136892

United States General Accounting Office

# GAO

Report to the Chairman, Subcommittee on Oversight and Investigations, Committee on Energy and Commerce, House of Representatives

August 1988

# AIR POLLUTION

## Reliability and Adequacy of Air Quality Dispersion Models

136892

GAO

United States
General Accounting Office
Washington, D.C. 20548

**Resources, Community, and Economic Development Division**

B-220184

August 24, 1988

The Honorable John D. Dingell
Chairman, Subcommittee on Oversight and Investigations
Committee on Energy and Commerce
House of Representatives

Dear Mr. Chairman:

This is the second of a series of three reports evaluating the Environmental Protection Agency's (EPA) management of its air quality dispersion models.[1] The issues contained in these reports surfaced in your review of our report Air Pollution: Improvements Needed in Developing and Managing EPA's Air Quality Models, (GAO/RCED-86-94, Apr. 22, 1986). Although EPA provided you certain information regarding its air models, you still had some questions about the accuracy and adequacy of air quality dispersion models and requested that we provide additional information. This report presents the status and results of EPA's (1) testing of models to determine their reliability and ranges of reliability and (2) efforts to develop additional models to meet identified needs.

Air quality dispersion models provide estimates of pollution concentrations. Since the mid-1970s, such models have been used as a basis for regulatory decisions under the Clean Air Act and have had a major impact on the cost and extent of pollution controls and the selection of sites for new industrial sources. EPA developed a program in the early 1980s to evaluate model reliability by comparing model estimates with actual measured pollutant concentrations at selected sites. The evaluation program was intended to provide a consistent basis for assessing model reliability. As part of a separate effort, EPA has determined that better and additional models are needed for it to fully implement and monitor the programs required by the Clean Air Act.

[1] Air Pollution: EPA's Efforts to Develop a New Model for Regulating Utility Emissions, (GAO/RCED-88-57, Jan. 22, 1988), addressed EPA's policy for purchasing computerized models and its efforts to obtain the Advanced Utility Simulation Model. We will issue a third report at a later date on EPA's answers to your questions regarding whether model estimates are appropriately used in regulatory and attainment decisions.

## Results in Brief

We found that 45 of the 48 air quality dispersion models used in making regulatory decisions have been evaluated either by EPA or other organizations as shown in table 1. However, 11 of the evaluations did not conform to the statistical approach EPA currently uses. This nonconformity occurred because the models were evaluated (1) by organizations other than EPA or (2) by EPA before the current approach was adopted. The evaluations conducted, to date, indicate that model reliability (i.e., a model's percentage of over or underestimation) has ranged from underestimated pollution concentrations of 50 percent for one model to overestimated pollution concentrations of 1,720 percent for another model. These results, however, are not conclusive measures of model reliability, as substantially different results are obtained for the same model evaluated at different sites.

**Table 1: Status of Models Evaluated**

| | Models included in EPA's Guideline on Air Quality Models (revised)[a] | Models not included in guideline | Total |
|---|---|---|---|
| Models evaluated: | | | |
| by EPA using uniform statistical approach | 29 | 5 | **34** |
| by organizations other than EPA | 9 | 0 | **9** |
| by EPA before adoption of the uniform statistical approach | 2 | 0 | **2** |
| **Total models evaluated** | **40** | **5** | **45** |
| Models not evaluated | 3 | 0 | **3** |
| **Total** | **43** | **5** | **48** |

[a]EPA's Guideline on Air Quality Models (revised) provides guidance to model users concerning recommended air quality modeling techniques and identifies the models that should be used in regulatory applications.

In 1985, EPA updated an earlier study which determined that to fully implement the requirements of the Clean Air Act, it had to meet 241 modeling needs. These include the need to estimate various pollutants in rural, urban, and complex-terrain areas and to estimate pollution generated by mobile sources. Models under development as of September 1987 will satisfy 65 of the identified needs. EPA considers the remaining 176 modeling needs to be generally of lower priority than the needs that will be met by models being developed.

Model evaluations and modeling needs are discussed further in the following sections.

## Model Evaluations

EPA's evaluation program, which assesses models based on a uniform statistical approach recommended by the American Meteorological Society, has covered 29 of the 43 models identified in EPA's Guideline on Air Quality Models. The statistical approach is intended to provide a common basis for model evaluation. The evaluations of models contained in the guideline have covered

- 5 of 9 preferred models. Preferred models are designated by EPA as usable in regulatory decisions without prior special justification.
- 22 of 29 alternate models. Alternate models can be used in regulatory decisions but require special justification and the approval of the EPA regional administrator.
- 2 of 5 screening models. Screening models are used for initial, conservative estimates to determine whether refined modeling is needed.

In addition to these 29 models, EPA has evaluated 5 models not included in its guideline.

Test results of the 34 models indicate that model accuracy varies widely, depending on the type of model (i.e., complex terrain, rural, etc.) and its specific application. Rural and urban models that simulate pollutant sources in simple terrain generally appear more accurate than complex terrain models that simulate conditions where surrounding terrain exceeds smokestack heights. The percentage of over and underestimations ranged from an urban model that underestimated pollutant concentrations by 50 percent to a complex-terrain model that overestimated pollutant concentrations by 1,720 percent. For a listing of models and their evaluation results see appendix I.

Model evaluations yield inconclusive estimates of model reliability since substantially different reliability indications are obtained for the same model evaluated at different sites. For example, one model underestimated by 70 percent at one site and overestimated by 270 percent at another site. Accordingly, while evaluations provide evidence of model reliability at the specific sites evaluated, EPA cannot extrapolate model reliability results to new sites. The closer conditions can be matched between the sites actually evaluated and the simulation sites, the more confident EPA is of the results.

EPA has begun using newer statistical techniques that involve simulation of many conditions using the same data and estimating the distribution of the results. It has concluded that the new techniques yield more reliable estimates. Further evaluation of the new statistical techniques are planned in late fiscal year 1988. From this research, EPA hopes to increase its understanding of model reliability.

## Models Not Evaluated Under Uniform Statistical Approach

In addition, 11 other models identified in EPA's guideline have been evaluated by organizations not affiliated with EPA or by EPA prior to adoption of the uniform approach for model evaluations. These evaluations were not based on consistent data bases (collection of actual measured pollution concentrations, meteorological and other data) or the statistical measures as recommended by the American Meteorological Society. Therefore, there is little basis for comparing the models' performance against one another or against models evaluated by EPA. Appendixes II and III summarize the available information on the independent evaluations of these models.

According to EPA's Chief, Source Receptor Analysis Branch, EPA plans to evaluate one of the 11 models. The model selected is from the reactive plume category, which estimates concentrations of chemically reactive pollutants such as ozone. The four models in this category have not been evaluated because each requires a separate, costly evaluation data base. Urban Airshed, the preferred model in this category, has undergone a relatively comprehensive evaluation, but the evaluation did not employ the uniform statistical approach adopted by EPA.

For the reasons discussed below, EPA has no plans to evaluate the remaining 10 models not uniformly evaluated.

- Two models (one a short-term, long-range transport model and one an offshore coastal dispersion model) have been recently added to EPA's guideline and have not yet been scheduled for model evaluations.
- Four models (three reactive models and one screening model) were not scheduled to be evaluated because similar models had been scheduled for evaluations, and EPA officials did not believe the additional costs to evaluate these models were justified.
- Two visibility models have not been evaluated because the impact of source emissions on visibility is not a controversial or high-priority issue at this time.
- Two industrial complex models have not been uniformly evaluated, and EPA currently has no plans to do so because (1) the data base for the evaluation was found to be incomplete and (2) EPA has been unable to locate another suitable data base.

EPA also has no plans to evaluate any of three additional models in the EPA guideline that have not been evaluated to date.[2]

## EPA Believes Model Estimates Are Useful Despite Uncertainties

EPA believes that while evaluations revealed that models estimate pollutant concentrations with varying degrees of reliability, the models recommended in its guideline can be useful as regulatory and research tools if users recognize their limitations. EPA also believes that their comparative economy of use makes models more desirable than using air quality monitors. Additionally, modeling is the only method available to estimate future pollution concentrations. According to a previous EPA Assistant Administrator of Air and Radiation, modeling is EPA's only viable alternative to the extremely costly process of monitoring pollutant-emitting sources.

# EPA's Modeling Needs

In 1985, EPA updated an early study of modeling needs that had identified 241 modeling needs for regulatory and research applications required to fully implement the Clean Air Act during the next 10 years. The update estimated the cost of meeting the higher priority needs. EPA's current model development program addresses 65 of the identified needs. EPA considers the remaining 176 modeling needs to be generally of lower priority than the needs that will be met by models being developed.

EPA did not, however, identify the actual number of models that would be needed to fulfill the identified modeling needs. According to EPA officials, EPA has not determined the number of models needed because model development generally involves modifications to existing models or development of modules (model components) for use in several models, rather than development of new models.

EPA has not estimated the funds needed to fully address its modeling needs. However, in 1985 EPA did estimate that it would cost $97 million to develop and evaluate models to meet 6 urgent and 70 high-priority needs. Since 1985 some models have been fully or partially funded; however, EPA has not updated the estimate of funds needed to address its urgent and high-priority needs. EPA's annual budget for model development and refinements has averaged about $21 million over the last 4 years. The total amount is not directly focused on particular models as it includes cost for administrative personnel and basic research. The basic

[2]The three models not evaluated are PTPLU-2, LONGZ, and APRAC-3.

research is generally intended to advance air dispersion modeling without focusing on a particular model. According to EPA, as advances in model development become available, models under development and completed models may require more research, development or refinement, and testing.

## Objective, Scope, and Methodology

As agreed with your office, we examined EPA's responses to your request that it

- explain the ranges of reliability of models, identify the uncertainties of each, and indicate the status of testing and
- explain EPA's modeling needs, including the status of the needs being met by models under development.

To obtain information on the first issue, we relied on evaluations done by EPA's consultants. We interviewed EPA officials concerning the interpretation of model evaluation reports and limitations surrounding model evaluation data. To obtain information on the second issue, we reviewed documentation and interviewed EPA officials on modeling needs.

We discussed the information contained in this report with EPA officials and have included their comments where appropriate. However, as you requested, we did not obtain official agency comments on a draft of this report. We conducted our review between November 1987 and June 1988 in accordance with generally accepted government auditing standards.

As arranged with your office, unless you publicly release its contents earlier, we plan no further distribution of this report until 30 days from the date of this letter. At that time, copies of the report will be sent to appropriate congressional committees; the Administrator, EPA; and the Director, Office of Management and Budget.

Major contributors to this report are listed in appendix IV.

Sincerely yours,

Hugh J. Wessinger

Hugh J. Wessinger
Senior Associate Director

# Contents

## Abbreviations

| | |
|---|---|
| EPA | Environmental Protection Agency |
| GAO | General Accounting Office |
| RCED | Resources, Community, and Economic Development Division |

# Models Evaluated by EPA Evaluation Program Using Uniform Statistical Approach

**Table I.1: Models Evaluated by EPA Evaluation Program Using Uniform Statistical Approach**

| Model type | Model | Category of model in EPA guideline | Time interval evaluated (hours) | Number of evaluations | Average percent over and under estimation[a] | Range of over and under estimates (percent) |
|---|---|---|---|---|---|---|
| Rural | MPTER | Preferred | 3 hrs. | 6 | 0 | −30 to +20 |
| Models | | | 24 hrs. | 6 | −20 | −30 to 0 |
| | CRSTER | Preferred | 3 hrs. | 6 | 0 | −30 to +20 |
| | | | 24 hrs. | 6 | −20 | −30 to 0 |
| | MPSDM | Alternate | 3 hrs. | 6 | 20 | −10 to +40 |
| | | | 24 hrs. | 6 | −10 | −40 to +30 |
| | COMPTER | Alternate | 3 hrs. | 2 | 10 | −20 to +30 |
| | | | 24 hrs. | 2 | 20 | +20 to +20 |
| | SCSTER | Alternate | 3 hrs. | 2 | −10 | −10 to −10 |
| | | | 24 hrs. | 2 | −30 | −40 to −30 |
| | Model | Alternate | 3 hrs. | 2 | −40 | −40 to −40 |
| | 3141 | | 24 hrs. | 2 | −40 | −50 to −40 |
| | Model[b] | Alternate | 3 hrs. | 2 | −40 | −40 to −40 |
| | 4141 | | 24 hrs. | 2 | −40 | −50 to −40 |
| | TEM-8A[b] | Alternate | 3 hrs. | 6 | −20 | −50 to −10 |
| | | | 24 hrs. | 6 | −40 | −60 to −10 |
| | MULTI- | Alternate | 3 hrs. | 2 | −20 | −20 to −20 |
| | MAX | | 24 hrs. | 2 | −30 | −40 to −30 |
| | PPSP | Alternate | 3 hrs. | 6 | 260 | +60 to +530 |
| | | | 24 hrs. | 6 | 280 | +40 to +900 |
| Short-Term Long Range Transport | MESOPUFF | Not Listed | 1 hr. | 1 | 40 | N/A |
| | | | 10 hrs. | 1 | 40 | N/A |
| | MSPUFF | Not Listed | 1 hr. | 1 | 250 | N/A |
| | | | 10 hrs. | 1 | 30 | N/A |
| | MESO-PLUME | Not Listed | 1 hr. | 1 | 80 | N/A |
| | | | 10 hrs. | 1 | 40 | N/A |
| | MESOPUFF II | Alternate | 1 hr. | 1 | 20 | N/A |
| | | | 10 hrs. | 1 | 10 | N/A |
| | RTM-II | Alternate | 1 hr. | 1 | −30 | N/A |
| | | | 10 hrs. | 1 | 0 | N/A |
| | RADM | Alternate | 1 hr. | 1 | 230 | N/A |
| | | | 10 hrs. | 1 | 370 | N/A |
| | ARRPA | Alternate | 1 hr. | 1 | 80 | N/A |
| | MTDDIS | Alternate | 10 hrs. | 1 | 40 | N/A |

(continued)

| Model type | Model | Category of model in EPA guideline | Time interval evaluated (hours) | Number of evaluations | Average percent over and under estimation[a] | Range of over and under estimates (percent) |
|---|---|---|---|---|---|---|
| Urban | TEM-8A[b] | Alternate | 3 hrs. | 1 | 110 | N/A |
| | | | 24 hrs. | 1 | 100 | N/A |
| | RAM | Preferred | 3 hrs. | 1 | −40 | N/A |
| | | | 24 hrs. | 1 | −50 | N/A |
| | CDM | Preferred | Annual | 1 | −10 | N/A |
| | AQDM | Alternate | Annual | 1 | 0 | N/A |
| | TCM | Alternate | Annual | 1 | −10 | N/A |
| | ERTAQ | Alternate | Annual | 1 | 20 | N/A |
| Complex Terrain | 4141[b] | Alternate | 3 hrs. | 1 | 320 | N/A |
| | | | 24 hrs. | 1 | 140 | N/A |
| | PLUME 5 | Alternate | 3 hrs. | 1 | 590 | N/A |
| | | | 24 hrs. | 1 | 380 | N/A |
| | SHORTZ | Alternate | 3 hrs. | 1 | 460 | N/A |
| | | | 24 hrs. | 1 | 290 | N/A |
| | IMPACT | Alternate | 1 hr. | 1 | −50 | N/A |
| | RTDM | Screening | 3 hrs. | 1 | 10 | N/A |
| | | | 24 hrs. | 1 | −20 | N/A |
| | COMPLEX-I | Screening | 3 hrs. | 1 | 980 | N/A |
| | | | 24 hrs. | 1 | 810 | N/A |
| | COMPLEX-II | Not Listed | 3 hrs. | 1 | 1720 | N/A |
| | | | 24 hrs. | 1 | 1170 | N/A |
| | COMPLEX-PFM | Not Listed | 3 hrs. | 1 | 590 | N/A |
| | | | 24 hrs. | 1 | 350 | N/A |
| Mobile Source | GMLINE | Alternate | 1 hr. | 3 | −10 | −90 to +50 |
| | | | 8 hrs. | 2 | −30 | −90 to +30 |
| | CALINE-3 | Preferred | 1 hr. | 3 | 60 | −70 to +270 |
| | | | 8 hrs. | 2 | 20 | −80 to +110 |
| | PAL | Alternate | 1 hr. | 3 | 300 | +20 to +790 |
| | | | 8 hrs. | 2 | 150 | −60 to +350 |
| | HIWAY-2 | Alternate | 1 hr. | 3 | 50 | −70 to +190 |
| | | | 8 hrs. | 2 | 30 | −80 to +130 |

[a]Hourly averages are based on 25 highest observed concentrations unpaired in time or location. Annual averages are based on all observed concentrations. Zero percent over/underestimation indicates model's estimations are same as observed concentrations; percentage greater than 0 indicates overestimation; percentage less than 0 indicates underestimation. For example, COMPLEX I model estimations (3 hrs.) are 980 percent greater than observed concentrations.

[b]Model 4141 was evaluated twice—once when used as a rural model and again when used as a complex-terrain model. Also, model TEM-8A was evaluated twice—as a rural model and as a urban model.

N/A = Not Applicable.

## GAO Notes to Appendix I

1. EPA's evaluation reports contain voluminous amounts of data for different data bases, averaging intervals (1-, 3-, 24-hour, and annual; or other intervals), and concentration ranges (all, highest, etc.). The table in appendix I summarizes the uniform evaluation results based on (1) the 25 highest concentrations unpaired in time or location, (highest concentrations are used in regulatory decisions), (2) time intervals corresponding to air quality standard intervals (i.e., SO2 standards are written for 3-and 24-hour intervals), and (3) the average and range across data bases rather than separate values for each data base.

2. Evaluation results indicating the reliability of predictions at particular locations within the modeling area are not included in this table, although EPA has published this data, because regulatory decisions are generally based on highest estimated concentrations independent of location. Location-dependent estimates are less reliable than estimates of 25 highest concentrations shown in appendix I. For example, MPTER estimates of the 25 highest concentrations for a particular monitoring location typically differ from observations by plus or minus 40 percent, compared with differences of approximately 20 percent for estimations of the 25 highest concentrations independent of monitoring location.

3. Evaluation results provide inconclusive indications of a model's reliability since substantially different reliability indications are obtained for the same model evaluated at different sites. While evaluations provide direct evidence of model accuracy only at the specific sites evaluated, the results may also apply in other situations provided that the environmental and physical conditions are reasonably similar. For example, EPA believes that accuracy statements derived for rural models applied to plants with tall stacks in relatively flat terrain would also be applicable to model estimates derived for similar sources located in similar terrains under similar weather patterns. The credibility of extrapolating model accuracy results to new sites cannot be quantified; however, the more closely conditions are matched between the evaluation site and the intended application, the more confident EPA is of the results.

Since EPA initiated its comprehensive model evaluation program, advances have been made in the statistical methodology used to measure or determine model accuracy at a specific site. However, these advances do not solve the problem of extrapolating to new sites. EPA has begun testing these newer statistical techniques using one of the available rural evaluation data bases. Based on these limited tests, EPA has concluded that the resampling technique known as "bootstrapping" yields more accurate confidence intervals for model reliability estimates than was possible when the evaluation program began. A confidence interval is a range around the predicted value which is likely to contain the true value at some specified probability. Bootstrapping involves a simulation of many samples using the same data and estimating the distribution of the results.

Because resampling techniques are relatively new and costly to apply, EPA plans to further evaluate them in fiscal year 1988 using additional data bases. An important potential use of model reliability data is the ability to translate estimates of model reliability into confidence statements regarding the probable success of regulatory decisions. For example, EPA is exploring ways of using the bootstrap outputs to make statements such as "Control of emissions from power plant A will result in 95 percent confidence that pollutant concentrations around this power plant will be in compliance with EPA's ambient standards."

# Models Evaluated Using Various Evaluation Approaches by Organizations Not Affiliated With EPA

| Model type | Model | Category of model in EPA guideline | Summary of available accuracy indicators |
|---|---|---|---|
| Industrial Complex | BLP | Preferred | Generally overestimates highest observed concentrations, usually by less than 30 percent. |
| | ISC | Preferred | Estimates 25 highest 3-hour and 24-hour concentrations within 10 percent of observed levels at one powerplant. |
| Reactive | UAM | Preferred | Accuracy of hourly estimates measured against various data bases ranges from -47 to +29 percent. |
| | RPM-II | Alternate | Mean estimated hourly concentrations differs by less than 4 percent from mean observed concentrations. |
| | PLMSTAR | Alternate | Model estimations agree within 20 percent of observations at key receptor locations. |
| Visibility Model | ERT Visibility | Alternate | Evaluation results not comparable to other models. |
| | PLUVUE II | Alternate | Same results as ERT Visibility above. |
| Short-Term Long-Range Transport | AVACTA II | Alternate | Overestimates maximum concentrations by 100 to 200 percent at one site. |
| Offshore Coastal Dispersion | OOD | Preferred | Estimates 25 highest 1-hour concentrations agrees within 25 percent of observed concentrations. |

# Models Evaluated Using Various Evaluation Approaches by EPA Before Uniform Approach Adopted

| Model type | Model | Category of model in EPA guideline | Summary of available accuracy indicators |
|---|---|---|---|
| Reactive | EKMA | Screening | Estimated peak concentrations generally agree within 30 percent of observed levels provided data on ambient hydrocarbons are available. |
| Screening Models | Valley | Screening | Overestimated 24 hour averages by 200 to 1,500 percent at one site. |

# Major Contributors to This Report

## Resources, Community, and Economic Development Division

Richard L. Hembra, Associate Director (202) 252-0600
William F. McGee, Group Director
Judy K. Pagano, Operations Research Analyst
Cynthia G. Lewis, Secretary

## Norfolk Regional Office

Everett O. Pace, Evaluator-In-Charge
Robert R. Floren, Site Supervisor
Patricia F. Sawyer, Staff Member
Michelle Y. Hall, Typist

Requests for copies of GAO reports should be sent to:

U.S. General Accounting Office
Post Office Box 6015
Gaithersburg, Maryland 20877

Telephone 202-275-6241

The first five copies of each report are free. Additional copies are $2.00 each.

There is a 25% discount on orders for 100 or more copies mailed to a single address.

Orders must be prepaid by cash or by check or money order made out to the Superintendent of Documents.

United States
General Accounting Office
Washington, D.C. 20548

Official Business
Penalty for Private Use $300

First-Class Mail
Postage & Fees Paid
GAO
Permit No. G100