APPLICATION OF ASTM D6589 TO EVALUATE DISPERSION MODEL PERFORMANCE TO SIMULATE AVERAGE CENTERLINE CONCENTRATION VALUES

John S. Irwin[1], *David Carruthers*[2], *James Paumier*[3], *Jenny Stocker*[2]
[1]NOAA Atmospheric Science Modeling Division, RTP, NC, 27711;
[2]Cambridge Environmental Research Consultants, 3 King's Parade, Cambridge CB2 1SJ, UK;
[3]Pacific Environmental Services, Inc., RTP, NC 27709

INTRODUCTION

Models of atmospheric processes simulate only certain spatial and temporal scales. The observations used to test the performance of these models are individual realizations (which can be envisioned as coming from ideal ensembles), and are affected by variations unaccounted for within the simulation model. Given this, it makes sense to ask the models to replicate variations having certain spatial and temporal scales, but it does not make sense to ask the models to replicate the variations unaccounted for within the simulation model (e.g., observed maxima, or total variance).

The "Standard Guide for Statistical Evaluation of Atmospheric Dispersion Model Performance," published by the American Society for Testing and Materials (ASTM D6589, 2000), provides a framework for developing techniques that are useful for comparison of modeled and observed concentrations. To address the concern that models simulate only certain spatial and temporal variations and observations are individual realizations from imperfectly defined ensembles, the Guide suggests a two step process: Step 1) analyze the observations to provide smoothed observed patterns in time or space for comparison with modeled patterns (i.e., it is at this step that one attempts to remove variations from the observations that the models do not characterize), and Step 2) employ bootstrap resampling when comparing these patterns, which accounts for uncertainties in performing Step 1, and provides a means for objectively testing whether differences seen are statistically significant. An example procedure is provided in the Annex of D6589 for evaluating performance of models to estimate the average maximum ground-level centerline concentration. In the example procedure, centerline concentration values having similar meteorological conditions are grouped together at each downwind distance and comparisons are made of observed and modeled group averages. The procedure in the Annex is a test of only one dimension of a three dimensional problem. Thus for a complete evaluation, an additional test (say of the lateral diffusion) is needed, but for now we are focusing on the one test procedure that has been drafted.

In the following discussion, we summarize the results obtained in applying the D6589 example procedure to test the performance of four plume dispersion models: ADMS 3.1(*Carruthers et al*., 1994), AERMOD (versions 98022 and 02161, *Cimorelli et al*., 1996), HPDM (version 4.3, level 920605, *Hanna and Paine*, 1989), and ISCST3 (version 00101, *U.S. Environmental Protection Agency*, 1995), with tracer field data from three studies: Prairie Grass (*Barad*, 1958; *Haugen*, 1959), Kincaid (*Bowne et al.*, 1983), and Indianapolis (*Murray and Bowne*, 1988).

ANALYSIS

The Prairie Grass included 10-minute near-surface sampling along five arcs, 50 to 800 m, downwind from a near-surface point source release of sulfur dioxide, SO_2. The 20-minute releases were conducted during July and August of 1956, with an equal number of cases occurring during the daytime and nighttime. The sampling was for the 10-minute period in the middle of the 20-minute release. The mixing heights, Zi, were determined from on-site data.

The 68 experiments were sorted into 35 groups (5 arcs and 7 stability regimes defined in terms of 1/L, where L is the Monin-Obukhov length) as suggested by *Irwin and Rosu* (1998).

The sulfur-hexafluoride, SF_6, tracer experiments conducted at Kincaid (*Bowne et al*., 1983), involved a release from a 183-m stack with a buoyant plume rise on the order of 200 m. There were 171 experiments conducted during April, May and August of 1980, and May and June of 1981, with measurements of near-surface hourly concentrations and hourly meteorology. There were twelve roughly-defined receptor arcs ranging from 0.5-km to 50-km from the release. We divided the daytime data into four stability ranges defined in terms of Zi/L, where Zi is the mixing height. There were insufficient data at nighttime for our analyses. This provided 27 groups where centerline concentration values could be compared with modeling results. In defining these groups, the AERMET estimated values of Zi and L were used.

The SF_6 tracer experiments conducted at Indianapolis (*Murray and Bowne*, 1988) involved a release from an 84-m stack with a buoyant plume rise. There were 170 experiments conducted during September and October of 1985, with measurements of near-surface hourly concentrations and hourly meteorology. The mixing heights were determined from on-site data. There were twelve roughly-defined arcs ranging from 0.2-km to 12-km from the release. We divided the data using Zi/L into four daytime stability ranges and one nighttime stability range. This provided 29 groups where centerline concentration values could be compared with modeling results. In defining these groups, the AERMET estimated values of L were used with the observed Zi values.

The meteorology for each model was generated using each model=s meteorological processor, using the available on-site observations, hourly National Weather Service (NWS) weather observations, and twice-daily NWS upper air observations, to characterize the meteorological conditions for each of the three tracer dispersion sites (*Paumier*, 2001). The observed Zi values were used by all models for Prairie Grass and Indianapolis. For each release, the models were run so that the simulated centerline concentration was obtained for all possible downwind arcs for each field study. We then used these simulated centerline concentration values, C, divided by the emission rate, Q, for each experiment, for comparison with the observed C/Q values.

The ASTM procedure combines the observed C/Q values along arcs within a group for analysis, using the computed center of mass from each arc as a common reference point. Once combined, bootstrap sampling is applied to each group; samples are generated of observed and modeled centerline C/Q values, and the samples are averaged to produce observed and modeled average C/Q values for each group. The Annex to D6589 provides the details of how “centerline” C/Q values are selected. At the end of a bootstrap sampling pass, we have a sample average of observed and modeled centerline C/Q values for every group within a particular field experiment. The software computes several comparison statistics. Following ASTM D6589, we used the Normalized Mean Squared Error (NMSE) as an overall measure of bias and scatter. The NMSE results are stored for later use. The above processing is repeated for each bootstrap sample. We used 500 bootstrap samples. The model with the smallest average value of NMSE is the >base= model. We test whether the results from each of the other models is significantly different using the saved bootstrap NMSE values. The NMSE values are used to compute an average difference (>base=-model) and variance of the differences. A student-t test is then used to test whether the average difference from 500 boot samples is significantly different than zero.

DISCUSSION OF RESULTS

Table 1 summarizes the NMSE comparisons for the five models over the three field experiments. The >base= models (those having the lowest value of NMSE) are seen to be AERMOD (Version 02161) for Prairie Grass, and HPDM for Kincaid and Indianapolis.

Table 1. Summary of NMSE comparisons.

	Prairie Grass	EPRI KINCAID	EPRI Indianapolis
ADMS	1.375*	0.554	0.514*
AER98022	0.353	0.439	0.345
AER02161	0.210^{b}	0.489*	0.453
HPDM	6.295*	0.346^{b}	0.341^{b}
ISC3	2.915*	0.366	0.423

An asterisk (*) indicates that the value is significantly different with 90% confidence from the base model (which is indicated with a A**b**@). Results for Prairie Grass are for 30 of the original 35 groups, where the 5 groups excluded are the most stable stability group for each arc.

Figure 1 shows there is a bias in ADMS and ISC to overestimate concentration values for the unstable stability regimes and for ADMS, HPDM and ISC to underestimate concentration values for stable stability regimes. The variation of the bias with stability regime is seen at all arcs. L is < 0 for stability regimes 1 through 4, and L > 0 for stability regimes 4 and 5. HPDM is not recommended for low-level releases, and was included in the Prairie Grass comparisons purely as a matter of curiosity.

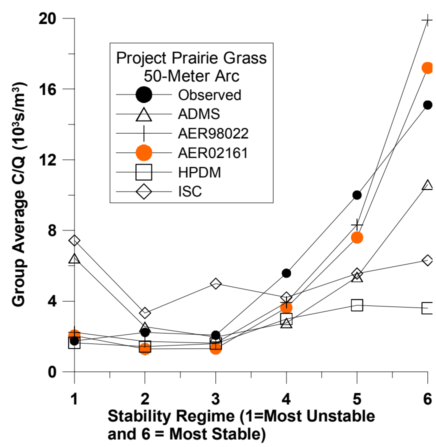

Figure 1. Project Prairie Grass group-average normalized concentration values (C/Q) as a function of stability regime for the 50-meter arc.

Figure 2A shows a bias in the modeling results for Kincaid which is most apparent for the slightly unstable (near-neutral) group. ADMS, AERMOD (Version 02161) and HPDM are overestimating the concentration values for distances downwind of 2 to 4 km. If we eliminate these three arcs from the overall analysis (i.e., run with 24 rather than 27 groups), the NMSE is smaller for HPDM end AERMOD (Version 02161), such that HPDM is still the >base= model, and the NMSE for AERMOD (Version 02161) is no longer deemed significantly different from HPDM. Elimination of these three arcs does not alter the NMSE for ADMS significantly, but the NMSE for AERMOD (Version 98022) is increased such that it is now deemed significantly different from HPDM.

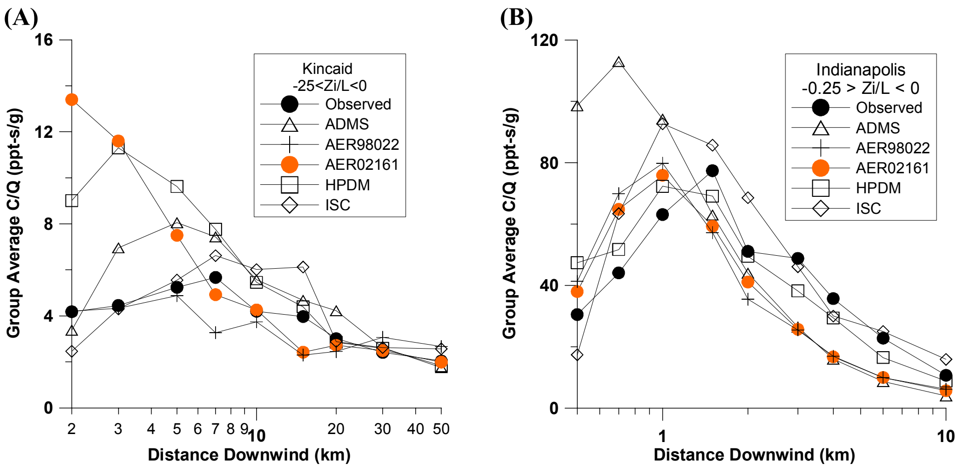

Figure 2. Group-average normalized concentration values (C/Q) for (A) Kincaid and (B) Indianapolis as a function of downwind distance for the slightly unstable (near-neutral) group.

Figure 2B shows for Indianapolis for the slightly unstable (near-neutral) group, that ADMS overestimates concentration values for the first three arcs. If we eliminate these three arcs from the analysis (i.e., run with 26 rather than 29 groups), the NMSE is smaller for ADMS and larger for AERMOD (Version 02261), but all models are accepted as performing as well as HPDM.

Several actions have been left for further research. A detailed review of the experiments placed into each group needs to be conducted to insure that questionable observations are not unduly influencing the results. As mentioned in ASTM D6589, given that the results may be sensitive to how data are sorted into groups, several different sorts should be tested to see if conclusions reached are robust. Finally, the results shown in Figures 1 and 2 suggest that a review of the modeling algorithms might indicate areas where further research could improve the performance of ADMS and AERMOD (Version 02161).

Disclaimer: The information in this document in part was funded by the US EPA under IAG 13938483-01 to NOAA. It has completed Agency review for approval for presentation and publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

REFERENCES

American Society for Testing and Materials, (2000): Standard Guide for Statistical Evaluation of Atmospheric Dispersion Model Performance (D 6589), (Available at http://www.astm.org), 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428, 17 pages.

Barad, M.L. (Editor), (1958): *Project Prairie Grass, A Field Program In Diffusion*, Geophysical Research Paper, No. 59, Vol 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 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. Thompson and F.B. Smith, (1994): UK-ADMS: A new approach to modelling dispersion in the earth's atmospheric boundary layer. *Journal of Wind Engineering and Industrial Aerodynamics*. (52):139-153.

Cimorelli A.J., S.G. Perry, R.F. Lee, R.J. Paine, A. Venkatram, J.C. Weil and R.B. Wilson, (1996): Current progress in the AERMIC model development program. *Preprints, 89th Annual Meeting and Exhibition of the Air and Waste Management Association, Air and Waste Management Association*, Nashville, TN, June 24-28, 1996, Preprint No. 96-TP24B.04, 27 pp.

Hanna, S.R., and R.J. Paine, (1989): Hybrid Plume Dispersion Model (HPDM) development and evaluation, *Journal of Applied Meteorology*, (28):206-224.

Haugen, D.A. (Editor), (1959): *Project Prairie Grass, A Field Program In Diffusion*, Geophysical Research Papers, No. 59, Vol. III, AFCRC-TR-58-235, Air Force Cambridge Research Center, 673 pp.

Irwin, J.S. and M-R. Rosu, (1998), Comments on a draft practice for statistical evaluation of atmospheric dispersion models. *Proceedings of the 10th Joint Conference on the Application of Air Pollution Meteorology with the A&WMA*. Phoenix, AZ., p 6-10.

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.

Paumier, J., (2001) *Meteorological Data Preparation*, EPA Contract No. 68D98006, Work Assignment No. 4-02. Technical Memorandum December 12, 2001, 6pp.

U.S. Environmental Protection Agency, (1995): *Industrial Source Complex (ISC3) Dispersion Model User's Guide*, EPA-454/B-95-003a (Volume I, 390 pp.) and EPA-454/B-95-003b (Volume II, 128 pp.), U.S. Environmental Protection Agency, Research Triangle Park, NC.