact, both refer to the same document. The KB should be clear that Appendix W and the GAQM are the same. Both the KB and the developer's web sites should provide links to the GAQM.

- **CalPUF (PG)**

The KB, though extremely useful concept/tool, should not be considered as providing a substitute (e.g., in summary report form) of the detailed information that has to be retrieved from the open literature in order to compare potentially relevant models for an application.

The summary report should provide a very simple summary of the "applicability range" of the models. For example the summary report states that "CALPUFF is intended for use on scales from tens of meters from a source to hundreds of kilometers" but does not mention the fact that the minimum temporal resolution of the model (hourly averages) restricts its applicability to a range of simulation that do not include important short-term phenomena (e.g., emergency events such as accidental spills), dispersion of heavy gases, etc.

Especially important information that should be provided in the KB include i) all input/output formats, ii) all software tools (public domain and proprietary – as well as potential substitutes) that are needed in order to fully utilize the model's capabilities, iii) available databases of inputs (potentially outputs from other models), iv) past evaluations (especially cross-evaluation) studies involving the model(s) of concern.

The KB provides the opportunity to turn abstract discussions in the Guidance into specific examples; however, in order to achieve this, more information needs to be included in the Kbase.

The models in the Kbase differ widely in terms of ranges, attributes, objectives etc. The completeness/focus of the "model report" information also varies widely (but the amount of information provided.

The structure needs to be enhanced to include i) more specific information on the format (ASCII, NetCDF, DXF, Shapefiles, etc.) of inputs and outputs, as well as on ranges and resolution/aggregations attributes and ii) direct (i.e., not through external sites) of documentation (not only user guides but also application/evaluation studies) for each model.

– AQUATOX (JD)

I have not had an opportunity yet to review the AQUATOX documentation material in depth, but I have scanned the material to allow some preliminary thoughts on the material provided with regard to its adherence to the REM Guidance, including model evaluation, uncertainty analysis, and model transparency. Also, there are four specific sub-questions that have been asked.

7a) Easily accessible resource material for new model developers?

A new model developer would find the documentation and descriptive material on the technical and theoretical aspects of AQUATOX very helpful in elimination of duplication. Processes in the model are well documented.

7b) Presentation of model assumptions and data used for model formulation?

The technical documentation of release 2.0 is reasonably thorough with regard to process documentation and assumptions inherent in the model. However, the format of the report **does not follow** the recommended elements for model documentation given in the draft guidance. I would prefer seeing a separate "Model Development" chapter that includes a conceptual model, a complete disclosure of all model assumptions and resulting caveats, and data used to convert the conceptual model to a mathematical model.

Release 2 does specify that it can only be used in a non-dimensional or one-dimensional mode and does discuss the temporal scales of use. There are certainly limitations to the model use imposed by these assumptions; the document does discuss these.

7c) Examples of successful model applications?

This model has not had a long history of application in its current form, although it does have a long history of application of previous incarnations of the model (e.g., as CLEAN or CLEANER or PEST). The user manual presents several examples of applications of the model; however, only one of them shows system data that allows the user to assess the success of these applications – Onondaga Lake. On the web site, they do offer model "validation" examples in an EPA report published in 2000 that includes Onondaga Lake, PCBs in Lake Ontario, and agricultural runoff in Coralville Reservoir. I have not yet thoroughly reviewed these model evaluation exercises yet, but it does appear that they compare AQUATOX with data and previous models for these systems, which is good.

No discussion of regulatory use of the model. Does make point that this is a multi-stressor, multi-response model.

7d) There is an opportunity to become a registered user on the web site; however, it is not clear that this is the portal to provide feedback to the agency on outside application experience or suggestions.

The CalPUFF site under the "user information" the section on Technical Requirements" is missing. To facilitate identification of all relevant models for an application, each model should have the exact same major sections. AquaTox and CalPuff are both missing the Framework section on the model science page. Again, even if sections are left blank, they should be included for every model to facilitate use.

Figures are very helpful in the model conceptual basis as used in the IPM. Otherwise, the information provided by the three models is not necessarily in line with the definition of "Conceptual basis" in the guidance. The descriptions range in detail from providing a statement of what the model does to what inputs are required but not always clear on what the conceptual basis is (i.e., is it mechanistic or empirical or something in between). The BLP model only has two of the four sections in the model use section. There also appears to be some confusion between "Scientific Basis" and "Model Framework" which is illustrated by the similar level of information provided in the SB section for CalPUFF and the MF section of the IPM. With the IPM it appears that the text was just pasted in to the sections on conceptual basis and the framework was used as overflow indicating that it was not clear to the imputer (model development team?) exactly what information was being requested.

It would be useful if the web page on "User Information" provide an indication of level of user expertise required to apply the model. For example, the IPM states that "The model's core LP code is run by ICF Consulting…" while at the other extreme, the dBase states that "User needs only moderate level of technical education and/or modeling experience.". This type of information is valuable for users planning to actually apply the models beyond just learning what is available.

The level of detail provided is very different across the models. An example of a model that is very sparse is TRACI. Scientific detail is often just a statement of units used in the model (e.g., the SWIMODEL includes only the following statement under Scientific Detail "The model uses fixed units (S.I.)." and is missing Conceptual Basis all

together). The NWPCAM report is missing the model evaluation section. This speaks to the issue of quality control across the Knowledge Base. Is the Agency responsible for the quality of information provided on these pages? If so, there will need to be some oversight provided to the various people inputting the data to get some level of consistency in the information provided.

Return to previous page link on the exit to disclaimer does not return to previous page but returns user to the search page.

A good example of a version tracking matrix or table is given on the PRIZM version index page that is found by following the links to the model web site that goes through the EPA Center for Exposure Assessment Modeling site at http://www.epa.gov/ceampubl/products.htm by selecting the model from the menu.

It would be helpful for keeping the information up to date if an annual automated message was sent to individuals listed as the model contacts requesting updates or reviews of the material on the data base. As an incentive, this could be accompanied with a report on the number of accesses that were made to the specific model.

- The user community may provide a very effective policing mechanism to maintain model quality, especially when money is at stake, which provides a clear opportunity and incentive for improving the models contained in the Kbase. However, this requires feedback mechanism, which is currently lacking.

It may be appropriate to consider technology transfer as an option in the long-term plan for the resource.

This appendix summarizes comments related to the form and function of the knowledge base with specific emphasis on models selected to facilitate the review and response for charge question 7.

-CALPUF:

The CALPUFF example evaluation starts from the "Models Knowledge Base" page, and then goes to the listing of available models, and from that to the CALPUFF model report. With respect to Question **7(a),** if we weren't going to a specific model, it would be hard to decide, using this list alone, how to choose from among the several seemingly air-related models listed (however, the keyword search capability is helpful for this). A model overview on the "general information" page provides information that addresses, in part, Question **7(b).** Going to the "user information" page gives us information on downloading and the availability of user's guides. Here the heading "Using the Model" is slightly misleading in that it implies information on how the model is used to make decisions but is actually about how a modeler would run the model. This section also provides no citations or links as to application of model results in actual decision making. Moving to the "model science" page, we find much of the information relevant to Question **7(b).** Although the "Recommendations for Regulatory Use" section is informative, it also provides no citations or links as to how model results have faired in actual decision making. The "Model Evaluation" section is clearly about evaluation of the model as a model and not as a decision support tool.

The Knowledge Base does provide sufficient information to accomplish goal 7a for the CalPUF model in that it allows users of the data base to locate candidate models which might serve their purpose. The Knowledge Base should not be considered as providing a substitute (e.g., in summary report form) of the detailed information that has to be retrieved from the open literature in order to compare potentially relevant models for an application. It would be impractical for the Knowledge Base to provide the level of information necessary for users to determine which models are suitable for every application, but the Knowledge Base can certainly help eliminate duplication by providing a limited number of candidates to consider. Evaluating these candidate models requires consistency in the presentation of information.

The Knowledge Base cannot reasonably be expected to provide sufficient detail to fully address a model users/developer's questions about CalPUF. However, the Knowledge Base can and should answer basic questions such as "at what temporal and spatial scales has the model been shown to operate successfully?" and (for air models in the GAQM) at what scales are these models considered to be "preferred" or acceptable alternatives to preferred models. This information should be sufficient to allow users of the Knowledge Base to ask the right questions, but probably cannot provide complete answers, since understanding the "endogenous assumptions made during model formulation" will require detailed understanding of the model algorithms beyond the scope of the Knowledge Base.

The models in the Knowledge Base differ widely in terms of ranges, attributes, objectives etc... The completeness/focus of the "model report" information also varies widely relative to the amount of information provided. For example, under User Information, essentially all that is provided for CalPUFF is links to the SCRAM and to the developer's web site, but for some other vendor-supplied models, summary information is provided n the Knowledge Base itself (plus appropriate links). Because vendors may provide information on models as they see fit, it would be beneficial to have at least a summary of basic information about each model in the Knowledge Base. As indicated in the Panels Report, this information should include computational requirements (including operating systems supported and requirements for other software), descriptions of input data requirements, and descriptions of model output. Additional useful information could include some examples where the model was successfully applied, along with references and contact information to facilitate further research into the suitability of models for specific applications.

As another example of the need for consistency, the CalPUFF site under the "user information" section, the link to "Technical Requirements" is missing. To facilitate identification of all candidate models for a specific task, each model should have the exact same major sections. AquaTox and CalPuff are both missing the Framework section on the model science page. Even if sections are left blank, they should be included for every model to facilitate use of the Knowledge Base. The top page of developer's website provides little information about the science of the model but does nicely summarize model updates, provides links to its regulatory status, a download, and training opportunities. The "regulatory status" page provides information similar to that found on the EPA "model science" page but goes further by offering links to notices and reports on regulatory use. This also highlights the need for some support by the Agency to synthesis information provided by the model developer to provide a consistent format and level of detail.

Navigating the CalPUF pages was somewhat awkward. The "environmental indicators" search was the least useful since it presupposes knowledge of how the Agency defines and uses such indicators. One of the download links from the "user information" page takes us to EPA's SCRAM website, as does a similar link for "model homepage" on the "general information" page. The SCRAM website is apparently the only point at which it is possible to access the critical "Model Change Bulletin" and "Model Status" records, which are obscurely included only as "Notes" in smaller font. There appears to be considerable overlap in these two sets of information and the question arises why they couldn't be combined in one obviously accessible location (e.g., on the "user information" page). The link to the NTIS site is probably necessary but models without online documentation would appear to be at a disadvantage. Getting to CALPUFF on the SCRAM website from either the "general information" or "user information" pages provides you with a link to the model developer's website, who is a contractor and not EPA. A link directly to this website is also on the "user information" page. Thus you have three apparently different links on two different pages all leading to the same destination, a non-EPA website. This seems unnecessarily convoluted. It is not entirely clear until this point that the real substance on the model resides with a contractor and not with EPA. Something seemed to be wrong in the keyword search feature on the Knowledge Base primary panel, since entering "air dispersion" produced only three results, all related to the RAIMI. This search should produce several hits including CalPUFF. The Panel recognizes that the search is only performed on the title and abstract so if the word or phrase is missing from this field it will not be found. In CalPUFF, the abstract does not include the word "air" so it is not picked up by searching for "air dispersion". The "browse for models by selecting for environmental indicators" seems to have no search criterion which locates CalPUFF either. Also, after inadvertently selecting "Exit Disclaimer" on the CalPUFF User Information page, the "Return to Previous Page" takes the user to the "Browse to Knowledge Base" page rather than the previous page.

On the CalPUFF model developer's website, a reference is made to the GAQM, while in the Knowledge Based, there is a reference to Appendix W. In fact, both refer to the same document. The Knowledge Base should be clear that Appendix W and the GAQM are the same. Both the Knowledge Base and the model developer's web sites should provide links to the GAQM.

The Knowledge Base obviously includes many highly successful models (including CalPUFF), but it is not clear how users of the Knowledge Base will be able to determine for themselves which ones are "successful". Clearly models "preferred" in the GAQM qualify, but a similar gold standard may not exist for other media. Other GAQM models may be assumed to have achieved some measure of "success". A list of the successful applications of a model could be useful in providing a measure of its success. To allow one to judge the level of success of a particular model, the summary report should provide a very simple summary of the "applicability range" of the model. For example the summary report states that "CALPUFF is intended for use on scales from tens of meters from a source to hundreds of kilometers" but does not mention the fact that the minimum temporal resolution of the

model (hourly averages) restricts its applicability to a range of simulation that do not include important short-term phenomena (e.g., emergency events such as accidental spills), dispersion of heavy gases, etc.. As indicated in the Panel's report, especially important information that should be included in the Knowledge Base are i) all input/output formats, ii) all software tools (public domain and proprietary – as well as potential substitutes) that are needed in order to fully utilize the model's capabilities, iii) available databases of inputs (potentially outputs from other models), iv) past evaluations (especially cross-evaluation) studies involving the model(s) of concern. The Knowledge Base provides the opportunity to turn abstract discussions in the Guidance into specific examples; however, in order to achieve this, more detailed and consistent information needs to be included in the Kbase.

The role of the EPA as the "model contact" is somewhat unclear for the feedback forum. The appropriate or desired role of the model contact as either an internal (Agency) or external (public) interface for the model is not clear at this stage of the development of the Knowledge Base. It also seems that a more direct link to the actual developer and maintainer of the model would be helpful. The Knowledge Base appears to have no formal feedback mechanism other than contacting Mr. Pasky Pascual. Feedback from model users could be extremely valuable to others who have specific modeling needs. The information would help users answer the questions posed in 7a-c. The Knowledge Base could solicit comments from users of the models, and post these comments on a bulletin board linked to the Knowledge Base. Postings should allow for anonymity, as some model users might not want to be identified personally as users of the models – it's not unusual for busy modelers to get phone calls from graduate students wanting help running complex environmental models for thesis projects.

- **IPM**

Before going through the charge, we focus on the IPM write-up as it appears on the CREM site. This write-up is very thorough on what is asked. It is clear, concise and helpful as a first description of what this model contains and what it is used for. It turns out that almost all of the write-up is a verbatim cut and past from the IPM Model Documentation. Cut and past from existing model documentation is sufficient as long as the appropriate items are covered in the appropriate depth. However, in examining the IPM Model Documentation, page 2-5 begins a section on Key Methodological Features (e.g., details of how the load duration curve is specified and information on how the dispatch order is determined) that could be simplified and incorporated into CREM to give the reader one level further down in detail. Therefore, to maintain consistency in the level of detail presented in the Knowledge Base it may be necessary for existing documentation to be re-written with a consistent format across all models. We recognize that this would likely require a dedicated scientific editor/webmaster that is charged with the task of working with the model developer to prepare the documentation for upload onto the Knowledge Base.

Although the Panel recognizes that the Knowledge Base alone will not likely be able to provide sufficient information for new model developers that require a detailed understanding of potentially competing models. This type of information can only be obtained, if at all, from model documentation. The IPM site, which can be accessed from the Knowledge Base, does contain links to such detailed documentation. So, in this sense, new modelers may benefit. But, to be frank, an internet search or a search of the EPA site would immediately bring up such documentation without the need for the Knowledge Base. New developers would be particularly keen on knowing the IPM's limitations, questionable assumptions, and the like, none of which seems to be available in the Knowledge Base. As for IPM, in particular, this model is extremely well entrenched in the Air Office and would be, therefore, unlikely to attract "new model developers."

The level of detail on "endogenous assumptions" for a given model is completely dependent on the information provided by the model developer so at some level this may be out of the realm or control of the developers of the Knowledge Base. Evaluating the utility in combination with other models, or more appropriately in contrast to other models requires first that competing models be identified through the Knowledge Base and second that the Knowledge Base provide enough information at a comparable level of detail so that appropriate choices on which model to use can be made. A high spatially resolved model is expected to be more accurate than one of lower resolution, but choices about resolution always involve tradeoffs, such as in model complexity, data availability, and model flexibility, and the types of questions a model is designed to answer. The charge question does not encourage this kind of thinking (although earlier questions may) and the database is silent on providing information to aid in this type of thinking as well.

For IPM, spatial resolution is clearly given – all 48 states plus DC are covered along with the number of coal producing regions are identified. Temporal resolution is less clear. The time step for the model is not explicit but the forecasting horizon of the model is clear. Exogenous assumptions are not fully provided directly on the Knowledge Base model page, but model documentation accessed through links would surely provide this

information for this model. There is a list of key assumptions (e.g., perfect foresight, pure competition) in the IPM Model Documentation document and this information should be provided in the Knowledge Base. Again, as noted earlier, modelers should be asked to provide a write-up for the Knowledge Base of significant limitations of their models in terms of simplifications, strong assumptions, and factors that have been ignored and/or are outside the scope that could be argued should be in the model, and the like.

The Panel agrees that the IPM model is a good example of a "successful" model but a forum for feedback on model uses outside Agency applications and external suggestion for updating/improving model structure is currently inadequate.

– AQUATOX

A new model developer would find the documentation and descriptive material on the technical and theoretical aspects of AQUATOX very helpful in elimination of duplication. Processes in the model are well documented on the Knowledge Base and the associated model documentation provided on the model web site.

The technical documentation of release 2.0 is reasonably thorough with regard to process documentation and assumptions inherent in the model. However, the format of the report **does not follow** the recommended elements for model documentation given in the draft guidance. The Panel would prefer seeing a separate "Model Development" chapter that includes a conceptual model, a complete disclosure of all model assumptions and resulting caveats, and data used to convert the conceptual model to a mathematical model.

Release 2 does specify that it can only be used in a non-dimensional or one-dimensional mode and does discuss the temporal scales of use. There are certainly limitations to the model use imposed by these assumptions; the document does discuss these.

This model has not had a long history of application in its current form, although it does have a long history of application of previous incarnations of the model (e.g., as CLEAN or CLEANER or PEST). The user manual presents several examples of applications of the model; however, only one of them shows system data that allows the user to assess the success of these applications – Onondaga Lake. On the web site, they do offer model "validation" examples in an EPA report published in 2000 that includes Onondaga Lake, PCBs in Lake Ontario, and agricultural runoff in Coralville Reservoir. It does appear that these evaluation exercises compare AQUATOX with data and previous models for these systems, which is good.

No discussion of regulatory use of the model. The documentation does make the point that this is a multi-stressor, multi-response model.

Finally**,** the model web site does provide an opportunity to become a registered user; however, it is not clear that this is the portal to provide feedback to the agency on outside application experience or suggestions.

- **Other Models**

As noted in the Panel's report, it was necessary to evaluate other models in the Knowledge Base to assess level and consistency of detail an ease of use. The following comments are general observations from this survey.

The Panel found that figures and diagrams were particularly helpful in the section describing the model conceptual basis as used in the IPM. The information provided by a number of the models is not necessarily in line with the definition of "Conceptual basis" in the guidance. The descriptions range in detail from providing a statement of what the model does to what inputs are required but not always clear on what the conceptual basis is (i.e., is it mechanistic or empirical or something in between). The BLP model only has two of the four sections in the model use section. There also appears to be some confusion between "Scientific Basis" and "Model Framework" which is illustrated by the similar level of information provided in the Scientific Basis section for CalPUFF and the Model Framework section of the IPM. With the IPM it appears that the text was just pasted in to the sections on conceptual basis and the framework was used as overflow indicating that it was not clear to the imputer (model development team?) exactly what information was being requested.

It would be useful if the web page on "User Information" provide an indication of level of user expertise required to apply the model. For example, the IPM states that "The model's core LP code is run by ICF Consulting…" while at the other extreme, the THERdbASE states that "User needs only moderate level of technical education and/or modeling experience.". This type of information is valuable for users planning to actually apply the models beyond just learning what is available.

The Panel found that the level of detail provided in the Knowledge Based is very different across the models. An example of a model that is very sparse is TRACI. Scientific detail is often just a statement of units used in the model (e.g., the SWIMODEL includes only the following statement under Scientific Detail "The model uses fixed units (S.I.)." and is missing Conceptual Basis all together). The NWPCAM report is missing the model evaluation section. This speaks to the issue of quality control across the Knowledge Base. If the Agency is going to take responsibility for the quality of information provided on these pages, then there will need to be some oversight provided to the various people inputting the data to get some level of consistency in the information provided. Or, as indicated earlier, there may be a need for a dedicated Scientific Editor.

The Panel has recommended that the Knowledge base include more detail on model version. A good example of a version tracking matrix or table is given on the PRIZM version index page that is found by following the links to the model web site that goes through the EPA Center for Exposure Assessment Modeling site at http://www.epa.gov/ceampubl/products.htm by selecting the model from the menu.

It is important that the information on the Knowledge Base be current. It would be helpful for keeping the information up to date if an annual automated message was sent to individuals listed as the model contacts requesting updates or reviews of the material on the data base. As an incentive, this could be accompanied with a report on the number of accesses that were made to the specific model.

The user community for the Knowledge base may provide a very effective policing mechanism to maintain model quality, especially when money is at stake. This provides a clear opportunity and incentive for improving the models contained in the Knowledge Base. However, this requires a more transparent feedback mechanism, which is currently lacking.

Once this resource is developed, the Panel recognizes that the Knowledge Base may be a good candidate for technology transfer in the long-term plan for the resource. The resource has value and maintaining current information and continuing to improve the tool may be better left to the private sector, possibly in the form of a non-profit organization.

Appendix C (TLT note: incomplete)
Documentation "nits"

1. Foreword: " …adopt different approaches". Under what circumstances?

2. Page 8. "models is either empirical or mechanistic" add "or both"?

3. Page 10 (and others) "sound science" This term begs he question of other ideas associated with the level of approximation, both in model development and in model selection for a given task. How does the term "sound science" relate to concepts like "best available science" or "appropriate level of process complexity" or "spatial and temporal aggregation", etc. These and other similar terms are not used much in the document. Will best modeling practice (as described in this guidance) lead to the above, i.e. "sound science", and its implications?

4. Page 13 discussion of "object oriented platforms modeling systems" seems circular and is very confusing. The guidance should revise the description, or leave it out altogether (could provide a reference).

5. Page 20, P2, L2: "which was…" to "which is…"

6. Page 27. In section 4.2.2 the guidance discusses measures that can be taken to support modeling decisions, particularly in case of later court challenges. The guidance should also

discuss communicating information about the modeling to stakeholders early and often during the modeling process. Doing this provides an opportunity for objections to be raised before final regulatory actions are taken, and weakens the case of challengers who did not raise objections until late in the process.

7. Page 43 last P, L1: consistent

8. Page 44, last P. Why is kurtosis mentioned, and not skew? Has anyone ever used kurtosis for model corroboration?

9. Page 50 P2, L2: Figure C.5.1

10. Page 60 Ref 61 has no date.

11. Model Knowledge Base. Has the information in the knowledge base been subject to Agency QA protocols? The Panel believes it should.