

MODEL EVALUATION RESULTS FOR AERMOD

draft document

December 17, 1998

R. J. Paine, ENSR Corporation

R. F. Lee, Trinity Consultants

Roger Brode, Pacific Environmental Services

R. B. Wilson, U.S. Environmental Protection Agency, Region X

A. J. Cimorelli, U.S. Environmental Protection Agency, Region III

S.G. Perry¹, Atmospheric Sciences Modeling Division,
Air Resources Laboratory, EPA/ NOAA

J. C. Weil, Cooperative Institute for Research in Environmental Sciences,
University of Colorado

A. Venkatram, College of Engineering, University of California at Riverside

W. D. Peters, U.S. Environmental Protection Agency, OAQPS

¹ On assignment to the Atmospheric Research and Exposure Assessment Laboratory, U.S. Environmental Protection Agency.

TABLE OF CONTENTS

EXECUTIVE SUMMARY.....	i
INTRODUCTION	1
MODEL DESCRIPTION	2
EVALUATION DATABASES	6
Developmental Evaluation	6
Independent Evaluation	7
EVALUATION PROCEDURES	12
ANALYSIS AND RESULTS OF THE DEVELOPMENTAL EVALUATION	15
ANALYSIS AND RESULTS OF THE INDEPENDENT EVALUATION	18
SUMMARY AND CONCLUSIONS	21
DISCLAIMER	21

LIST OF TABLES

Table 1 Summary of AERMOD Evaluation Results	iv
Table 2 Comparison of Dispersion Model Features:	3
Table 3 Summary of AERMOD Evaluation Databases	11

EXECUTIVE SUMMARY

In the 1980s, several reviews of dispersion models used in regulatory applications reported that these techniques were generally many years behind the state of the art and produced predictions that did not agree well with observations (see Smith (1984), Weil (1985), and Hayes and Moore (1986)). One of the results of a workshop on the parameterization of the planetary boundary layer held by the United States Environmental Protection Agency (U.S. EPA), in conjunction with the American Meteorological Society (AMS) was the formation of the AMS/EPA Regulatory Model Improvement Committee (AERMIC), which was to build upon earlier modeling developments to provide a state-of-the-art dispersion model for routine regulatory applications. The resulting model, AERMOD (Cimorelli et al., 1998), is the subject of this model evaluation report.

The intended purpose of AERMOD is to replace ISCST3 (U.S. EPA, 1995a,b), the current widely-used short-range dispersion model recommended by the United States Environmental Protection Agency (EPA). AERMOD represents an advance in the formulation of a steady-state, Gaussian plume model. It is apparent that AERMOD has an advantage over ISCST3 when the various scientific components are compared. However, to be acceptable as a guideline regulatory model, AERMOD must also perform at least as well as or better than the existing guideline models.

One important aspect of the performance evaluation of AERMOD is the availability of two major components of the model evaluation process: one set of data bases was used during the model formulation and development, and another set was used for final (“independent”) evaluation with limited changes made to the model. Taken together, these studies involve four short-term tracer studies and six conventional long-term SO₂ monitoring data bases in a variety of settings. The purpose of these studies is to be sure that AERMOD has been tested in the various types of environments for which it will be used. Compared with past models that have been proposed for regulatory use, AERMOD has been subjected to a significantly greater degree of testing with these two groups of evaluation data bases, as well as other independent tests conducted by the public.

The **developmental evaluation**, conducted during the model formulation and initial testing, involved the following five data sets:

- Prairie Grass, a rural flat terrain, near-surface tracer release experiment in Nebraska with sampler measurements out to 800 meters;
- Kincaid, a rural flat terrain, tall stack tracer release experiment in Illinois with sampler measurements out to 50 kilometers;
- the Kincaid site with several months of SO₂ monitors at 30 locations;
- Indianapolis, an urban flat terrain tall stack tracer release experiment in Indiana with sampler measurements out to 12 kilometers; and
- Lovett, a rural complex terrain, tall stack SO₂ network in New York State with sampler measurements mostly on a nearby hill.

A subsequent “**independent**” **evaluation** initially involved the following three data sets, each with one full year of data for a limited number of fixed SO₂ monitoring sites in the vicinity of an electric utility source:

- Baldwin, a rural flat terrain site in Illinois with ten monitors ranging in distance from 2 to 10 km from the source location;
- Clifty Creek, a rural moderately hilly site (with the source in the Ohio River Valley, but the stack top well above the surrounding plateau) in Indiana with six monitors ranging in distance from 4 to 15 km from the source location; and
- Martins Creek, a rural complex terrain site near the Pennsylvania/New Jersey border with seven monitors ranging in distance from 2.5 to 8 km from the nearest source location.

After initial evaluation results for these three data bases were examined by a peer review panel, AERMIC received recommendations from the panel that additional data bases involving complex terrain features should be examined. In addition, there was a concern expressed about an apparent tendency toward

underprediction by AERMOD for one of the independent databases. As a result of these recommendations, AERMIC made selected minor revisions to AERMOD involving the complex terrain dispersion algorithms and conducted **additional evaluations** on two other independent data sets:

- Westvaco, a rural complex terrain site at the Maryland/Virginia border with eleven monitors ranging in distance to 3 km from the paper mill source (for a full 1-year data set); and
- Tracy, a rural complex terrain site east of Reno, NV, which was the site of EPA's operational facility tracer study in complex terrain as part of the development of CTDMPPLUS.

The results for these two additional databases looked favorable, so the revised AERMOD was re-run for all ten developmental and independent evaluation databases, resulting in mostly small changes to the results for the databases already evaluated. Evaluation statistics presented in this report include quantile-quantile (Q-Q) plots, residual plots, and robust highest concentration results. A summary of the results for the Robust Highest Concentration (RHC) statistic is presented in Table 1. These results show that AERMOD is nearly unbiased, on average, across all averaging times. For 1-hour averages (the tracer databases), the ratio of predicted to observed values for the RHC ranges from 0.76 to 1.20, with a geometric mean of 0.96. For 3-hour averages, the RHC predicted to observed ratio ranges from 1.00 to 1.31, with a geometric mean of 1.11. The same ratio for 24-hour averages ranges from 0.72 to 1.72, with a geometric mean of 1.06. Annual average statistics are less reliable because background concentrations, which are removed from the measured values in many cases, are uncertain and approach the value of the source-caused impact. The AERMOD RHC predicted to observed ratio for annual averages ranges from 0.30 to 1.64, with a geometric mean of 0.73. For all averaging times in general and in most cases, AERMOD's model performance was better than that of ISCST3.

As a result of the superior technical formulation of AERMOD and its better evaluation performance relative to ISCST3, the AERMIC committee concludes that AERMOD can be justified as a replacement for ISCST3 for regulatory modeling applications.

Table 1 Summary of AERMOD Evaluation Results

Data Base	Ratio of Modeled/Observed Robust Highest Concentrations*
Prairie Grass (SO ₂) Flat, grassy field (Nebraska, USA)	AERMOD: 0.87 (1-hr avg) ISCST3: 1.50 (1-hr avg)
Kincaid (SF ₆) Flat, rural (Illinois, USA)	AERMOD: 0.76 (1-hr avg) ISCST3: 0.68 (1-hr avg)
Kincaid: (SO ₂) Flat, rural (Illinois, USA)	AERMOD: 1.01 (3-hr avg) ISCST3: 0.56 (3-hr avg) AERMOD: 0.97 (24-hr avg) ISCST3: 0.45 (24-hr avg) AERMOD: 0.30 (annual peak) ISCST3: 0.14 (annual peak)
Baldwin (SO ₂): Flat, rural (Illinois, USA)	AERMOD: 1.31 (3-hr avg) ISCST3: 1.48 (3-hr avg) HPDM: 1.06 (3-hr avg) AERMOD: 1.02 (24-hr avg) ISCST3: 1.13 (24-hr avg) HPDM: 1.02 (24-hr avg) AERMOD: 0.97 (annual peak) ISCST3: 0.63 (annual peak) HPDM: 1.15 (annual peak)
Indianapolis (SF ₆) Flat, urban (Indiana, USA)	AERMOD: 1.20 (1-hr avg) ISCST3: 1.30 (1-hr avg)
Clifty Creek (SO ₂) Moderately hilly terrain, rural (Indiana, USA)	AERMOD: 1.25 (3-hr avg) ISCST3: 0.98 (3-hr avg) HPDM: 1.33 (3-hr avg) AERMOD: 0.72 (24-hr avg) ISCST3: 0.67 (24-hr avg) HPDM: 1.46 (24-hr avg) AERMOD: 0.54 (annual peak) ISCST3: 0.31 (annual peak) HPDM: 0.96 (annual peak)

Data Base	Ratio of Modeled/Observed Robust Highest Concentrations*
Tracy (SF ₆): Hilly terrain, rural (Nevada, USA)	AERMOD: 1.07 (1-hr avg) CTDMPLUS: 0.77 (1-hr avg)
Martins Creek (SO ₂): Hilly terrain, rural (Pennsylvania/New Jersey, USA)	AERMOD: 1.06 (3-hr avg) CTDMPLUS: 4.80 (3-hr avg) ISCST3: 7.25 (3-hr avg) RTDM: 3.33 (3-hr avg) AERMOD: 1.72 (24-hr avg) CTDMPLUS: 5.56 (24-hr avg) ISCST3: 8.88 (24-hr avg) RTDM: 3.56 (24-hr avg) AERMOD: 0.74 (annual peak) CTDMPLUS: 2.19 (annual peak) ISCST3: 3.37 (annual peak) RTDM: 1.32 (annual peak)
Lovett (SO ₂)	AERMOD: 1.00 (3-hr avg) CTDMPLUS: 2.36 (3-hr avg) ISCST3: 8.20 (3-hr avg) AERMOD: 1.00 (24-hr avg) CTDMPLUS: 2.02 (24-hr avg) ISCST3: 9.11 (24-hr avg) AERMOD: 0.78 (annual peak) CTDMPLUS: 1.71 (annual peak) ISCST3: 7.49 (annual peak)
Westvaco (SO ₂): Hilly terrain, rural (Maryland/Virginia, USA)	AERMOD: 1.08 (3-hr avg) CTDMPLUS: 2.14 (3-hr avg) ISCST3: 8.50 (3-hr avg, estimated) AERMOD: 1.14 (24-hr avg) CTDMPLUS: 1.54 (24-hr avg) ISCST3: N/A (24-hr avg) AERMOD: 1.64 (annual peak) CTDMPLUS: 0.93 (annual peak)

*Notes:

The Robust Highest Concentration (RHC) is a statistical estimator for the highest concentration. It is determined from a tail exponential fit to the high end of the frequency distribution of observed and predicted values. The number of points used for the fit is arbitrary, but usually ranges between 10 and 25.

The estimated ISCST3 result for Westvaco is derived from the EPA Complex Terrain Model Development study (Strimaitis et al., 1987) in which several models, including CTDMPPLUS and COMPLEX-I (now part of ISCST3), were evaluated.

INTRODUCTION

In 1991, the American Meteorological Society and the United States Environmental Protection Agency initiated a formal collaboration with the designed goal of introducing recent scientific advances in our understanding of the planetary boundary layer into applied dispersion models. A working group (AMS/EPA Regulatory **Model Improvement Committee, AERMIC**) of three AMS scientists and four EPA meteorologists was formed to facilitate this collaborative effort. The focus of the AERMIC group has been on applied models designed for estimating near-field impacts from industrial source types. The primary products of the ongoing AERMIC development work are the AERMOD (**AERMIC Model**) dispersion model, the AERMET meteorological preprocessor, and the AERMAP terrain preprocessor.

The development of a new model is generally dependent not only on published research in atmospheric diffusion, but also on model development work that has gone on before. This is certainly true with AERMOD. AERMOD may become one of the first “new generation plume models” to achieve regulatory status in the U.S. A “new generation plume model” is not simply a variation on the traditional Gaussian plume model, but, instead, takes advantage of more recent research on turbulence and diffusion in the atmosphere. Other models in this category include PPSP (Weil and Brower, 1984), HPDM (Hanna and Paine, 1989), TUPOS (Turner et al., 1986), CTDMPLUS (Perry et al., 1989), and, more recently, ADMS (developed in the United Kingdom; see Carruthers et al., 1992) and OML (developed in Denmark; see Olesen, 1991). AERMIC members were involved in the development of three of these models, PPSP, CTDMPLUS and HPDM

As AERMOD was developed, algorithms used by these models were considered along with other published approaches. In addition, the developers of OML met with AERMIC to discuss their experiences. As with most technological developments, much credit is due to the developers of earlier models, since AERMOD was built on the knowledge and experience gained from their development .

The evaluation of AERMOD has been accomplished in two phases. The first phase, the “developmental evaluation,” was performed concurrently with the development of the model. As each feature of the model was added, a relevant portion of the developmental evaluation was repeated with five databases to identify any problems that might have been introduced at that stage of the model’s development. Because of the possibility that the model may have been inadvertently biased to fit particular characteristics of the

developmental databases used, a second phase, the “independent evaluation,” was conducted using three additional data sets. This second evaluation was conducted with a minimum of model changes (only those required to fix run-time errors or to correctly implement the model formulation).

The results of both evaluations were submitted to a peer review panel assembled by the US EPA. The panel recommended that additional databases be evaluated for complex terrain impacts, and a concern was raised about an underprediction for one averaging time for one of the complex terrain data sets. To address these concerns AERMIC made a small number of changes to the model formulation (mostly focused upon the complex terrain dispersion algorithms), added two additional data sets (used previously in the development of CTDMPPLUS) to the independent evaluation, and then re-ran the evaluation on the five developmental and the five independent data sets. The results of this final series of evaluation runs are described in this report.

MODEL DESCRIPTION

The AERMOD modeling system is composed of one main model (AERMOD) and two preprocessors—a meteorological preprocessor (AERMET) and a terrain preprocessor (AERMAP). AERMET calculates hourly boundary layer parameters for use by AERMOD, including friction velocity, Monin-Obukhov length, convective velocity scale, temperature scale, convective boundary layer (CBL) height, stable boundary layer (SBL) height, and surface heat flux. In addition, AERMET passes all observed meteorological parameters to AERMOD including wind direction and speed (at multiple heights, if available), temperature, and, if available, measured turbulence. AERMOD uses this information to calculate concentrations in a manner that accounts for changes in dispersion rate with height, allows for a non-Gaussian plume in convective conditions, and accounts for a dispersion rate that is a continuous function of meteorology. In contrast, ISCST3 assumes that the dispersion rate is constant with height, that the plume is always Gaussian in form, and is based on discrete dispersion (stability) categories that were developed in the 1960's and can result in jumps in calculated concentrations with small changes in meteorology. AERMAP prepares terrain data for use by AERMOD in complex terrain situations. This allows AERMOD to account for terrain using a simplification of the procedure used in the CTDMPLUS model (Perry, et al., 1989). Table 2 summarizes the differences between AERMOD and the current regulatory model, ISCST3. Detailed descriptions of the formulations are presented in Perry, et al. (1998).

Table 2 Comparison of Dispersion Model Features:

AERMOD vs. ISCST3

Feature	ISCST3	AERMOD	Comments
Types of sources modeled	Point, area, and volume sources	Same as ISCST3	Models are comparable
Plume Rise	Uses Briggs equations with stack-top wind speed and vertical temperature gradient	In stable conditions, it uses Briggs equations with winds and temperature gradient at stack top and half-way to final plume rise; in convective conditions, plume rise is superposed on the displacements caused by random convective velocities	AERMOD is better because in stable conditions it factors in wind and temperature changes above stack top, and in unstable conditions it accounts for convective updrafts and downdrafts
Meteorological Data Input	One level of data accepted	An arbitrarily large number of data levels can be accommodated	AERMOD can adapt multiple levels of data to various stack and plume heights
Profiling Meteorological Data	Only wind speed is profiled	AERMOD creates profiles of wind, temperature, and turbulence, using all available measurement levels	AERMOD is much improved over ISCST3 in this area
Use of Meteorological Data in Plume Dispersion	Stack-top variables for all downwind distances	Variables measured throughout the plume depth (averaged from plume centerline to 2.15 sigma-z below centerline; changes with downwind distance)	AERMOD treatment is an advancement over that of ISCST3; and accounts for meteorological data throughout the plume depth
Plume Dispersion: General Treatment	Gaussian treatment in horizontal and vertical	Gaussian treatment in horizontal and in vertical for stable conditions; non-Gaussian probability density function in vertical for unstable conditions	AERMOD's unstable treatment of vertical dispersion is a more accurate portrayal of actual conditions

Feature	ISCST3	AERMOD	Comments
Urban Treatment	Urban option either on or off; no other specification available; all sources must be modeled either rural or urban	City size and population are specified, so treatment can consider a variety of urban conditions; sources can individually be modeled rural or urban	AERMOD provides variable urban treatment as a function of city population, and can selectively model sources as rural or urban
Characterization of the Modeling Domain Surface Characteristics	Choice of rural or urban	Selection by direction and month of roughness length, albedo, and Bowen ratio, providing much user flexibility	AERMOD provides the user with considerably more options in the selection of the surface characteristics
Boundary Layer Parameters	Wind speed, mixing height, and stability class	Friction velocity, Monin-Obukhov length, convective velocity scale, mechanical and convective mixing height, sensible heat flux	AERMOD provides parameters required for use with up-to-date planetary boundary layer (PBL) parameterizations; ISCST3 does not
Mixed Layer Height	Holzworth scheme; uses interpolation based upon maximum afternoon mixing height	Has convective and mechanical mixed layer height; convective height based upon hourly accumulation of sensible heat flux	AERMOD's formulation is more advanced than that of ISCST3, includes a mechanical component, and in using hourly input data, provides a more realistic sequence of the diurnal mixing height changes
Terrain Depiction	Elevation at each receptor point	Controlling hill elevation <u>and</u> point elevation at each receptor, obtained from special terrain pre-processor (AERMAP) that uses digital elevation model (DEM)	AERMOD's terrain pre-processor provides information for advanced critical dividing streamline height algorithms and uses digital data to obtain receptor

		data	elevations
Feature	ISCST3	AERMOD	Comments
Plume Dispersion: Plume Growth Rates	Based upon 6 discrete stability classes only; dispersion curves (Pasquill-Gifford) are based upon surface release experiments (Prairie Grass)	Uses profiles of vertical and horizontal turbulence (from measurements and/or PBL theory); variable with height; uses continuous growth functions rather than a discrete (stability-based) formulation	Use of turbulence-based plume growth with height dependence rather than that based upon stability class provides AERMOD with a substantial advancement over the ISCST3 treatment
Plume Interaction with Mixing Lid: convective conditions	If plume centerline is above lid, a zero ground-level concentration is assumed	Three plume components are considered: a “direct” plume that is advected to the ground in a downdraft, an “indirect” plume caught in an updraft that reaches the lid and eventually is brought to the ground, and a plume that penetrates the mixing lid and disperses more slowly in the stable layer aloft (and which can re-enter the mixed layer and disperse to the ground)	The AERMOD treatment avoids potential underpredictions suffered by ISCST3 due to its “all or nothing” treatment of the plume; AERMOD’s use of convective updrafts and downdrafts in a probability density function approach is a significant advancement over ISCST3
Plume Interaction with Mixing Lid: stable conditions	The mixing lid is ignored (assumed to be infinitely high)	A mechanically mixed layer near the ground is considered. Plume reflection from an elevated lid is used to account for the effects of sharply reduced turbulence aloft..	AERMOD’s use of a mechanically mixed layer is an advancement over the very simplistic ISCST3 approach

EVALUATION DATABASES

Developmental Evaluation

AERMOD is being developed as a regulatory tool and, thus, is intended to handle a variety of pollutant source types (including surface and buoyant elevated sources) in a wide variety of modeling situations (including rural, urban, flat terrain and complex terrain). With this in mind, data from five diverse field studies were selected for the developmental evaluation. A summary of the features of each database is provided in Table 3.

The **Prairie Grass** study (Barad, 1958; Haugen, 1959) used a near-surface, non-buoyant tracer release in a flat rural area. The Prairie Grass study involved a tracer of SO₂ released at 0.46 m above the surface. Surface sampling arrays (arcs) were positioned from 50 m to 800 m downwind (see Figure 1). Meteorological data included 2-m wind speed, sigma-theta, and delta T (2 m - 16 m). Other surface parameters, including friction velocity, Monin-Obukhov length, and σ_y were estimated. A total of 44 sampling periods were used, including both convective and stable conditions.

The **Kincaid SF₆** study (Liu and Moore, 1984; Bowne, et al., 1983) consisted of an elevated, buoyant tracer release in a flat rural area (see Figure 2 for an area map). Six weeks of intensive study was conducted during the spring and summer of 1980 and 1981. During this study, approximately 200 monitors were placed in arcs from about 500 m to 50 km downwind of the single 187-m stack (an example of the tracer sampler deployment is shown in Figure 3). Meteorological data included wind speed and direction, u-v-w winds, delta T from a 100-m instrumented tower, delta T from a 10-m instrumented tower, and nearby National Weather Service (NWS) data. Estimates of lateral plume spread (σ_y) are available from the sampling arcs.

The **Indianapolis** study (Murray and Bowne, 1988) consisted of an elevated, buoyant tracer (SF₆) released in an urban area (see Figures 4 and 5). The site is a flat-terrain, urban to suburban area with a single 84-m stack. Data are available for approximately a four- to five-week period with 177 monitors in arcs from 250 m to 12 km downwind. Meteorological data included wind speed and direction, σ_θ on a 94-meter tower; and

wind speed, ΔT (2m - 10m) and other supporting surface data at three other towers. Observed plume rise and estimates of plume σ_y are also available from the database.

The **Kincaid SO₂** study (Liu and Moore, 1984; Bowne, et al., 1983) consisted of a buoyant, continuous release of SO₂ from a 187-m stack. The site is in a rural area in flat terrain. The study includes about six months of data between April 1980 and June 1981. There were 30 SO₂ monitoring stations from about 2 km to 20 km downwind of the stack (see Figure 6). The meteorological data are the same as in the Kincaid tracer study.

The **Lovett Power Plant** study (Paumier et al., 1992) consisted of a buoyant, continuous release of SO₂ from a 145-m tall stack. The site is located in complex terrain in a rural area (see Figure 7). The data spans one year from December 1987 through December 1988. Data were collected from 12 monitoring sites (10 on terrain, 2 as background) located about 2 to 3 km from the plant. The important terrain features rise approximately 250 m to 330 m above stack base. The monitors on terrain are generally about 2 to 3 km downwind from the stack. Meteorological data include winds, turbulence, and delta T from a tower instrumented at 10 m, 50 m, and 100 m. NWS surface data were obtained from a station 45 km away.

Independent Evaluation

PES (1998a) performed the independent evaluation of AERMOD using the first three databases described below. Results for two additional databases are described by PES (1998b).

The **Baldwin Power Plant** is located in a flat terrain setting of southwestern Illinois (see Figure 8). The terrain slopes gently upward east of the facility. Three 184-meter stacks aligned approximately north-south with a horizontal spacing of about 100 meters between each stack were modeled for this evaluation (Hanna and Chang, 1993). Building widths and heights developed for the Hanna and Chang study were used in this evaluation. Although the stacks are slightly less than the Good Engineering Practice height, sufficient momentum rise was present to avoid building downwash effects under most conditions. There were 10 SO₂ monitors surrounding the facility ranging in distance from two to ten kilometers, as shown in Figure 8.

On-site meteorological data from the Baldwin field study covered the period from April 1, 1982 through March 31, 1983 and consisted of hourly wind speed, wind direction, and temperature measurements taken at 10 meters and hourly wind speed and wind direction at 100 meters.. Both the National Weather Service (NWS) upper air soundings and hourly surface observations were provided by EPA. Upper air sounding data for Salem, Illinois were provided in a format compatible with SIGPRO, the meteorological preprocessor for the Hybrid Plume Dispersion Model (HPDM), and reformatted to the input format requirements for AERMET quality assurance.

The **Clifty Creek Power Plant** is located in southern Indiana on the north side of the Ohio River (see Figure 9). The area immediately north of the facility is characterized by cliffs rising about 115 meters above the river and intersected by creek valleys.. The tops of the stacks extend about 80 to 100 meters above the top of the cliffs on both sides of the river. Three 208-meter stacks meters were modeled in this evaluation. This same database was used in a major EPA-funded evaluation of rural air quality dispersion models in the early 1980s (Mills, 1979; TRC, 1982).

There were six SO₂ monitors on the surrounding terrain. One was located in the river valley approximately eight kilometers to the east (upriver) at about the same elevation as the sources. Another was located three kilometers south of the facility on Liberty Ridge near the meteorological tower on the south side of the Ohio River and about 110 meters above stack base. The remaining four monitors were located on the terrain north and northeast of the facility at about 125 meters above the base of the stacks, ranging in distance from four to 15 kilometers.

Meteorological data from the Clifty Creek field study covered the two year period from January 1, 1975 through December 31, 1976, although only the data from 1975 were used in this evaluation. The on-site meteorological data were recorded on an instrumented meteorological tower three kilometers south of the facility (across the river in Kentucky) on Liberty Ridge.

The **Martins Creek Steam Electric Station** (MCSES) is located on the Pennsylvania/New Jersey border, approximately 30 kilometers northeast of Allentown, PA and 95 kilometers north of Philadelphia, PA on the Delaware River. In addition to the MCSES, there are three other major, more distant facilities that contributed to the monitored SO₂ concentrations. These facilities are the Metropolitan Edison (ED) Portland Station, Hoffman-LaRoche (HL), and the Warren County Resource Recovery Facility (WCRRF). Stack heights among the facilities range in height from 60 to 183 meters. The area is characterized by complex terrain rising above the stacks toward the southeast. Figure 10 shows the facilities and surrounding topographic features.

The seven SO₂ monitors used in this evaluation (TRC, 1994) were on located Scotts Mountain, which is about 2.5 - 8 kilometers southeast of the Martins Creek facility. The monitors were about 90-120 meters above the top of the Martins Creek sources. The hourly background concentration was removed from the observed concentrations and was defined as the lowest value monitored value each hour at any of the monitors. An eighth SO₂ monitor was located about six kilometers northeast of the facility for purposes of estimating background concentrations. Since this monitor was at an elevation below the Martins Creek stack heights, it was not included in the evaluation.

On-site meteorological data for the Martins Creek station covered the period from May 1, 1992 through May 19, 1993. Hourly temperature, wind speed, wind direction, and σ_A at 10 meters were recorded from an instrumented tower located in a flat area approximately 2.5 kilometers west of the Martins Creek power generation station. In addition, hourly multi-level wind measurements were taken by a SODAR located approximately three kilometers southwest of the Martins Creek station.

The **Westvaco** Corporation's pulp and paper mill in Luke, Maryland is located in a complex terrain setting in the Potomac River valley in western Maryland (Strimaitis et al., 1987). Figure 11 shows the location of the facility, topographic features, and locations of the ambient monitors and on-site meteorological data towers. A single 190-m stack was modeled for this evaluation.

There were 11 SO₂ monitors surrounding the facility, with eight monitors well above stack top on the high terrain east and south of the mill at a distance of 800 - 1500 meters. Two monitors, one on Luke Hill and the

other at Stony Run, were at elevations approximately equal to stack top and were 900 meters NNW and 3300 meters NE, respectively. One monitor at Bloomington was located 1500 meters northwest at about stack base elevation.

Hourly meteorological data were collected between December 1980 and November 1991 at three instrumented towers: the 100-meter Beryl tower in the river valley about 400 meters southwest of the facility; the 30-meter Luke Hill tower on a ridge 900 meters north-northwest of the facility; and the 100-meter Met tower 900 meters east-southeast of the facility on a ridge across the river.

The **Tracy Power Plant** is located 27 kilometers east of Reno, Nevada in the Truckee River valley with mountainous terrain on all sides (DiCristofaro et al., 1985). Figures 12 and 13 show the location of the facility, topographic features, locations of the ambient monitors, and on-site meteorological data tower. A field tracer study was conducted at the power plant in August 1984 with SF₆ being released through the 91-m stack servicing unit 3. A total of 128 hours of data were collected over 14 experimental periods. Most of the hours were during stable atmospheric conditions.

There were 106 monitors used in this evaluation. Most of the monitors were primarily on the high terrain east and northwest of the mill, although there were about 20 in the river valley near or below stack base. The most distant receptor was about 9.5 kilometers from the source and the closest was about 1.5 kilometers.

On-site meteorological data for Tracy were collected from an instrumented 150-m tower located 1.2 kilometers east of the power plant for the 128-hour period. The wind measurements from the tower were extended above 150 meters using a Doppler acoustic sounder and temperature measurements were extended with tether sonde data.

Table 3 Summary of AERMOD Evaluation Databases

Database	Description of Field Study	Available Meteorological Data
Prairie Grass (SO ₂)	Flat, grassy field (Nebraska, USA) Nonbuoyant source, single source Near-surface release; 44 data hours SO ₂ tracer measurements out to 800 meters	16-m meteorological tower, instrumented at several levels (wind, turbulence, and temperature data)
Kincaid (SF ₆)	Flat, rural (Illinois, USA) Highly buoyant single source Tall stack release (187 m); 375 hours of data SF ₆ tracer out to 50 km	100-m tower, instrumented at 2, 10, 50, and 100 meters (wind, turbulence, and temperature data)
Kincaid (SO ₂)	Flat, rural (Illinois, USA) Highly buoyant single source Tall stack release (187 m) 30 fixed monitors out to 20 km 4,613 hours of data	100-m tower, instrumented at 2, 10, 50, and 100 meters (wind, turbulence, and temperature data)
Baldwin (SO ₂)	Flat, rural (Illinois, USA) Highly buoyant source, 3 identical stacks Tall stack release (184 m) 10 fixed monitors out to 10 km 1 year of data	100-m tower, instrumented at 10 and 100 meters (wind and temperature data)
Indianapolis (SF ₆)	Flat, urban (Indiana, USA) Highly buoyant single source Tall stack release (84 m); 170 hours of data SF ₆ tracer out to 12 km	Urban tower on a bank building (94-m level data), 10-m towers placed in suburban and rural areas. All towers had wind, temperature, and turbulence data.
Clifty Creek (SO ₂)	Moderately hilly terrain, rural (Indiana, USA) Highly buoyant source, 3 identical stacks Tall stack release (208 m) 6 fixed monitors out to 15 km 1 year of data	60-m tower on the plateau, instrumented at the 10-m and 60-m levels (wind and temperature data)
Tracy (SF ₆)	High terrain, rural (Nevada, USA) Moderately buoyant single source Tall stack release (91 m); 128 hours of data SF ₆ tracer out to 8 km	150-m tower, instrumented at multiple levels (wind, temperature and turbulence data)
Martins Creek (:SO ₂)	Hilly terrain, rural (Pennsylvania/New Jersey, USA) Highly buoyant steam electric utility sources Tall stack releases (122-183 m) 7 fixed monitors out to 8 km 1 year of data	10-m tower plus sodar (wind, temperature, and turbulence data)
Lovett (SO ₂)	Hilly terrain, rural (New York, USA) Highly buoyant single source Tall stack release (145 m) 12 fixed monitors out to 3 km 1 year of data	100-m tower instrumented at the 10-m, 50-m, and 100-m levels (wind, temperature, and turbulence data)
Westvaco (SO ₂)	Hilly terrain, rural (Maryland/Virginia, USA) Highly buoyant single source Tall stack release (183 m) 11 fixed monitors out to 3 km 1 year of data	Three meteorological towers; the primary 100-m tower was instrumented at multiple levels (wind, temperature, and turbulence data)

EVALUATION PROCEDURES

The model evaluation was designed to provide diagnostic as well as descriptive information about the model performance. Highlights of the evaluation results for the current model are presented here using residual plots and quantile-quantile (Q-Q) plots (provided in Appendix A). The residual plots feature box and whisker symbols that show the distribution of cases along the y-axis domain for various “bins” or domain segments along the x-axis. The y-axis in this case is the ratio of the predicted to observed concentration, and the x-axis is distance (although it could be wind speed, stability parameter, or some other independent variable). The center of each box denotes the 50% y-axis value, and the bottom and top of the box correspond to the 25% and 75% values, respectively. The extremes of the box represent the 10% and 90% values.

For all data sets, quantile-quantile (Q-Q) plots are presented. Q-Q plots are simple ranked pairings of predicted and observed concentrations, such that any given quantile of the predicted concentration is plotted against the same quantile of the observed concentration. A solid line has been added to the Q-Q plots to indicate an unbiased prediction and two dotted lines have been added to indicate a factor of two under- and over-prediction. The Q-Q plot is an effective method for comparing the frequency distributions of two data sets.

Cox and Tikvart (1990) proposed a robust test statistic that represents a smoothed estimate of the highest concentrations, based on a tail exponential fit to the upper end of the distribution. With this procedure, the effect of extreme values on model comparison is reduced. This statistic is the robust highest concentration (RHC) and is given by:

$$\text{RHC} = X(N) + [\bar{X} - X(N)] \ln \left[\frac{3N - 1}{2} \right]$$

where

- M_0 = number of values used to characterize the upper end of the distribution
- M = # values exceeding a threshold value

$$\begin{aligned}
N &= \min(M_0, M) \\
\bar{X} &= \text{average of the } N-1 \text{ largest values, and} \\
X(N) &= N^{\text{th}} \text{ largest value}
\end{aligned}$$

In this evaluation, the value of M_0 was taken to be 26.

Many of the statistical tests and comparisons with observations were applied to analyze the performance of the model and various model algorithms. The observed peak concentration for a given arc of samplers was compared to the predicted arc maximum. The comparisons included time and downwind-distance pairings to significantly challenge the model components. Residual plots (predicted/observed, paired in time and downwind distance) of concentration estimates were used to judge whether AERMOD was working correctly and was yielding better results (than existing applied models) for the right reasons. Generally, residuals were plotted as a function of distance, although residuals versus other parameters, such as friction velocity, Monin-Obukhov length, and mixing height, proved to be extremely valuable diagnostic tools. In addition, quantile-quantile plots (that pair ranked concentrations that are unpaired in time and space) were used to examine the ability of the model to reproduce the distribution of observed concentrations over a wide range of environmental conditions. Reproducing the measured distribution (particularly the high concentration end) is important in regulatory applications of the model.

For the intensive tracer data sets (Prairie Grass, Kincaid SF₆, Indianapolis, and Tracy), concentration residuals of the form $\langle C_p/C_o \rangle$ were plotted as a function of downwind distance for each of two stability regimes (convective and stable). Here, C_o is the maximum observed concentration, and C_p the maximum predicted, on an arc at a given time. The brackets, $\langle C_p/C_o \rangle$, indicate the median of the ratio. These data were paired in time and downwind distance.

For the other data sets (Kincaid SO₂, Lovett, Baldwin, Clifty Creek, Martins Creek, and Westvaco), where the sampler array was not sufficiently dense to arrange the data in arcs, the concentration measures used were the maximum observed and predicted over the entire receptor array at a given time. For these, residual plots by distance were not meaningful.

Since AERMOD is being designed as a possible replacement for the Industrial Source Complex Short Term (ISCST3) Model, comparisons between AERMOD and ISCST3 are included for both the residual plots and the Q-Q plots for those data sets for which ISCST3 is applicable. Comparisons were also made with the CTDMPPLUS model (Perry, 1992), RTDM (Paine and Egan, 1987) for complex terrain, and with the HPDM model (Hanna and Paine, 1989; Hanna and Chang, 1993) for selected data sets.

For the tracer databases, results for 1-hour averages are reported. For the 1-year SO₂ data sets, 3-hour, 24-hour, and annual results are reported. All of the observed concentrations for the long-term databases are subject to uncertainty because a background concentration is subtracted from the actual monitored observations to obtain a “source-caused” impact. In addition, it should be realized that SO₂ monitors typically have a 6 ppb (16 µg/m³) detection limit, and baseline (zero) drifts of up to 10 ppb (26 µg/m³) are not corrected (Gendron, 1998). Concentrations below the detection limit are typically set to half of the limit (8 µg/m³), even though they may actually be zero. Another factor that could result in overestimates of “observed” concentrations is the acceptance without correction of nonzero concentrations caused by baseline drift that should actually be reported as zero. Therefore, the combined potential errors in SO₂ measurements from the detection limit treatment, ignored baseline drifts, and background concentration estimates can result in significant uncertainties in “observed” annual averages. Peak short-term averages are not affected significantly because the uncertainty is typically a small percentage of the reported value. However, the reader should interpret evaluation results for annual averages with considerable caution, even though these values are of considerable interest for chronic health risk studies.

ANALYSIS AND RESULTS OF THE DEVELOPMENTAL EVALUATION

Prairie Grass

Q-Q plots for the Prairie Grass data set for AERMOD (Figure A-1) and ISCST3 (Figure A-2) indicate that both models predict well within a factor of 2. The AERMOD concentrations in Figure A-1 parallel the 1-1 line for concentrations above 1000 $\mu\text{g}/\text{m}^3$, while ISCST3 concentrations do not parallel the 1-1 line until the observed concentration exceeds 10000 $\mu\text{g}/\text{m}^3$. AERMOD results for convective conditions for Prairie Grass are shown in Figures A-3 and A-4. The results show a slight underprediction tendency on the Q-Q plot, with no substantive trend notable on the residual plot. Similar results are evident for stable conditions (see Figures A-5 and A-6), with less of an underprediction tendency for AERMOD that was evident for unstable conditions.

The 1-hour RHC results (see Table 1) indicate a slight underprediction by AERMOD (0.87 ratio of predicted to observed RHCs), and an overprediction by ISCST3 (1.50).

Kincaid SF₆

Q-Q plots for all cases from the Kincaid SF₆ tracer study are plotted in Figure A-7 for AERMOD and in Figure A-8 for ISCST3. AERMOD's performance is clearly superior, with substantial underpredictions noted for ISCST3. An analysis of convective conditions for AERMOD (Figures A-9 and A-10) shows good performance on the Q-Q plot and no substantial trend on the residual plot. (In Figure A-10, while the middle of the box plots show a modest underprediction trend for AERMOD, the upper "whisker" crosses the Cp/Co ratio of 1.0 consistently.) Note that the peak unstable concentrations are significantly higher than the peak stable concentrations (see Figures A-11 and A-12). AERMOD's inability to match the comparatively lower observed stable concentrations may be partially due to a limited sample size in this database, and this behavior is not evident in the Kincaid SO₂ results discussed below.

The 1-hour RHC results (see Table 1) indicate a modest underprediction by both AERMOD (0.76 ratio of predicted to observed RHCs) and by ISCST3 (0.68).

Indianapolis

The Indianapolis data set provides a database on which to test the behavior of the models in an urban setting. It should be noted, however, that Indianapolis is the only urban study that was used in the AERMOD evaluations, and the confirmation of model components such as the nocturnal urban mixed layer height was made with the help of the field data from this specific site.

The Q-Q plots that include the entire database (Figures A-13 and A-14) show a nearly unbiased trend for AERMOD over the entire range of concentrations, while ISCST3 exhibits an overprediction tendency over the whole range. In convective conditions (Figures A-15 and A-16), AERMOD shows a slight underprediction tendency, with only a small trend with distance. The Q-Q plot for stable conditions (Figure A-17) indicates a nearly unbiased performance for AERMOD for a large portion of the concentration domain. Figure A-18 indicates a notable trend with distance, with underpredictions especially evident in the near field (within 1 km). However, these distances are generally associated with low observed concentrations (near the observation threshold), so an underprediction ratio involving two small values is not of significant concern.

The 1-hour RHC results (see Table 1) indicate a modest overprediction by AERMOD (1.20 ratio of predicted to observed RHCs), and by ISCST3 (1.30).

Kincaid SO₂

The Kincaid SO₂ database provides data from the same stack source as the Kincaid SF₆. There are, however, three main differences in that study: 1) The data base contains several months of continuous observations, 2) the sampler network is less dense, and 3) the pollutant being measured is the SO₂ that is emitted due to the sulfur contained in the fuel instead of the SF₆ tracer. Because the samplers are not arranged in arcs, residual plots by distance are not meaningful, and therefore have not been included. However, the database does allow for computation of 1-hour, 3-hour and 24-hour average concentration statistics. For this data set, the single highest concentration for each 1-, 3-, and 24-hour period was used. Q-Q plots for both AERMOD and ISCST3

are presented in Figures A-19, A-20, and A-21 for the 1-hour, 3-hour, and 24-hour averages, respectively. In each case, AERMOD's curve parallels the 1-1 line more closely, while ISCST3 is shown to consistently underpredict. Analyses of the convective (Figure A-22) AERMOD predictions show good results that are consistent with those of the Kincaid SF₆ results. The Q-Q plot of the stable hours (Figure A-23) indicate reasonably good AERMOD performance, in contrast with the poor showing of AERMOD in the sample size-limited Kincaid SF₆ database.

The 3-hour and 24-hour RHC results and the annual peak results (see Table 1) indicate a nearly unbiased predicted to observed ratio for AERMOD for the 3-hour and 24-hour averages (1.01 and 0.97, respectively) as opposed to underpredictions by ISCST3 (ratios of 0.45 and 0.45 for the 3-hour and 24-hour averages). Both models underpredict for the annual RHC statistic (0.30 for AERMOD and 0.14 for ISCST3). However, the low annual concentrations (near the instrument threshold) and the uncertainties in subtracting background concentrations make the "observed" average concentrations subject to considerable uncertainty.

Lovett

The Lovett data set provides a test on the AERMOD treatment of complex terrain. In terms of the complexity of its theoretical formulation, AERMOD lies between the current screening models and the CTDMPPLUS refined model (Perry, 1992). Q-Q plots comparing results for AERMOD and ISCST3 are shown in Figures A-24, A-25, and A-26 (for 1-hour, 3-hour, and 24-hour averaging times, respectively). Q-Q plots of AERMOD results show a curve very close to the 1-1 line for each averaging time. ISCST3, on the other hand, substantially overpredicts these concentrations for all three averaging times. (ISCST3 uses the COMPLEX-I screening model and the EPA Intermediate Terrain Procedures in these calculations, which is inherently "conservative," that is, it tends to overpredict.) The CTDMPPLUS results show a consistent overprediction tendency, by about a factor of 2.

Q-Q plots for AERMOD for convective and stable conditions are shown in Figures A-27 and A-28, respectively. In convective conditions, the AERMOD curve parallels the 1-1 line with very little bias for most of the concentration domain. In stable conditions, the AERMOD curve overstates concentrations except for the top few, which indicate a modest underprediction tendency.

The 3-hour and 24-hour RHC results and the annual peak results (see Table 1) indicate an overall unbiased predicted to observed ratio for AERMOD for the 3-hour and 24-hour averages (1.00 for both averaging times) as opposed to overpredictions by ISCST3 (ratios of 8.20 and 9.11 for the 3-hour and 24-hour averages) and overpredictions by CTDMPLUS (ratios of 2.36 and 2.02 for the 3-hour and 24-hour averages). AERMOD shows a slight underprediction for the annual average (ratio of 0.79), while ISCST3 continues to show a large overprediction (ratio of 7.51), and CTDMPLUS overpredicts within a factor of two (ratio of 1.71).

ANALYSIS AND RESULTS OF THE INDEPENDENT EVALUATION

Baldwin

The Baldwin site is a test of the model performance for tall stacks in flat terrain. Q-Q plots comparing results for AERMOD, ISCST3, and HPDM are shown in Figures A-29, A-30, and A-31 (for 1-hour, 3-hour, and 24-hour averaging times, respectively). Q-Q plots of AERMOD results show overpredictions (but within a factor of 2) for 1-hour averages, and nearly unbiased results for the 3-hour and 24-hour averages. The HPDM and AERMOD curves show a similar behavior for the Q-Q plots, except that for the highest concentrations, the HPDM predictions are slightly closer than the AERMOD predictions to being unbiased. In each figure, ISCST3 shows nearly unbiased concentrations for the top end of the concentration domain, but AERMOD's performance is better (closer to the 1:1 line) for a much larger range of the concentration domain in each case. ISCST3 underpredicts at the lower concentration values in each case.

The 3-hour and 24-hour RHC results and the annual peak results (see Table 1) indicate a modest overprediction tendency for AERMOD for the 3-hour average (ratio of 1.31) and a nearly unbiased 24-hour and annual average set of ratios (1.02 for the 24-hour average and 0.97 for the annual average). HPDM performs very well with the Baldwin database for all averaging times, with nearly unbiased estimates for the 3-hour and 24-hour averages (RHC ratios of 1.06 and 1.02, respectively), and a slight overprediction tendency for the annual average (predicted to observed ratio of 1.15). ISCST3 results indicate higher overpredictions for the 3-hour and 24-hour averages (ratios of 1.48 and 1.13, respectively), and underpredictions for the annual average (a predicted to observed ratio of 0.63).

Clifty Creek

This case features a tall stack with terrain extending at least halfway to stack top. Q-Q plots comparing results for AERMOD, HPDM, and ISCST3 are shown in Figures A-32, A-33, and A-34 (for 1-hour, 3-hour, and 24-hour averaging times, respectively). Q-Q plots of AERMOD results show overpredictions (but well within a factor of 2) for 1-hour averages, nearly unbiased results for the 3-hour average, and a modest underprediction

for 24-hour averages. HPDM shows overpredictions over most of the concentration range in the Q-Q plots, but it is still within a factor of 2 in each case. For the 3-hour and 24-hour averages, ISCST3 shows nearly unbiased concentrations for the top end of the concentration domain, but AERMOD's performance is once again better for a larger range of the concentration domain.

The 3-hour and 24-hour RHC results and the annual peak results (see Table 1) indicate a modest overprediction tendency for AERMOD for the 3-hour average (ratio of 1.25) and a modest underprediction for the 24-hour average (0.72). HPDM shows overpredictions for these short-term averaging times (ratios of 1.33 and 1.46 for the 3-hour and 24-hour periods, respectively). ISCST3 results for the same averaging times are 0.98 and 0.67 for the 3-hour and 24-hour averages, respectively. Both models show underpredictions for the annual peaks (ratios of 0.54 for AERMOD and 0.31 for ISCST3).

Martins Creek

This site represents a test of the complex terrain algorithms of AERMOD, ISCST3, RTDM, and CTDMPLUS. Q-Q plots comparing results for AERMOD, ISCST3, RTDM, and CTDMPLUS are shown in Figures A-35, A-36, and A-37 (for 1-hour, 3-hour, and 24-hour averaging times, respectively). Q-Q plots of AERMOD results show a similar trend in each case, featuring overpredictions of less than a factor of 2 over most of the concentration domain, but showing that the curve approaches the 1-1 line at the top, or has two peak points below the line. On the other hand, predictions of ISCST3, RTDM, and CTDMPLUS show significant overpredictions (with turbulence data for CTDMPLUS coming from AERMOD internally-generated profiles).

The 3-hour and 24-hour RHC results and the annual peak results (see Table 1) indicate a nearly unbiased result for AERMOD for the 3-hour average (ratio of 1.06) and an overprediction for the 24-hour average (1.74). AERMOD shows a modest underprediction ratio for the annual average (0.74). In contrast, the 3-hour and 24-hour ratios for ISCST3 are 7.25 and 8.88, showing significant overprediction. The CTDMPLUS resulting ratios are 4.80 and 5.56 for the same averaging times. RTDM overpredictions are somewhat less, with RHC predicted to observed ratios of 3.33 and 3.56 for the 3-hour and 24-hour averages. For annual averages, ISCST3, CTDMPLUS, and RTDM are still overpredicting, with predicted to observed ratios of 3.37, 2.19, and 1.32, respectively.

Westvaco

Westvaco is another complex terrain database. It was one of the independent evaluation data sets for CTDMPPLUS. Q-Q plots comparing results for AERMOD and CTDMPPLUS are shown in Figures A-38, A-39, and A-40 (for 1-hour, 3-hour, and 24-hour averaging times, respectively). Q-Q plots of AERMOD results show a nearly unbiased trend for the upper part of the concentration domain for each averaging time. For the short-term averages, CTDMPPLUS shows a factor-of-2 overprediction trend, with a less overprediction for the annual average.

The 3-hour and 24-hour RHC results and the annual peak results (see Table 1) indicate a nearly unbiased result for AERMOD for the 3-hour and 24-hour averages (ratios of 1.08 and 1.14, respectively), and an overprediction for the annual average (1.64). For the short-term averages, CTDMPPLUS shows overpredictions (ratios of 2.14 and 1.54 for the 3-hour and 24-hour averages). The CTDMPPLUS annual average ratio is 0.93.

Tracy

The Tracy Power Plant database was a developmental evaluation data set for CTDMPPLUS. Figure A-41 shows a Q-Q plot for 1-hour averages for both AERMOD and CTDMPPLUS. Both curves parallel the 1-1 line for the entire concentration domain, but AERMOD shows nearly unbiased results at the top end of the concentration range, while CTDMPPLUS exhibits a modest underprediction tendency. This trend is consistent with the results of the 1-hour RHC analysis, for which the AERMOD ratio of predicted to observed concentrations is 1.02, as opposed to a ratio of 0.77 for CTDMPPLUS.

SUMMARY AND CONCLUSIONS

The model evaluation results show a general consistency for AERMOD concentrations on the Q-Q plots to parallel the 1-1 line over a larger range of the concentration domain than other models tested. The AERMOD prediction bias exhibited on the Q-Q plots and in the RHC statistics shows an overall slight overprediction tendency. This trend was seen among the diverse set of databases that were evaluated. Apparent underpredictions for annual averages are, in part, probably artifacts of the low concentrations (close to the instrument thresholds) and the uncertainty in determining background concentrations that need to be subtracted from the reported total concentrations.

For simple terrain data bases, AERMOD's performance is comparable to that of HPDM, which is an advanced model that was expected to do well for these databases. This comparable result for AERMOD is another confirmation that the model's performance is consistent with expectations for state-of-the-art modeling techniques.

The overall results indicate that AERMOD is protective of air quality in view of the RHC values for 3-hour and 24-hour concentrations that are above 1.00. The better technical formulation of the model and its ability to provide better Q-Q plots statistics over a large concentration range provide the US EPA with adequate evidence to propose AERMOD as a guideline model to replace ISCST3.

REFERENCES

Barad, M. L., (Ed.), 1958: *Project Prairie Grass, A Field Program in Diffusion*. Geophysical Research Papers, No. 59, Vols. I and II, Report AFCRC-TR-58-235, Air Force Cambridge Research Center, 439 pp.

Bowne, N. E., R. J. Londergan, D. R. Murray, and H. S. Borenstein, 1983: *Overview, Results, and Conclusions for the EPRI Plume Model Validation and Development Project: Plains Site*. EPRI Report EA-3074, Project 1616-1, Electric Power Research Institute, Palo Alto, CA, 234 pp.

Carruthers, D.J., R.J. Holroyd, J.C.R. Hunt, W.S. Weng, A.G. Robins, D.D. Apsley, D.J. Thomson, and F.B. Smith, 1992. UK-ADMS – a New Approach to Modelling Dispersion in the Earth's Atmospheric Boundary Layer. In Proceedings of the Workshop: Objectives for Next Generation of Practical Short-Range Atmospheric Dispersion Models, May 6-8, 1992, Riso, Denmark, pp. 143-146.

Cimorelli, A. J., S. G. Perry, A. Venkatram, J.C. Weil, R. J. Paine, R. B. Wilson, R. F. Lee, and W.D. Peters, 1998: AERMOD: Description of Model , EPA web site (<http://www.epa.gov/scram001>).

Cox, W. and Tikvart, J., 1990: A Statistical Procedure for Determining the Best Performing Air Quality Simulation Model. *Atmos. Envir.*, 24A, 2387-2395.

DiCristofaro, D. C. , D. G. Strimaitis, B. R. Green, R. J. Yamartino, A. Venkatram, D. A. Gooden, T. F. Lavery and B. A. Egan, 1985. EPA Complex Terrain Model Development: Fifth Milestone Report - 1985. EPA-600/3-85-069, U. S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Gendron, L., 1998. Personal Communication (ENSR Corporation).

Hanna, S.R. and J.C. Chang, 1993. Hybrid Plume Dispersion Model (HPDM) Improvements and Testing at Three Field Sites. *Atmos. Envir.*, 27A, 1491-1508.

Hanna, S.R. and R.J. Paine, 1989: Hybrid Plume Dispersion Model (HPDM) Development and Evaluation. *J. Appl. Meteor.*, **28**, 206-224.

Haugen, D. A. (Editor), 1959: Project Prairie Grass, A field program in diffusion. Geophysical Research Paper, No. 59, Vol. III. Report AFCRC-TR-58-235, Air Force Cambridge Research Center, 439 pp.

Hayes, S.R. and G.E. Moore, 1986. Air Quality Model Performance: a Comparative Analysis of 15 Model Evaluation Studies. *Atmos. Envir.*, **20**, 1897-1911.

Liu, M. K., and G. E. Moore, 1984: Diagnostic validation of plume models at a plains site. EPRI Report No. EA-3077, Research Project 1616-9, Electric Power Research Institute, Palo Alto, CA.

Mills, M.T., 1979. Data Base for the Evaluation of Short Range Dispersion Models. R-001-EPA-79, Teknekron, Inc., Berkeley, CA.

Murray, D. R., and N. E. Bowne, 1988: Urban power plant plume studies. EPRI Report No. EA-5468, Research Project 2736-1, Electric Power Research Institute, Palo Alto, CA.

Olesen, H.R., P. Lofstrom, R. Berkowicz, and A.B. Jensen, 1991. An Improved Dispersion Model for Regulatory Use – the OML Model. In Proceedings of the 19th International Technical Meeting on Air Pollution Modeling and its Applications, Ierapetra, Crete, Greece. 29 September – 4 October 1991.

Paine, R.J. and B.A. Egan, 1987. User's Guide to the Rough Terrain Diffusion Model (RTDM), (Rev. 3.20). NTIS # PB88-171467/REB. U.S. Environmental Protection Agency, Research Triangle Park, NC.

Paumier, J. O., S. G. Perry, and D. J. Burns, 1992: CTDMPPLUS: A dispersion model for sources near complex topography. Part II: Performance characteristics. *J. Appl. Meteor.*, **31**, 646–660.

Pacific Environmental Services, 1998a. Final (Phase II) Evaluation of the AERMOD Dispersion Model (Draft). EPA Contract No. 68D70069. Prepared for U.S. Environmental Protection Agency, Research Triangle Park, NC (February, 1998).

Pacific Environmental Services, 1998b. Addendum to Performance (Phase II) Evaluation of the AERMOD Dispersion Model (Draft). EPA Contract No. 68D70069. Prepared for U.S. Environmental Protection Agency, Research Triangle Park, NC (October, 1998).

Perry, S. G., 1992: CTDMPPLUS: A dispersion model for sources near complex topography. Part I: Technical formulations. *J. Appl. Meteor.*, **31**, 633–645.

Perry, S. G., A. J. Cimorelli, R. F. Lee, R. J. Paine, A. Venkatram, J. C. Weil, and R. B. Wilson, 1994: AERMOD: A dispersion model for industrial source applications. *Preprints, 87th Annual Meeting of the Air and Waste Management Association*, Cincinnati, OH, June 19–24, Preprint No. 94-TA23.04.

Smith, M.E., 1984. Review of the Attributes and Performances of 10 Rural Diffusion Models. *Bull. Amer. Soc.*, **65**, 554-558.

Strimaitis, D. G., R. J. Paine, B. A. Egan and R. J. Yamartino, 1987. EPA Complex Terrain Model Development: Final Report. Contract No. 68-02-3421, U. S. Environmental Protection Agency, Research Triangle Park, North Carolina.

TRC, 1982. Evaluation of Rural Air Quality Simulation Models. EPA-450/4-83-003 (NTIS # PB83-182758), prepared for Environmental Protection Agency, Research Triangle Park, NC.

TRC, 1994: Air Quality Model Performance Evaluation and Comparison Study for Martins Creek Steam Electric Station. TRC Project No. 14715-R61. TRC Environmental Corporation, Windsor, CT.

Turner, D.B., T. Chico and J. Catalano, 1986. TUPOS—A Multiple Source Gaussian Dispersion Algorithm Using On-site Turbulence Data. EPA/600/8-86/010. US EPA, Research Triangle Park, NC, 39 pp.

US EPA, 1982: Evaluation of Rural Air Quality Simulation Models. EPA-450/4-83-003, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, 27711.

US EPA, 1987: User's Guide to the CTDM Meteorological Preprocessor Program. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, 27711.

US EPA, 1993: User's Guide for the Model Evaluation Methodology (MEM) System for Comparing Model Performance, Version 1.0 (Draft). U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, 27711.

U. S. Environmental Protection Agency, 1995a: User's guide for the Industrial Source Complex (ISC3) Dispersion Models. Volume 1—user instructions. EPA-454/B-95-003a, 320pp. [NTIS PB95-222741.]

U. S. Environmental Protection Agency, 1995b: User's guide for the Industrial Source Complex (ISC3) dispersion models. Volume II: Description of model algorithms. EPA-454/B-95-003b, 120 pp. [NTIS PB95-222758.]

Weil, J.C. and R.P. Brower, 1984. An Updated Gaussian Plume Model for Tall Stacks. *J. Air Pollut. Control Assoc.*, **34**, 818-827.

List of Figures

- Figure 1 Layout of the Prairie Grass SO₂ Tracer Experiment
- Figure 2 Topography in the Vicinity of the Kincaid Generating Station
- Figure 3 Example of Kincaid SF₆ Tracer Sampling Array for Neutral Conditions
- Figure 4 Location of Meteorological Sensors Deployed for the Indianapolis SF₆ Tracer Study
- Figure 5 Map Showing the Relationship of the Perry-K Station (A), the Hoosier Dome Sport Stadium (B), and the Central Indianapolis Business District (C)
- Figure 6 Locations of Fixed SO₂ Monitoring Stations in the Vicinity of the Kincaid Generating Station
- Figure 7 Depiction of the Monitoring Network Used for the Lovett Complex Terrain Model Evaluation Study
- Figure 8 Monitoring Locations in the Vicinity of the Baldwin Power Plant
- Figure 9 Monitoring Locations in the Vicinity of the Clifty Creek Power Plant
- Figure 10 Locations of SO₂ Monitors, Meteorological Stations, and Emissions Sources for the Martins Creek Model Evaluation Study
- Figure 11 Locations of SO₂ Monitors and Meteorological Towers in the Vicinity of the Westvaco Luke Mill
- Figure 12 Field Experiment Layout for the Tracy Power Plant Tracer Study
- Figure 13 Deployment of Tracer Sampling Sites for the Tracy Power Plant Study

Appendix A

Q-Q and Residual Plots Associated with the AERMOD Performance Evaluation

Listing of figures

- Figure A-1: 1-hour Quantile-Quantile plot for all hours evaluated in the Prairie Grass database, AERMOD predictions
- Figure A-2: 1-hour Quantile-Quantile plot for all hours evaluated in the Prairie Grass database, ISCST3 predictions
- Figure A-3: 1-hour Quantile-Quantile plot for convective hours evaluated in the Prairie Grass database, AERMOD predictions (χ/q)
- Figure A-4: Residual box and whisker plot for Prairie Grass convective hours paired in time and distance – ratio of AERMOD predicted to observed concentrations as a function of downwind distance
- Figure A-5: 1-hour Quantile-Quantile plot for stable hours evaluated in the Prairie Grass database, AERMOD predictions (χ/q)
- Figure A-6: Residual box and whisker plot for Prairie Grass stable hours paired in time and distance – ratio of AERMOD predicted to observed concentrations as a function of downwind distance
- Figure A-7: 1-hour Quantile-Quantile plot for all hours evaluated in the Kincaid SF₆ database, AERMOD predictions
- Figure A-8: 1-hour Quantile-Quantile plot for all hours evaluated in the Kincaid SF₆ database, ISCST3 predictions
- Figure A-9: 1-hour Quantile-Quantile plot for Kincaid SF₆ convective hours evaluated in the database, AERMOD predictions (χ/q)
- Figure A-10: Residual box and whisker plot for Kincaid SF₆ convective hours paired in time and distance – ratio of AERMOD predicted to observed concentrations as a function of downwind distance

Figure A-11: 1-hour Quantile-Quantile plot for Kincaid SF₆ stable hours evaluated in the database, AERMOD predictions (chi/q)

Figure A-12: Residual box and whisker plot for Kincaid SF₆ stable hours paired in time and distance – ratio of AERMOD predicted to observed concentrations as a function of downwind distance

Figure A-13: 1-hour Quantile-Quantile plot for all hours evaluated in the Indianapolis SF₆ database, AERMOD predictions

Figure A-14: 1-hour Quantile-Quantile plot for all hours evaluated in the Indianapolis SF₆ database, ISCST3 predictions

Figure A-15: 1-hour Quantile-Quantile plot for Indianapolis SF₆ convective hours evaluated in the database, AERMOD predictions (chi/q)

Figure A-16: Residual box and whisker plot for Indianapolis SF₆ convective hours paired in time and distance – ratio of AERMOD predicted to observed concentrations as a function of downwind distance

Figure A-17: 1-hour Quantile-Quantile plot for Indianapolis SF₆ stable hours evaluated in the database, AERMOD predictions (chi/q)

Figure A-18: Residual box and whisker plot for Indianapolis SF₆ stable hours paired in time and distance – ratio of AERMOD predicted to observed concentrations as a function of downwind distance

Figure A-19: 1-hour Quantile-Quantile plot for all hours in the Kincaid SO₂ database, AERMOD and ISCST3 predictions

Figure A-20: 3-hour Quantile-Quantile plot for all hours in the Kincaid SO₂ database, AERMOD and ISCST3 predictions

Figure A-21: 24-hour Quantile-Quantile plot for all hours in the Kincaid SO₂ database, AERMOD and ISCST3 predictions

Figure A-22: 1-hour Quantile-Quantile plot for Kincaid SO₂ convective hours evaluated in the database, AERMOD predictions (chi/q)

Figure A-23: 1-hour Quantile-Quantile plot for Kincaid SO₂ stable hours evaluated in the database, AERMOD predictions (chi/q)

Figure A-24: 1-hour Quantile-Quantile plot for all hours in the Lovett SO₂ data base - AERMOD, CTDMPLUS and ISCST3 predictions

Figure A-25: 3-hour Quantile-Quantile plot for all hours in the Lovett SO₂ database – AERMOD, CTDMPLUS and ISCST3 predictions

Figure A-26: 24-hour Quantile-Quantile plot for all hours in the Lovett SO₂ database – AERMOD, CTDMPLUS and ISCST3 predictions

Figure A-27: 1-hour Quantile-Quantile plot for Lovett SO₂ convective hours evaluated in the database, AERMOD predictions (chi/q)

Figure A-28: 1-hour Quantile-Quantile plot for Lovett SO₂ stable hours evaluated in the database, AERMOD predictions (chi/q)

Figure A-29: 1-hour Quantile-Quantile plot for all hours in the Baldwin SO₂ database - AERMOD, HPDM, and ISCST3 predictions

Figure A-30: 3-hour Quantile-Quantile plot for all hours in the Baldwin SO₂ database – AERMOD, HPDM, and ISCST3 predictions

Figure A-31: 24-hour Quantile-Quantile plot for all hours in the Baldwin SO₂ database – AERMOD, HPDM, and ISCST3 predictions

Figure A-32: 1-hour Quantile-Quantile plot for all hours in the Clifty Creek SO₂ database - AERMOD, HPDM, and ISCST3 predictions

Figure A-33: 3-hour Quantile-Quantile plot for all hours in the Clifty Creek SO₂ database – AERMOD, HPDM, and ISCST3 predictions

Figure A-34: 24-hour Quantile-Quantile plot for all hours in the Clifty Creek SO₂ database – AERMOD, HPDM, and ISCST3 predictions

Figure A-35: 1-hour Quantile-Quantile plot for all hours in the Martins Creek SO₂ database - AERMOD, CTDMPLUS, RTDM, and ISCST3 predictions

Figure A-36: 3-hour Quantile-Quantile plot for all hours in the Martins Creek SO₂ database - AERMOD, CTDMPLUS, RTDM, and ISCST3 predictions

Figure A-37: 24-hour Quantile-Quantile plot for all hours in the Martins Creek SO₂ database - AERMOD, CTDMPLUS, RTDM, and ISCST3 predictions

Figure A-38: 1-hour Quantile-Quantile plot for all hours in the Westvaco SO₂ database – AERMOD and CTDMPPLUS predictions

Figure A-39: 3-hour Quantile-Quantile plot for all hours in the Westvaco SO₂ database – AERMOD and CTDMPPLUS predictions

Figure A-40: 1-hour Quantile-Quantile plot for all hours in the Westvaco SO₂ database – AERMOD and CTDMPPLUS predictions

Figure A-41: 1-hour Quantile-Quantile plot for all hours in the Tracy SF₆ database -- AERMOD, CTDMPPLUS, and ISCST3 predictions