5.4 ASSESSING DISPERSION MODEL PERFORMANCE TO SIMULATE AVERAGE CENTERLINE CONCENTRATION VALUES

John S. Irwin[1*], David Carruthers[2], James Paumier[3], Jenny Stocker[2]
[1]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

1. INTRODUCTION

Most operational atmospheric simulation models are deterministic. They provide estimates of the average time- and space-variations in the conditions (e.g., mesoscale meteorological models), or they provide estimates of the average time- and space-variations of effects (e.g., air quality models). The observations used to test the performance of these models are individual realizations (which can be envisioned as coming from some ideal ensemble), and are affected by stochastic variations unaccounted for within the simulation model. If we believe this, then it makes sense to ask the models to replicate average patterns seen in the observations, but it does not make sense to ask the models to replicate the effects of stochastic variations unaccounted for within the simulation model (e.g., observed maxima, or total variance).

The American Society for Testing and Materials (ASTM), published in December 2000 D 6589, entitled, “Standard Guide for Statistical Evaluation of Atmospheric Dispersion Model Performance”, to provide a framework for developing techniques that are useful for comparison of modeled and observed concentrations, that addresses the concern that models provide deterministic estimates of ensemble averages, and observations are individual realizations from imperfectly defined ensembles. The Guide suggest a two step process: Step 1) analyze the observations to provide average patterns for comparison with modeled patterns, 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 D 6589 for evaluating performance of models to estimate the average maximum ground-level centerline concentration. In the example procedure, observations having similar meteorological conditions are grouped together at each downwind distance.

In the following discussion, we summarize the results obtained in applying the example procedure to test the performance of four plume dispersion models: ADMS 3.1(Carruthers et al., 1994), AERMOD (version 01247, 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).

2. Analyses

The Prairie Grass data included sampling along five arcs, 50 to 800 m, downwind from a near-surface point source release of sulfur dioxide, SO_2. The Obukhov length, L, was determined from the onsite meteorology. The 68 experiments were sorted into 35 groups from most unstable to most stable, with stability defined as 1/L, and divided the data into seven stability groups, for each downwind arc.

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 near-surface hourly concentrations with reasonably complete meteorology. There were twelve roughly-defined receptor arcs ranging from 0.5-km to 50-km from the release. We divided the data into four stability ranges defined in terms of Zi/L (Zi is the modeled mixing height). This provided 29 regimes where centerline concentration values could be compared with modeling results.

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 near-surface hourly concentrations, with reasonably complete meteorology. 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 36 regimes where centerline concentration values could be compared with modeling results.

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). For each tracer release, the models were run so that the simulated centerline concentration was obtained for all possible downwind arcs for each field study. These simulated centerline concentration values, C, divided by the emission rate, Q, were then used for each experiment, for comparison with the

---

[*] *Corresponding author address:* John S. Irwin, Office of Air Quality Planning and Standards (Mail Code D243-01), U.S. Environmental Protection Agency, Research Triangle Park, NC 27711

average of the observed C/Q values selected as being representative of centerline concentration values for a particular regime.

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. At the end of a bootstrap sampling pass, we have for each group a set of sample averages of observed and modeled centerline C/Q values for every group. We compute any (or several) comparison statistics of choice. We have use 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 this average difference is significantly different than zero.

3. Results

Table 1 summarizes the NMSE comparisons for the four models over the three field experiments. We presently have ADMS results for only Prairie Grass. It is seen that the "next" generation plume models (ADMS, AERMOD, and HPDM) are consistently the best performing models. AERMOD's results are not testing to be significantly different from the 'base' model's results. ISC3's results for Prairie Grass and Kincaid do test to be significantly different from the other models.

4. References

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. Elsevier Science B.

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.

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.

Table 1. Summary of NMSE comparisons

	Prairie Grass	EPRI Kincaid	EPRI Indianapolis
ADMS	1.117^{b}	Not Completed	Not Completed
AERMOD	1.144	0.400	0.368
HPDM	6.676*	0.298^{b}	0.341^{b}
ISC3	4.581*	1.652*	0.479

An asterisk (*) indicates that the value is significantly different with 90% confidence from the base model, which is indicated with a "**b".** ADMS results are for 25 regimes, excluding the most stable stability groupings.

Disclaimer. The information in this document in part was funded by the US EPA under IAG 13938483-01 to NOAA. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.