2

INPUT TO CHAPTER SIX OF NRC/SKI PAPER: "REGULATORY VIEWS ON VALIDATION OF PERFORMANCE ASSESSMENT MODELS"

Prepared for

Nuclear Regulatory Commission
Contract NRC-02-88-005

Prepared by

Stephen H. Spector
and
Patrick A. LaPlante

Center for Nuclear Waste Regulatory Analyses
Arlington, Virginia

June 1992

Table of Contents

6.2 EPA Guidance on Use of Models for Air Quality Programs

6.2.1 EPA Regulation Under the Clean Air Act

Air quality models, generally developed by the U. S. Environmental Protection Agency (EPA), are used extensively in demonstrating compliance with EPA approved state programs under the Clean Air Act (CAA)(42 U.S.C. 7401 et seq. as amended). The "Guideline on Air Quality Models" was originally published by EPA in April 1978. It was incorporated by reference into the EPA's regulations (40 CFR §52.21(l) and 40 CFR §51.166(l)), thus giving it the force of law. As stated in a subsequent proposed change to the regulations and Guideline, "the purpose of the Guidelines is to promote consistency in the use of modeling within the air management process."

The EPA's model evaluation process for air quality regulation involves ongoing refinement of both models and evaluation techniques. This has resulted in several technical conferences, a 1986 revision of the Guideline (51 F.R. 32176) (USEPA, 1986), and a February 13, 1991 Notice of Proposed Rulemaking (56 F.R. 5900) which proposed further modifications to the rule and model usage. In a paper to a conference on application of Air Pollution Meteorology, an EPA official stated "regulators need to know how accurately a model estimates pollutant concentrations, and modelers need a yardstick to evaluate and compare the performance of alternative models... The interest is in how well the model estimates compare to measured air quality data ('operational' evaluation); the issue of why they perform well or poorly ('diagnostic' evaluation) has been deferred to a latter time." (Tikvart & Cox, 1984). EPA uses the term "evaluation" because most of the models are valid to some extent; therefore, what is critical is how the model is used at a specific location and, thus, the EPA's emphasis is on evaluation of model application.

Currently, EPA evaluates models through a notice and comment process. EPA's general strategy for gaining broad acceptance of models involves: publication of developed models for public comment, analysis of comments and suggested alternative models, publishing final procedures and sponsoring periodic national conferences. When accumulated modeling experience identifies a need for further development, or insufficiencies are identified in approved models, the process is repeated. EPA is currently preparing the analysis of public comments on draft Supplement B to the Guideline and Final Rulemaking.

The published Guideline is intended for use by EPA regional offices in judging adequacy of modeling analyses performed by EPA, state and local agencies, and industry. The Guideline contains EPA approved modeling techniques and data bases plus specific procedures for submitting alternative models with necessary justification, documentation and evaluation (relationship of estimates to observations).

Each model offered for inclusion in the approved list undergoes a systematic performance evaluation and a scientific peer review. A standard set of statistical performance measures has been developed consistent with recommendations from an American Meteorological

Society (AMS) Workshop on performance measures for air quality models (Fox, 1981). This workshop was part of a cooperative EPA-AMS agreement to review regulatory air quality modeling. Since the primary interest of EPA is to determine how well model estimates compare with measured air quality data, performance measures include fractional residual difference between observed and predicted concentrations (as well as for their respective standard deviations), correlation analysis, and comparison of frequency distributions of observed and predicted data. At key points in the evaluation process, model developers are asked to comment on selection of data bases and parameters, options for application, and the impact of coding changes. The results of the performance evaluation serve as a basis for the peer review which concentrates on performance and theoretical aspects of the models. While technical limitations exist for quantification of all uncertainties in the process, EPA relies on the results of performance tests and theoretical evaluations to indirectly quantify an element of the total uncertainty for decision-making purposes.

Regarding application of preferred models, the EPA Guideline supports data base selection by providing an approved list of the minimum data necessary for standard air dispersion modeling and the basic level of analysis needed to comply with air quality regulations. Sensitivity analyses are encouraged as a way to provide information on the effect of data and model uncertainties. Use of confidence intervals for statistical values is also recommended. EPA suggests providing such information to the decision-maker to indicate the effect of uncertainties on results. Reliance by the decision-maker on the modelers "best estimates" is recommended as the basis for regulatory decisions ("best estimate" implies EPA preferred models or equivalents are used in accordance with the guidance). EPA is working on procedures for determining accuracy of models, quantifying uncertainty and expressing confidence levels in pollution control decisions (Hillyer & Burton, 1980; Thrall et al., 1985). No formal recommendations are provided on analysis of model uncertainty due to current technical information limitations.

While EPA recommends preferred models, alternatives may be used under certain circumstances. The Guideline specifies three conditions under which an alternative may be used: 1) if the model is shown to produce concentration estimates equivalent to estimates obtained using a preferred model; 2) if a statistical performance evaluation conducted using measured air quality data indicates the alternative model performs better for the given application than a comparable model in the EPA guidance; or 3) if there is no preferred model for the specific application but a refined model is needed to satisfy regulatory requirements.

In the absence of an EPA approved model for a given application, it is required that: (i) the selected model is shown to be applicable to the problem on a theoretical basis; (ii) the data bases which are necessary to perform the analysis are available and adequate; and (iii) performance evaluations of the model in similar circumstances have shown the model to be unbiased toward underestimates.

Additional EPA documents on evaluation of alternatives include "Interim Procedures for Evaluating Air Quality Models" (USEPA, 1984), which contains procedures and techniques on evaluating a model based on superior performance and "Protocol for Determining the Best Performing Model" (Cox, 1988), which contains statistical techniques for comparing models. Interim evaluation procedures with a focus on implementation experiences are also available (USEPA, 1985).

6.2.2 Legal Review of Validity of EPA Model Usage

The largest body of litigation involving challenges to a model or its specific application concerns CAA compliance. Most cases involve questions as to the relationship between a model's predictions and a specific site's characteristics. Generally, in the cases where presentation of an evaluation process is made and data shows facts at the site correspond to parameters of the model, the agency is upheld.

One leading case (Ohio v. EPA, 1986) involves a state challenge to EPA's use of a computer model (CRSTER) and the assertion that the "predictions are not accurate reflections of actual pollution concentrations." The court agreed and held "EPA's reliance on the CRSTER model without testing the model against any monitored emissions from the plants and ambient air quality data from the area around the plant is arbitrary under these circumstances." In this case the court does not require re-validation for every use of the model. "By no means does the court insist that all models be validated at all sites. We find only that the accuracy of CRSTER at the site has not been sufficiently demonstrated to meet the arbitrary and capricious standard of review." In this case, the 6th Circuit Court appears to establish a rule that "[i]n the absence of a record supporting the trustworthiness of agency decision-making tools as they were applied, we cannot uphold those tools application." Therefore, this case establishes what amounts to a "legal definition" of an acceptable application of a model as one which is supported by a record of evidence detailing the models trustworthiness in relation to specific site characteristics and data. The case was reaffirmed and remanded by the 6th Circuit Court of Appeals at 798 F.2d 880 [24 ERC 1817] (Aug., 1986). It was sent back to EPA to determine the appropriate evaluation procedure for the model.

Two of the additional cases which were cited in the above case Mision (Mision Industrial Inc., et al., 1976) and South Terminal (South Terminal Corp. et al v. EPA, 1974) also involve challenges to EPA's application of computer models and the relationship of the models to actual conditions. In both cases, the 1st Circuit Court of Appeals upheld EPA's use of the models. An example of the relevant arguments and rulings from Mision is:

> Finally, Mision Industrial claims that it [EPA] was in error to approve a revised plan based on an uncalibrated model, that no real life data has been used in arriving at or calibrating the accuracy of the predicted relationship between sulfur-in-fuel and air quality. The EPA counters that calibration on incomplete data is not good practice and that the agency supports Puerto Rico's decision to rely on theoretical data in the light of conservative assumptions and calculations for the lack of

precision. ... Petitioner's ... concerns were communicated in detail to the EPA Administrator during the federal comment period. He in turn appears to have considered their views and objections and determined, without at least obvious unreason, that the computer model was a satisfactory predictive tool on which to base Puerto Rico's revision. ... This is an area where EPA's "expertise is heavily implicated" Sierra Club v. EPA, ..., and we may not substitute our judgement for that of the Administrator.

Similarly, the appeals court in South Terminal stated that the "Court...was required to make a careful and searching inquiry into the facts and to assure itself that EPA's technical conclusions no less than others were founded on supportable data and methodology and that such conclusions met the minimal standards of rationality."

6.3 Use and Validation of Models Outside EPA Permitting

6.3.1 Use and Validation of Models by OSHA

The U.S. Department of Labor's Occupational Safety and Health Administration's (OSHA) Office of Risk Assessment uses modeling primarily to conduct risk assessments for hazardous materials in the workplace. Since exposures are often determined by actual monitoring data, the complex process of modeling exposure is not a factor in many OSHA assessments. A fundamental source of uncertainty in OSHA risk assessment modeling is use of dose-response factors based upon data which is extrapolated from animal studies to the human population (i.e., there is no available human toxicity data). Since OSHA has a limited budget and uses well established risk assessment techniques, they do not develop models "in house" or have a general OSHA procedure for model evaluation or validation. Most evaluation is done on a case-by-case basis. When modeling is used to establish specific standards, risk assessment procedures are published in the Federal Register for comment and revised or justified as necessary. Most comments pertain to the validity of risk factors and assumptions used in modeling. Mediation involving interested parties has been used by OSHA as a means to establish consensus on modeling parameters, assumptions, and scenarios (54 F.R. 20680).

6.3.2 Legal Review of Models Used Outside EPA Jurisdiction

While much of the applicable case law on use of computer models relates to EPA modeling, cases involving model usage by other government agencies (e.g., OSHA, DOT) provide additional evidence on court determinations.

OSHA refers to a Supreme Court case (IUD v. API, 1980) regarding an occupational standard for benzene to guide its understanding of what constitutes sufficient evidence for supporting a determination of risk to workers. While this case does not explicitly address the issue of model validation, it does provide insight into the court's views on scientific evidence which, in turn, can help bound the understanding of the level of validity appropriate for models which are used to produce evidence. The court stated:

> ... Although the Agency has no duty to calculate the exact probability of harm, it does have an obligation to find that a significant risk is present before it can characterize a place of employment as 'unsafe.'
>
> "Second, OSHA is not required to support its finding that a significant risk exists with anything approaching scientific certainty. Although the Agency's finding must be supported by substantial evidence, 29 U.S.C. § 655(f), § 6(b)(5) specifically allows the Secretary to regulate on the basis of the 'best available evidence.' As several Courts of Appeals have held, this provision requires a reviewing court to give OSHA some leeway where its findings must be made on the frontiers of scientific knowledge. ... Thus, so long as they are supported by a body of reputable scientific thought, the Agency is free to use conservative assumptions in interpreting the data with respect to carcinogens, risking error on the side of overprotection rather than underprotection. (IUD v. API, 1980)

In a subsequent proposed rulemaking for another occupational exposure standard, OSHA interprets the court findings in IUD v. API as an indication that use of risk assessment modeling, which may involve inherently uncertain mathematical estimates, is an acceptable means of supporting their determinations of risk (54 F.R. 20676).

Another case involves a transportation department's use of traffic modeling and air pollution predictions for a proposed highway widening project (Florida DOT v. JWC Co., Inc., 1981). This case indicates how questions can arise regarding the representativeness of input data to the site. Here, the Florida Department of Environmental Regulation successfully argued DOT assumptions regarding inputs to both the traffic and air pollution models did not correspond to the actual conditions at the site -- leading to erroneous modeling results. While this case does not directly address validation of models, it does represent the court's tendency to focus on the issue of whether valid models have been used appropriately.

6.4 Summary and Identification of Issues (for joint NRC/CNWRA section)

From the analysis of court cases discussed in Sections 6.2 and 6.3, it is evident important factors in establishing a case for validity of a model and its application are: 1) modeling is based on theories which are scientifically valid and have gained acceptance in the relevant scientific community, 2) use of a theoretically valid model must be shown to be consistent with known conditions at the site being modeled, and 3) to the extent possible, site specific information should be used as data inputs and results of modeling should be validated against "real world" measurements. When site specific data are incomplete or nonexistent, the courts have supported use of conservative models as a means to err on the side of safety. In general, the courts do not expect the high level of certainty common in the sciences to make legal determinations on acceptable use of models. Rather, a reasonable case must be made which explains the assumptions and methodology used by the modeler and shows they are within the state of the art.

Regarding model validation, the EPA continues to refine procedures for model evaluation and use which are generally consistent with the above points. Both EPA and OSHA incorporate public participation in their selection of modeling approaches to facilitate consensus building. The sponsoring of technical conferences and other forms of peer review by EPA helps ensure methods are consistent with the current state of the art. These agencies also rely on the use of conservative assumptions in their application of models when data or theoretical limitations are present.

The experience within these agencies is relevant to general issues regarding development of a useful strategy for model validation in the HLW program. Nonetheless, there are unique aspects of the HLW disposal program which are not present in these agencies' work and are not represented in the reviewed case law. The information obtained from legal database searches and agency contacts focusses on modeling conditions in present or past time frames and no examples were identified which address validation issues unique to modeling long-term future system behavior.

9

REFERENCES

American Textile Manufacturers Institute (ATMI), et al. v. Donovan, et al., 452 U.S. 488 (1981).

Cox, W.M. 1988. Protocol for Determining the Best Performing Model. U.S. EPA, Research Triangle Park, NC. (Docket No. A-88-04, II-I-19); Computer Sciences Corporation, 1990.

Cox, W.M. 1988. Bootstrap System Users's Manual. U.S. EPA, Research Triangle Park, NC. (Docket No. A-88-04, II-I-33); Computer Sciences Corporation, 1990.

Florida DOT v. J.W.C. Co.,Inc., 396 So.2d 778 (Fla. App. 1981).

Fox, D.G. 1981. *Judging Air Quality Model Performance.* Bulletin of the American Meteorological Society, 62(5):599-609.

Hillyer, M.J. and C.S. Burton. 1980. *The ExEx Methods: Incorporating Variability in SO_2 Emissions Into Power Plant Impact Assessment.* Systems Applications, Inc., San Raphael, CA. Prepared under Contract No. 68-01-3957 for EPA, Research Triangle Park, NC. (Docket Reference No. II-B-37).

Industrial Union Department (IUD), AFL-COI, et al. v. A.P.I, et al., 448 U.S. 607, 655, 656, (1980).

Mision Industrial, Inc., et al., v. EPA, 547 F.2d 123 (1st Cir. 1976).

Ohio v. EPA, 784 F.2d 224 [23 ERC 2091] (6th Cir. 1986).

South Terminal Corp., et al. v. EPA, 504 F.2d 646 (1st Cir. 1974).

Thrall, A.D., T.E. Stoeckenius and C.S. Burton. 1985. *A Method for Calculating Dispersion Modeling Uncertainty Applied to the Regulation of an Emission Source.* Systems Applications, Inc., San Raphael, CA. Prepared for the U. S. EPA, Research Triangle Park, NC. (Docket Ref. No. IV-G-I).

Tikvart, J.A. and W.M. Cox. 1984. *EPA's Model Evaluation Program.* Paper presented at the Fourth Joint Conference on Applications of Air Pollution Meteorology, Portland, OR.

USEPA. 1984. *Interim Procedures for Evaluating Air Quality Models (Revised).* EPA Publication #EPA-450/4-84-023. U.S. EPA, Research Triangle Park, NC. (NTIS No. PB 85-106060).

USEPA. 1985. *Interim Procedures for Evaluating Air Quality Models: Experience with Implementation*. EPA Publication #EPA 450/4-85-006. U.S. EPA, Research Triangle Park, NC. (NTIS No. PB 85-242477).

USEPA. 1986. *Guideline on Air Quality Models (Revised) including Supplement A*. EPA Publication #EPA-450/2-78-027R. U.S. EPA, Research Triangle Park, NC.