United States
Environmental Protection
Agency

Research and Development

Atmospheric Research and Exposure
Assessment Laboratory
Research Triangle Park NC 27711

EPA/600/S3-90/024 June 1990

EPA

Project Summary

Rocky Mountain Acid Deposition Model Assessment: ARM3 Model Performance Evaluation

Gary E. Moore, Ralph E. Morris, Sharon G. Douglas, and Robert E. Kessler.

The Acid Rain Mountain Mesoscale Model (ARM3), a hybrid acid deposition model, calculates short- and long-term acid deposition (sulfur and nitrogen) and PSD pollutant concentrations (SO_2 and TSP) resulting from emissions of a single source or group of sources at mesoscale distances in the complex terrain of the Rocky Mountain region. The ARM3 consists of two principal components: a mesoscale meteorological model, which includes a diagnostic wind model, and a Lagrangian puff model, which treats transport, dispersion, chemical transformation, and wet and dry deposition. This modeling approach was guided by comments of members of the Western Acid Deposition Task Force (WADTF), who desired a computationally efficient model capable of calculating long-term source-specific acid deposition of nitrogen and sulfur in complex terrain.

Previous reports from the Rocky Mountain Acid Deposition Model Assessment Project reviewed existing mesoscale meteorological and acid deposition models for complex terrain; selected and evaluated candidate meteorological and acid deposition models for complex terrain; and provided the technical formulation and user's guide to the ARM3. Any model intended for use in regulatory decision-making must be evaluated. This report presents the evaluation of the ARM3.

The evaluation of the ARM3 was accomplished in several tasks:

- A diagnostic (or scientific) evaluation examined the formulation of the model by evaluating the parameterization of the major processes of acid deposition in the Rocky Mountains (transport, dispersion, chemical transformation, dry deposition, wet deposition). The diagnostic evaluation was part of the development of the model and is reported in a previous report.
- The wind model component of the ARM3 was evaluated separately. A preliminary evaluation was reported earlier, and a more complete evaluation is included in an appendix of this report.
- The Lagrangian puff model component of the ARM3 (CONDEP) was compared with two EPA-approved Gaussian plume models: ISC and MPTER. This comparison is presented in an appendix.
- The performance of the complete ARM3 modeling system was evaluated using tracer data from the Oklahoma and Savannah River Plant data sets. The model performance was then compared with up to seven other mesoscale air quality simulation models.

In general, the model performance statistics indicate that the performance of the ARM3 is as good or better than the other mesoscale

air quality models. However, care should be taken in the interpretation of these statistical measures. As noted in our analysis, the performance of a model varies depending on which measures of model performance are used: the model's ability to predict observations matched in time and location or the ability to predict peak observations.

Because resources were limited, the evaluation of ARM3 was limited in scope. In particular, because of a lack of an appropriate data base, the ARM3 could not be evaluated for its primary purpose, i.e., calculating source-specific acid deposition impacts in complex terrain. However, the fact that the model performs as good or better than existing mesoscale air quality simulation models indicates that the model shows some promise for use as a regulatory decision-making tool and should be further evaluated and refined.

This Project Summary was developed by EPA's Atmospheric Research and Exposure Assessment Laboratory, Research Triangle Park, NC, to announce key findings of the research project that is fully documented in a separate report of the same title (see Project Report ordering information at back).

Introduction

Acid deposition has recently become an increasing concern in the western United States. Although this problem may not be as acute in the western United States as it is in the eastern United States, it is currently a concern of the public and regulatory agencies because of the high sensitivity of western lakes at high altitudes and the rapid industrial growth expected to occur in certain areas of the West. An example of such an area is the region known as the Overthrust Belt in southwestern Wyoming. Several planned energy-related projects including natural gas sweetening plants and coal-fired power plants may considerably increase emissions of acid precursors in northeastern Utah and northwestern Colorado and significantly affect ecosystems in the sensitive Rocky Mountain areas.

Under the 1977 Clean Air Act, the U.S. Environmental Protection Agency (EPA), along with other federal and state agencies, is mandated to preserve and protect air quality throughout the country. As part of the Prevention of Significant Deterioration (PSD) permitting processes, federal and state agencies are required to evaluate potential impacts of new emission sources. In particular, Section 165 of the Clean Air Act stipulates that, except in specially regulated instances, PSD increments shall not be exceeded and air quality-related values (AQRV's) shall not be adversely affected. Air-quality-related concerns range from near-source plume blight to regional-scale acid deposition problems. By law, the Federal Land Manager of Class I areas has a responsibility to protect air-quality-related values within those areas. New source permits cannot be issued by the EPA or the states when the Federal Manager concludes that adverse impacts on air quality or air-quality-related values will occur. EPA Region VIII contains some 40 Class I areas in the West, including two Indian reservations. Several of the remaining 26 Indian reservations in the region are considering similar designations. State and federal agencies, industries, and environmental groups in the West need accurate data concerning western source-receptor relationships.

To address this problem, EPA Region VIII needs to design an air quality model for application to mesoscale pollutant transport and deposition over the complex terrain of the Rocky Mountain region for transport distances ranging from several km to several hundred km. The EPA recognizes the uncertainties and limitations of currently available air quality models and the need for continued research and development of air quality models applicable over regions of complex terrain. Therefore, the objective of the project reported here is to select, assemble, and evaluate the best air quality models available for application to the Rocky Mountain area.

The primary objective of this project, the EPA Rocky Mountain Acid Deposition Model Assessment Project, is to assemble an air quality/acid deposition model based primarily on models or modules currently available for use by the federal and state agencies in the Rocky Mountain region. The EPA has formed an atmospheric processes subgroup of the Western Atmospheric Deposition Task Force, referred to as WADTF/AP, to develop criteria for model selection and subsequent model evaluations. This group comprises representatives from the National Park Service, U.S. Forest Service, EPA Region VIII, the National Oceanic and Atmospheric Administration, and other federal, state, and private organizations.

On the basis of our review of the modeling needs identified by the WADTF/AP, the specific requirements of the model proposed in this project are as follows:

- The anticipated use of the model is to analyze permit applications by calculating acid deposition impacts at sensitive receptors from specified sources. Thus the primary need is to estimate long-term averages of wet and dry nitrogen and sulfur deposition amounts. However, there is also a need to estimate short-term (3-hour, 24-hour) SO_2 and TSP impacts for PSD increment consumption. Thus the model should be mainly concerned with a mesoscale region within the Rocky Mountain region.

- The modeling system will include a mesoscale meteorological model, which creates wind fields in complex terrain, as a driver for an acid deposition model. Since the primary interest is in longer-term averages, this meteorological model will be required to generate these wind fields in a cost effective manner.

- The acid deposition model will be primarily concerned with estimating incremental acid deposition and ambient concentration impacts from the specified sources only.

A mathematical modeling system for describing the various physical and chemical processes associated with acid deposition must consist of several components or modules. These modules describe processes such as wind transport, chemical reactions, plume rise, and wet/dry deposition. The EPA Rocky Mountain Model Assessment Project has involved the following activities: (1) the review of existing mesoscale models for use in complex terrain; (2) the evaluation of mesoscale models for use in complex terrain; and (3) the assembly of a hybrid complex terrain acid deposition model, the Acid Rain Mountain Mesoscale Model (ARM3) and delivery of the model code to the EPA.

Before any model is used for regulatory decision making it needs to be evaluated and results from the model need to be compared against existing regulatory models. Limited funding for evaluating the ARM3 was available as part of EPA Rocky Mountain Model Assessment Project. This report documents the results of a preliminary evaluation of the ARM3 model performance.

Procedure

The ARM3 was evaluated in several different ways:

- A "scientific (or diagnostic) evaluation" in which each of the major components of the ARM3 are evaluated separately. This was performed as part of the development of the ARM3 and is reported in a previous report.
- A comparison of the ARM3 transport and dispersion module with those of two EPA-recommended steady-state Gaussian plume dispersion models. The ARM3 model predictions were compared with those obtained by MPTER(URBAN) and ISCST(RURAL).
- A separate evaluation of the Diagnostic Wind Model (DWM) component of the ARM3, which also examined the data requirements of the DWM.
- After a review of data sets that can be used to evaluate the ARM3 the complete ARM3 modeling system was evaluated against observed tracer data using data bases from Oklahoma and Savannah River Plant. This evaluation also included a comparison of the ARM3 model performance with the model performance of seven other mesoscale air quality simulation models.

Results and Discussion

Review of Existing Data Bases for Model Evaluation

Most of the data sets reviewed were generally suitable for evaluation of only one or two processes treated by the ARM3. In general, the available evaluation data sets can be divided into data sets capable of evaluating either: (1) the meteorological component; (2) the advection and dispersion component; or (3) the chemical transformation component of the ARM3.

Evaluation of Meteorological Calculations

The evaluation of the meteorological component of the ARM3 focuses mainly on evaluation of the Diagnostic Wind Model (DWM). The evaluation of the DWM should focus on determining how accurately the modeled wind fields reproduce observed trajectories. There are several ways of obtaining these observations:

- Tetroon release and tracking experiments
- High-resolution wind observations
- Tracer experiments under situations where transport dominates over dispersion
- Intensive meteorological measurement programs

Of the four categories of data sets it was elected to use an intensive measurement program to evaluate the DWM.

Evaluation of Transport/Dispersion Calculations

Tracer tests offer the best data base for evaluating the transport and dispersion component of the ARM3. After an extensive review of existing tracer data bases the Oklahoma and Savannah River Plant data sets appeared to be most appropriate for evaluating the ARM3 because: the receptor distances are at the spatial scales (mesoscale) for the intended application of the ARM3; there are both ground-level and elevated tracer releases; and the data sets have been used to evaluate other mesoscale air quality simulation models.

Evaluation of Chemical Transformation and Deposition Calculations

Experiments that can be used to evaluate the chemical transformation module of the ARM3 include: (1) simultaneous release of inert and chemically active compounds; (2) a plume that is isolated from other sources; and (3) measurements that are at downwind distances sufficient for significant plume chemistry to occur. There are currently no data bases available to evaluate the ARM3's ability to calculate source-specific acid deposition. Because of the lack of appropriate data sets and the limited resources available for evaluating the chemistry and deposition components of the ARM3, these components could not be evaluated at this time.

Comparison of ARM3 Model Predictions with Two EPA-Approved Models

ARM3 model predictions were compared with those obtained by two EPA-recommended, steady-state Gaussian plume models: ISCST(RURAL) and MPTER(URBAN). The three models were exercised for a set of 14 meteorological conditions which varied atmospheric stability, wind speed, and mixing height. On average the ARM3 predicted concentrations that lie between the two Gaussian plume models. The ARM3 model predictions were more like those of ISCST(RURAL) and MPTER(URBAN) then ISCST(RURAL) and MPTER(URBAN) were like each other. This result is not surprising since the ARM3 complex terrain dispersion rates lie somewhere between the slow rural dispersion rates of ISCST(RURAL) and the enhanced dispersion rates in MPTER(URBAN). It should be noted that when ISCST and MPTER are both exercised with the RURAL option they produce nearly identical results.

Evaluation of the DWM Using an Extensive Measurement Study

The DWM wind field component of the ARM3 was evaluated by comparing wind fields generated using an intensive measurement program, the South Central Coast Cooperative Aerometric Monitoring Program (SCCCAMP), with those produced by using routine data. In addition, the DWM wind fields were compared to aircraft observations and observations from dual-Doppler radar. In general the evaluation of the DWM wind fields was encouraging, however, fundamental differences in what the observations (point measurements) and predictions (mean flow) represent resulted in large discrepancies between some observations.

Evaluation of the Complete ARM3 Modeling System Against Tracer Data and Comparison of Model Performance with Other Mesoscale Models

The ARM3 was evaluated against tracer data from the Oklahoma (OK) and Savannah River Plant (SRP) data sets and the ARM3 model performance was compared with the model performance of up to seven other mesoscale air quality simulations models: MESOPUFF- I, MESOPLUME, MSPUFF, MESOPUFF-II, MTDDIS, ARRPA, RADM, (Randomwalk Advection and Dispersion Model, not to be confused with the Regional Acid Deposition Model), and RTM-II. The model was evaluated by comparing model predictions with observation for all data (matched by time and location) for the highest model predictions with observations (unmatched by time or location) and for the peak predictions and observations for each time period (matched by time but not location). Based on the models ability to simulate the tracer data, a relative ranking of the mesoscale air quality simulation models was obtained.

Due to the wide range of statistical measures, the relative ranking of model performance is somewhat subjective and based on how one weighs the merits of individual statistical performance measures. Ranking models ability to predict tracer data is particularly difficult because the natural variability of the atmosphere and the inability of a coarse network of meteorological observations to capture this variability. A case can be made for poor performance for all of the models discussed here. However, the purpose of this model ranking is to use the statistical measures to determine whether the ARM3 can predict the observed tracer concentrations as good as, better, or worse than the existing mesoscale air quality simulation models. Since the models behaved differently for the OK and SRP data sets, and the models also varied in their ability to reproduce the maximum observations versus all observations, the models are ranked separately for these categories.

Data Matched by Time and Location—Oklahoma Data

For the OK data set and all tracer data we get the following ranking of skill of the 8 models ability to reproduce the 45-minute tracer observations:

- Ranked 1: MESOPUFF-I and ARM3. Both models exhibit lower bias (48 and 60 percent) and the lowest average absolute error (199 and 126 percent) combined with high correlation coefficients (greater than 0.69). MESOPUFF-I was more accurate with a lower bias and absolute error, but ARM3 had a higher correlation coefficient.

- Ranked 2: RTM-II and MESOPLUME. These two models have an absolute error of about 150 percent, with MESOPLUME displaying a fairly high correlation coefficient (0.593) compared to RTM-II (0.179), however, RTM-II does have a positive correlation coefficient at the 95% confidence limit and the lowest bias (3 percent) of any model.

- Ranked 3: MESOPUFF-II and ARRPA. Both models have absolute errors greater than 200 percent and near zero correlation coefficients. MESOPUFF-II has the second lowest bias but is the only model exhibiting a negative correlation coefficient.

- Ranked 4: MSPUFF and RADM. A case can be made for ranking MSPUFF higher due to its high correlation coefficient (0.759), however, its bias is over 200 percent and absolute error is almost 300 percent. The RADM appears to systematically overpredict with a bias of over 250 percent and an absolute error over 400 percent.

Data Matched by Time and Location—Savannah River Plant Data

- Ranked 1: RTM-II and ARM3. Both models have the lowest bias (7 percent), average absolute error (165 and 171 percent), and, except for RADM, the highest correlation coefficients (0.132 and 0.101) of all the models.

- Ranked 2: MESOPUFF-II, MESO–PLUME, MSPUFF, and MESOPUFF-I. All four of these models have bias that range from 14 to 18 percent, absolute error of about 190 percent, and correlation coefficients that range from 0.010 to 0.096. Of these MSPUFF appears to have the least skill with the highest absolute error and lowest correlation coefficient, but not significantly worse than the other models in this class.

- Ranked 3: RADM. Although exhibiting the highest correlation coefficient of any model (0.264), the inaccuracy of the model (285 percent bias and 433 percent absolute error) indicates that the model contains a systematic tendency towards overprediction.

Data Unmatched by Time or Location—Oklahoma Data

- Ranked 1: MESOPUFF-II. The MESOPUFF-II predicts both the highest 25 and highest five observed tracer observations within 23 percent.

- Ranked 2: MESOPLUME, MESOPUFF-I, and ARM3. The MESOPLUME, MESOPUFF-I, and ARM3 predict the highest 25 and highest five observations to within, respectively, 79, 45, and 55 percent and 79, 76, and 52 percent. The ARM3 exhibits more skill at predicting the highest observations than the other two models in this ranking, but is still not showing as much skill as MESOPUFF-II for this category.

- Ranked 3: RTM-II. The RTM-II is the only model that is almost as accurate as the MESOPUFF-II in replicating the highest observations predicting the highest 25 and five observations to within 32 percent. However, the RTM-II is the only model that underpredicts the highest observed tracer observations. For regulatory purposes it is important to be conservative, i.e., tend towards overprediction of the peak observations; thus, the RTM-II is ranked below some of the other models that are less accurate in this category. Note that based on accuracy alone, the RTM-II would be ranked in the highest category.

- Ranked 4: RADM and MSPUFF. Both models overpredict the 25 and five highest observations by over a factor of two and are the least accurate of the models examined.

Data Unmatched by Time or Location—Savannah River Plant Data

- Ranked 1: MESOPUFF-II. The MESOPUFF-II predicts the 25 and five highest observations for the SRP data set to within 8 and 12 percent.

- Ranked 2: RTM-II and ARM3. These two models predict the 25 highest observations to within 4 and 2 percent, respectively. However, the RTM-II and ARM3 underpredict the five highest observations by, respectively, 9 and 15 percent. Despite the fact that the RTM-II and ARM3 are more accurate in predicting the peak observations than the MESOPUFF-II, the model attribute to be conservative in predicting peak observations is important enough to rank the models below MESOPUFF-II in this category.

- Ranked 3: MESOPUFF-I, MESOPLUME, and MSPUFF. The MESOPUFF-I, ARRPA, and MSPUFF predict the 25 and five highest observations to within, respectively, 33 to 44 percent and 38 to 58 percent.

- Ranked 4: RADM. This model overpredicts the 25 and five highest observations by about a factor of 4 and 6 respectively.

Data Matched by Time but Not Location—Oklahoma Data

- Ranked 1: MESOPUFF-II, MESOPUFF-I, and ARM3. All three models predict the average of the maximum observation for each sampling interval to within a factor of two and correlate well (0.64 to 0.84).

- Ranked 2: MESOPLUME and ARRPA. These two models predict the average of the maximum observations for each sampling interval by a little over a factor of two and both exhibit correlation coefficients of about 0.6.

- Ranked 3: RTM-II. The RTM-II underpredicts the average of the maximum observations for each experiment by a little over a factor of two and has the lowest correlation coefficient (0.354) of any model for this category.

- Ranked 4: MSPUFF and RADM. The MSPUFF and RADM overpredict the average of the maximum observations by a factor of 3.3 and 5.2, respectively.

Data Matched by Time but Not Location—Savannah River Plant Data

- Ranked 1: RTM-II. The RTM-II overpredicts the average of the maximum observations by nine percent and has the second highest correlation coefficient (0.280) in this category.

- Ranked 2: MESOPUFF-II and ARM3. The MESOPUFF-II predicts the average of the maximum observations to within 5 percent but is the only model that exhibits a negative correlation coefficient in this category (-0.074). The ARM3 predicts the average of the maximum observations to within 11 percent and has a slight positive correlation coefficient (0.084).

- Ranked 3: MESOPLUME and MESOPUFF-I. These two models both predict the average of the maximum observations to within about 30 percent and both have slightly positive correlation coefficients of less than 0.02.

- Ranked 4: RADM. RADM overpredicts the average of the maximum observations by almost a factor of five.

Final Ranking of the Models

In order to obtain an overall ranking of the eight mesoscale transport models ability to reproduce the tracer observations in the OK and SRP data bases we combine the above ranking in each category into a final model ranking. For each of the six categories above a model receives a score of four if it is ranked first, three if ranked second, two if ranked third, and one if ranked fourth. An overall model ranking is then obtained by adding up the models score in each of the six categories. Note that this is somewhat a subjective and arbitrary methodology for ranking the models. Some may want to score the models ability to replicate all observations (matched by time and location) higher than the categories involving maximum observation. However, most EPA-approved models are evaluated by their ability to reproduce the maximum observations, therefore we feel that this methodology is a fair and as objective as possible and is only intended to give a relative score for obtaining a relative ranking of the performance of the 8 models. Based on the above scoring system, in which a maximum 24 points is possible, we get the final model ranking:

Model	Score
ARM3	21
MESOPUFF-II	20
RTM-II	18 (tied)
MESOPUFF-I	18 (tied)
MESOPLUME	16
ARRPA	8 (out of a possible of 12 points)
MSPUFF	11
RADM	7

The best performing models are the ARM3 and MESOPUFF-II. The ARM3 is more accurate in predicting all observations (matched by time and location), whereas, the MESOPUFF-II is slightly better at predicting the very highest observations (unmatched by time or location).

The second best performing models in this study were the RTM-II and MESOPUFF-I. The RTM-II tended to match the observations from the SRP data set better, while, the MESOPUFF-I predicted a better match with the observations from the OK data set.

The next models in the ranking were the MESOPLUME and ARRPA, although the ARRPA was only exercised for the OK data base. The MSPUFF was ranked next and illustrated some skill in predicting the observations from the SRP data set but greatly overpredicted the observations from the OK data set. The worst performing model was the RADM which tended towards systematic overprediction.

Conclusions and Recommendations

In general, the model performance statistical results indicated that the performance of the ARM3 was as good or better than the other mesoscale air quality simulation models. However, care should be taken in the interpretation of these statistical measures. The performance of the model varies depending on which statistical measures of model performance is examined: ability to predict peak observations; or ability to predict the observations matched by time and location.

Although the model performance of the ARM3 was comparable or better than the other mesoscale models, further evaluation studies should be conducted. In particular the ARM3 should be evaluated for its primary intended purpose: the calculation of source specific acidic (sulfur and nitrogen) deposition and PSD pollutant concentrations at mesoscale distances in complex terrain. Furthermore, the ARM3 should be subjected to an extensive sensitivity analysis whose results should be used to improve the model.

G. E. Moore, R. E. Morris, S. G. Douglas and R. C. Kessler are with Systems Applications, Inc., San Rafael, California 94903.
***Alan H. Huber** is the EPA Project Officer (see below).*
The complete report, entitled "Rocky Mountain Acid Deposition Model Assessment: ARM3 Model Performance Evaluation," (Order No. PB90-188 871/AS; Cost: $39.00, subject to change) will be available only from:
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
Telephone: 703-487-4650
The EPA Project Officer can be contacted at:
Atmospheric Research and Exposure Assessment Laboratory
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

United States Environmental Protection Agency

Center for Environmental Research Information
Cincinnati OH 45268

Official Business
Penalty for Private Use $300

EPA/600/S3-90/024