EVALUATION OF THE ADMS, AERMOD, AND ISC3 DISPERSION MODELS WITH THE OPTEX, DUKE FOREST, KINCAID, INDIANAPOLIS, AND LOVETT FIELD DATA SETS

Steven R. Hanna*, Bruce A. Egan**, John Purdum, and Jen Wagler*******

****CSI MS 5C3, George Mason Univ., Fairfax, VA 22030-4444 USA***
*****Egan Environmental, 75 Lothrop St., Beverly, MA 01915 USA***
******17 Peters Way, North Attleboro, MA 02760 USA***
*******URS Greiner Woodward-Clyde, 38 Chauncy St., Boston, MA 02111 USA***

ABSTRACT The model evaluation exercise addresses the question whether the new models, ADMS and AERMOD, produce improvements over ISC3 when compared with a range of field observations. ADMS and AERMOD have similar state-of-the-art scientific components, whereas ISC3 contains 1960s technology. The five sets of field observations used in the statistical evaluation represent a cross-section of typical scenarios encountered by modelers. The OPTEX data base concerns non-buoyant tracer releases within an oil refinery complex, and the Duke Forest data base involves non-buoyant tracer releases from area and volume sources in an open field. The Kincaid, Indianapolis, and Lovett data bases all deal with buoyant plumes from tall stacks at power plants. However, the settings are quite different, since the Kincaid plant is surrounded by flat farmland, the Indianapolis plant is located in an urban environment, and the Lovett plant is sited in a valley surrounded by complex terrain with monitors at elevations higher than the stack. Analysis of the model performance measures suggest that ISC3 typically overpredicts, has a scatter of about a factor of three, and has about 33% of its predictions within a factor of two of observations. The ADMS performance is slightly better than the AERMOD performance and both perform better than ISC3. On average, ADMS underpredicts by about 20% and AERMOD underpredicts by about 40%, and both have a scatter of about a factor of two. ADMS and AERMOD have about 53% and 46% of their predictions within a factor of two of observations, respectively. Considering only the highest predicted and observed concentrations, ISC3 overpredicts by a factor of seven, on average, while ADMS and AERMOD underpredict by about 20%, on average.

KEYWORDS Model evaluation, atmospheric dispersion models, field tracer data sets

INTRODUCTION

ADMS and AERMOD are new state-of-the-art dispersion models recently proposed for use in regulatory applications in the U.K. and the U.S., respectively. ADMS (CERC, 1998) is a state-of-the-art model developed over the past five years by a government and industry consortium in the U.K. AERMOD (Cimorelli et al., 1998) is intended to replace the 30-year-old technology (e.g., Pasquill-Gifford stability classes and “all-or-nothing” penetration of elevated inversions by buoyant plumes) as represented by the current U.S. EPA regulatory model, ISC3 (EPA, 1995). ADMS and AERMOD include many of the same scientific algorithms. The U.S. EPA has recently evaluated ISC3 and AERMOD using several field data bases (Paine et al., 1998). Because the EPA exercise did not include field data sets concerned with low-level area and volume sources or

tracer sources influenced by buildings and structures typical of a refinery or chemical plant, and because the EPA did not include the ADMS model in their evaluations, the additional model evaluations were carried out in this project. The statistical evaluations are carried out using data from five field sites – OPTEX (a refinery complex), Duke Forest (a volume release in a field), Kincaid (a power plant stack in flat rural terrain), Indianapolis (a power plant stack in urban terrain), and Lovett (a power plant stack in complex terrain). Hanna et al. (1999) provide a complete description of the research project and the results, including many figures and tables.

MODEL DESCRIPTIONS

Three different models (ADMS, AERMOD, and ISC3) have been selected for evaluation and are briefly described below.

Advanced Dispersion Modeling System (ADMS) ADMS (Carruthers et al 1995, CERC 1998) is an advanced steady state, Gaussian-like dispersion model developed in the U.K. by a government-industry consortium led by Cambridge Environmental Research Consultant (CERC). It is capable of simulating continuous plumes and short duration puff releases. Although it was developed primarily by funds from UK public agencies, it is a proprietary model which needs to be licensed for commercial applications. The model can be applied to point, line, area and volume sources and has a module applicable to motor vehicle emissions in street canyon situations. Improvements to the model over ISC3 are most evident in the treatment of dispersion rate variations within the atmospheric boundary layer. In this regard it is similar to AERMOD. Verification of the model has been partially based upon the Kincaid and Indianapolis data bases, which were also used to verify AERMOD (Carruthers et al., 1995 and 1998). For sources located in complex terrain, ADMS makes use of the FLOWSTAR algorithm (Carruthers et al., 1988), which was developed to simulate wind flow and turbulence for a variety of types of mountain and valley scenarios.

AMS/EPA Regulatory Model (AERMOD) AERMOD (Cimorelli et al., 1998) is a model developed under the auspices of the AMS/EPA Regulatory Model Improvement Committee (AERMIC), whose objective has been to include state-of-the-art science in regulatory models. AERMOD's technical components are described by Cimorelli et al. (1998) and its evaluations with field data are described by Paine et al. (1998). It is being proposed as a replacement for ISC3 for many applications, and has been built on the framework of ISC3. It retains the single straight line trajectory limitation of ISC3 but contains advanced algorithms to describe turbulent mixing processes in the planetary boundary layer for both convective and stably stratified layers. It also includes a detailed treatment of the dynamics of plumes that rise to interact with elevated inversions at the top of the convective mixed layer. AERMOD also offers new and potentially improved algorithms for plume rise and buoyancy, and the computation of vertical profiles of wind, turbulence and temperature. These algorithms are similar to those in ADMS. AERMOD also is able to address complex terrain above stack release heights. Although the intent has been that AERMOD would also incorporate improved algorithms (over ISC3) for building downwash and deposition processes, such advances are not available in the version evaluated here.

Industrial Source Complex Model Version 3 (ISC3) ISC3 (EPA, 1995) is recommended in EPA's Guideline on Air Quality Modeling for applications to refinery-like sources and other industrial sources in simple terrain. It is a straight line trajectory, Gaussian-based model that has evolved for over two decades. It is typically used with a minimum of requirements for meteorological input data (e.g., nearest NWS airport wind speeds and directions, ceiling heights, cloud cover, and Pasquill-Gifford stability class for each hour). ISC3 is generally run with a sequence of hourly meteorological conditions to predict concentrations at receptors for averaging times of one hour up to a year. In some applications, many years of hourly data are used as inputs to develop a better understanding of the statistics of calculated short-term hourly peaks or

of longer time averages. ISC3 contains detailed sets of algorithms to handle building downwash, deposition of particles, and area and line source releases.

The major advantages of ISC3 over AERMOD and ADMS are its relative simplicity of use and its robust predictions (i.e., the same results are obtained by different users for the same scenario). The amount of meteorological input data required by ISC3 is relatively small, and the model can be run sequentially with routinely collected NWS airport data. For a single meteorological condition for a passive pollutant, the meteorological data needed are a single wind speed, a wind direction, a stability class determination, and an assumed mixing depth. Terrain elevations at receptor points, building dimensions in addition to emissions and stack parameters are also needed. The disadvantages of ISC3 are largely associated with the fact that an improved knowledge of the structure of the atmospheric boundary layer and resulting estimations of turbulent dispersion processes cannot be accommodated in the model.

Comparisons of Technical Components The scientific review of the technical documents for ADMS (CERC, 1998) and AERMOD (Cimorelli et al., 1998) suggests that many of their components are based on similar sets of state-of-the-art algorithms (e.g., both assume the bimodal distribution of turbulent vertical velocities for convective conditions). On the other hand, ISC3 (EPA, 1995) represents the typical Gaussian “workhorse” model that has been in wide use for 30 years. The AERMOD algorithms have been streamlined (e.g., the treatment of terrain) so the model runs relatively quickly compared with ADMS. The downwash algorithm in AERMOD is unchanged from that in ISC3, whereas the downwash algorithm in ADMS is based on recent wind tunnel experiments and model development. ADMS is unique in that it can treat the transport and dispersion of instantaneous releases. There are a few differences in requirements for input meteorology, since AERMOD will allow vertical profiles of wind and temperature to be input, whereas ADMS requires that only one level of near-ground observations be input. Because several of the components of AERMOD and ADMS are relatively new, it would appear to be wise to carry out a series of sensitivity tests with a wide range of source and meteorological and terrain conditions, in order to be sure that the solutions are robust.

ISC3 requires a determination of whether the area surrounding a facility is either rural or urban, thus establishing the set of horizontal and vertical dispersion curves (Pasquill-Gifford for rural or McElroy-Pooler for urban). There are no intermediate or other dispersion rates used. AERMOD and ADMS can include surface conditions such as soil moisture (via Bowen Ratio or Priestley parameter), surface albedo (for net radiation estimations), and the surface roughness length. Surface roughness affects the vertical profiles of wind and temperature and the dispersion rates in the surface layer, and is an important variable in assessing dispersion in the vicinity of refineries and other industrial sites.

ISC3 uses routine meteorological data to calculate the height of the well-mixed layer. For plume rises less than the mixing-height, the plume is “trapped” and continues to mix within the layer by the use of reflection concepts. For plume rises above the mixing-height, the plume can no longer diffuse to the ground. ADMS and AERMOD include algorithms which quantify partial penetration of an elevated plume. The amount that is left to diffuse to the ground depends upon the strength of the inversion and the plume buoyancy. This parameterization is important for very buoyant plumes or for moderately buoyant plumes interacting with relatively low level inversions.

STATISTICAL EVALUATION METHODS

The performance of the three models is assessed using two basic methodologies. The first methodology involves direct quantitative comparisons of observed and predicted maximum concentrations at the five field sites. The second methodology involves the application of rigorous statistical procedures (Hanna, 1989), which are used to quantify several relevant performance measures. There is interest in understanding how well the three models predict the observed concentrations for the following pairings of observed and predicted values: 1. Pairing in

Time and at a Fixed Downwind Distance – Except for the Lovett site, the sampling arrays were set out on arcs at different radial distances from the source area. In this pairing methodology, for a given time period, the maximum observed and predicted concentrations on an arc are compared independent of position on that arc; and 2. Pairing in Time Only - For this pairing methodology, the maximum values observed and predicted anywhere in the sampling array (i.e., on any arc) for a given time period are compared.

For the OPTEX, Duke Forest, Kincaid, and Indianapolis sites, the evaluations deal with the maximum observed concentration on a given downwind arc, C_o. For the Lovett site, where the 12 monitors were not arranged along arcs, but were generally placed on mountainous locations where maximum concentrations were expected, the evaluations deal with the maximum concentration observed at any of the 12 monitors. For the OPTEX and Duke Forest sites, evaluations were also carried out for the cross-wind summed concentration on each given downwind arc, C_{yo}.

Comparisons of Highest Predicted and Observed Concentrations Because of the way the regulations are written in the U.S., there is great interest in a model's ability to accurately predict the highest concentrations. For this reason, our evaluations include tabulations of the highest concentrations for the three models and the five field sites. These tables include the highest concentrations 1) over the entire group of data, 2) arranged by arc distance, and 3) arranged by stability class, where possible. For the Kincaid, Indianapolis, and Lovett sites, where large amounts of data were available, tables are provided by Hanna et al. (1999) for the ten-highest observed and model-predicted concentrations. No statistical methods are applied to these data because there are only single numbers (or at most ten numbers at Kincaid, Indianapolis and Lovett) and the data are therefore not amenable to statistical significance tests. Instead, the amount of relative over- or under-prediction is discussed and compared from model to model.

Evaluations using Statistical Model Evaluation Software Model evaluation software (Hanna, 1989) was used to calculate several performance measures for the field data sets. These performance measures tend to equally weight all data, with no special emphasis on the highest concentrations that are considered in the previous paragraph. The following general performance measures were calculated for the point maximum, C_o, or the cross-wind sum, C_{yo} (cross-wind sums were available only at the OPTEX and Duke Forest sites). Note that a <> bracket indicates an average over all points in the group.

FAC2 – Fraction of predictions within a factor of two of the observations. (1)

FB (Fractional Bias) = $(\langle C_o\rangle - \langle C_p\rangle)/(0.5(\langle C_o\rangle + \langle C_p\rangle))$ (2)

NMSE (Normalized Mean Square Error) = $\langle(C_o - C_p)^2\rangle/(\langle C_o\rangle\langle C_p\rangle)$ (3)

MG (Geometric Mean) = $\exp(\langle \ln C_o - \ln C_p\rangle) = \exp(\langle \ln(C_o/C_p)\rangle)$ (4)

VG (Geometric Variance) = $\exp(\langle(\ln C_o - \ln C_p)^2\rangle) = \exp(\langle(\ln(C_o/C_p))^2\rangle)$ (5)

Both FB and MG deal with mean biases; however, FB uses the arithmetic concentrations while MG uses the log of the concentration. A similar relation occurs between NMSE and VG, which both deal with variances or scatter.

In addition, the software allows plots to be made of MG (or FB) versus VG (or NMSE) for groups of models. An ideal model has MG = VG = 1, or FB = NMSE = 0.0. The software also allows residual plots to be made, where the model "residual" C_p/C_o is plotted versus downwind distance, wind speed, or stability. It is desirable that a model's residuals not show any trends with independent variables such as downwind distance.

When applying the model evaluation software to tables of observed or predicted concentrations whose magnitudes may be very low, it is useful to impose a "minimum concentration" below which the concentrations are not allowed to drop. Otherwise, the ratio C_p/C_o can "blow up" in either direction (a near-zero value leads to a very large negative value of ln (C_p/C_o)). Or, very low concentrations may overly influence biases and variances. Consequently, an arbitrary minimum concentration is selected for the five field sites based on 1) information in the data reports on instrument thresholds, and 2) visual inspection of the tables of observed and predicted concentrations

DESCRIPTION OF FIVE FIELD SITES AND INPUT DATA ASSUMPTIONS FOR MODELS

This section contains brief descriptions of the five field sites. Site maps are not given because they are available in published papers and reports (e.g., Hanna et al., 1999).

OPTEX - The OPTEX field study took place at a refinery in southeast Texas (ENSR, 1997). Concentrations from 16 releases of SF_6 tracer from two groups of sources were measured along four different arcs. The so-called "Matrix Source" was near a heater area, which is a structure of piping and scaffolding. The non-buoyant tracer gas was released simultaneously from several nearby point sources, intended to simulate point, area, line and volume releases. For the purposes of this study, the Matrix Source was modeled as a set of individual point sources. Concentrations were measured at samplers along the arcs at distances approximately 300, 500, and 700 meters from the source areas. The so-called "Tank Farm Source", was a point source at a height of 1.5 m in the midst of storage tanks about 160 m upwind of the heater area. The nearest storage tank to the "Tank Release" was located a distance of 2.4 tank heights (H = 18 m) from the release point. The nearest line of receptors to the storage tanks was located about 60 meters downwind of the release point and about 100 meters upwind of the heater area. No minimum value was assigned to the observed or predicted OPTEX concentrations because the observations of arc-maxima are all above the instrument threshold and the model predictions are all non-zero.

Duke Forest – The OPTEX and Duke Forest field tests were related in that both involved complex sources; however, the OPTEX study was carried out at a refinery while the Duke Forest study was carried out at an open field site. There were two primary objectives of the OPTEX and Duke Forest experiments. The first objective was to evaluate the performance of optical remote sensing instruments for measurement of releases of toxic air contaminants from industrial sources (this objective is not addressed in the current paper but is discussed in the project reports). The second objective was to evaluate the performance of area and volume source dispersion models for predicting the near-field impacts of low-level releases from complex industrial sources (this objective is the subject of the current paper).

The Duke Forest experiment site was an open field in the middle of a forested area in North Carolina (Ogden Environmental and Energy Services Co., Inc., 1995). The field measurement program yielded tracer data for 11 days, with non-buoyant tracer releases of SF_6 from multiple point sources, which were modeled as separate points. One-hour averaged concentrations were taken by Tedlar bags deployed along arcs at downwind distances of 50 to 200 meters. For the purposes of model evaluation, a minimum concentration of 5.21 $\mu g/m^3$ was assumed.

Kincaid Power Plant - The Kincaid Power Plant is located in relatively flat farmland in Illinois. A series of intensive field experiments took place in which SF_6 tracer gas was injected into the buoyant combustion gas emissions from the plant's 187-meter tall stack (Bowne, et al., 1983). Approximately 200 monitors were placed in arcs ranging from about 0.5 to 50 km downwind of the stack, yielding one-hour averaged concentrations. Onsite meteorological data included wind speed and direction, turbulence, and ΔT from a 100-m instrumented tower. The data used for modeling and evaluation are from the ISC3 and AERMOD model evaluation exercise reported by Paine et al. (1998) and from the ADMS model evaluation exercise reported by Carruthers et al.

(1995). Only those hours and arcs were used where model predictions were available from all three models. Furthermore, to make the statistical results more robust, a "minimum concentration" of 1.15 ng/m^3 was assumed for all observations and predictions (this avoids biasing of the statistics due to extremely large over and underpredictions). The data base consists of 473 arc-hours.

Indianapolis Power Plant - The Indianapolis power plant is located in a flat urban/commercial setting in Indianapolis, Indiana. SF_6 tracer gas was injected into the buoyant combustion gas released from a single 84-meter stack (TRC, 1986). Data are available for 89 hours, for which 177 monitors were deployed along arcs ranging from .25 to 12 km downwind. Meteorological data included wind speed and direction and turbulence at an elevation of 94 m on a tower mounted on a downtown building. Wind speed, ΔT, and other surface data were also observed at three other towers in the area. A sampling time of one hour was used for the SF_6 measurements. The observations and model predictions used for the model evaluation exercise at Indianapolis were taken from the report by Paine et al. (1998) for ISC3 and AERMOD and from the paper by Carruthers et al. (1998) for ADMS. The model evaluation used only those arc-hours for which predictions were available from all three models available. A minimum concentration of 44.8 ng/m^3 was imposed. The data base consists of 777 arc-hours.

Lovett Power Plant - The Lovett Power Plant is located in mountainous terrain in a rural area along the Hudson River north of New York City. The sulfur dioxide (SO_2) gas was routinely released as part of the combustion gases in a buoyant plume. The data base consists of a year of ambient air measurements of SO_2 obtained at 12 continuous monitors at sites both above and below stack top elevations. The stack is 145 meters tall, and nearby mountains rise 100 to 200 meters above stack top. Ten of the monitors were located on a mountain at distances about 2000 to 3000 meters from the 145-meter stack. Two of the monitors were located at low elevations for the determination of background concentrations. Meteorological data include winds, turbulence, and ΔT from the 100 m meteorological tower. This data set was used by Paine et al. (1998) to evaluate ISC3 and AERMOD, and the observations and model predictions were obtained in electronic format from the authors. Predictions of ADMS have been made separately by CERC using the inputs used by Paine et al. (1998). As at Indianapolis and Kincaid, a minimum concentration (of 0.1 ug/m^3) was imposed on both observations and predictions. The data base consists of 2595 hours.

RESULTS OF STATISTICAL EVALUATIONS AT THE FIVE FIELD SITES

This section presents the results of the statistical evaluation exercise. The results are given individually for each of the five sites. A comprehensive summary of the evaluation results is given in the next section, including summary tables. Detailed tables and figures are given by Hanna et al. (1999).

Table 1. Overview of Model Evaluation Results for ISC3, ADMS, and AERMOD for Arc Max for the Five Field Sites

	OPTEX Tanks	OPTEX Matrix	Duke Forest	Kincaid	Indianapolis	Lovett
Num Pts	10	25	89	473	777	2595
Max C Obs.	3080	152	468	319	5379	447
Max C ISC3	25850	1567	1660	175	6780	4420
Max C ADMS	1121	469	440	211	5736	267
Max C AERMOD	1091	502	251	152	6076	441
MG ISC3	0.86	0.55	0.32	1.33	0.85	-1.68
MG ADMS	2.12	0.89	1.41	-0.03	1.14	0.14
MG AERMOD	2.47	1.02	1.83	0.75	1.54	-0.37
VG ISC3	2.8	2.9	11.6	8.5	6.8	46
VG ADMS	3	1.59	1.7	0.7	5.6	3.6
VG AERMOD	4.4	1.8	2	2.2	13	3.6
FAC2 ISC3	0.8	0.64	0.17	0.13	0.49	0.064
FAC2 ADMS	0.8	0.76	0.63	0.59	0.42	0.3
FAC2 AERMOD	0.7	0.76	0.53	0.29	0.39	0.25

OPTEX Tanks – For the OPTEX tank releases, which involved a ground-level point source in the midst of several tanks at a refinery, ISC3 overpredicts by about 15 %, on average, while ADMS and AERMOD underpredict by about 50 %, on average (see Table 1). But these average results mask some problems on individual arc distances. In particular, there are major discrepancies on the closest (x = 125 m) arc, where it is believed that the differences are due to different model assumptions concerning whether to account for the downwash effects of nearby tanks, which are located about 2.4 tank heights from the release position. As shown in Table 1, the peak or maximum C (which always occurs on the 125 m arc) is overpredicted by a factor of eight by ISC3, whereas ADMS and AERMOD both underpredict the peak by a factor of three. The models' performances improve on the second and third arc (x = 450 and 700 m, not shown in Table 1), where the predictions are nearly always within a factor of two.

OPTEX Matrix – For the OPTEX matrix releases, which involved releases from several points adjacent to a heater unit at a refinery, Table 1 shows that ISC3 overpredicts the peak or maximum concentration by a factor of 10, while ADMS and AERMOD overpredict this value by a factor of about three. This result may be due to differences in accounting for the porous nature of the heater structure just downwind of the release points. Table 1 also shows that, on average, ISC3 overpredicts by a factor of two, ADMS overpredicts by about 10 %, and AERMOD has little bias. ADMS and AERMOD have similar scatter. These relative differences are similar at all three downwind arcs, although there is a more scatter at the closest arc (x = 250 m). ISC3 is more of an outlier for this data set, with a tendency to overpredict by a factor of two and has a larger scatter. ADMS and AERMOD show fairly good overall performance, with little reason to distinguish them. The residual plots (not included here but presented in the report by Hanna et al., 1999) show little trend with arc distance, but do show a trend with wind speed, for there is a tendency to overpredict by a factor of two or three at the lower wind speed class. This difference may be due to the plume downwash algorithms and the method of accounting for the porous structure.

Duke Forest – The Duke Forest releases were from a set of point sources near the ground in a field and were unobstructed by tanks or buildings. Table 1 suggests that there is a clear and consistent tendency for ISC3 to overpredict the maximum and the average concentration by a factor of three. ADMS, on the other hand, tends to predict the maximum concentration very well (within 10 %), and underpredicts the mean concentrations by about 40 %. AERMOD underpredicts the maximum concentration by about a factor of two. When all data are considered, AERMOD's performance is similar to ADMS's, although AERMOD tends to

underpredict by a larger amount than ADMS (see Table 1). The residual plots (not included here) do not suggest any major trends with stability, downwind distance, and wind speed for ISC3, ADMS, or AERMOD (Hanna et al., 1999).

Kincaid – For the Kincaid data, Table 1 shows that the all three models underpredict the highest concentration (by about 45 % for ISC3, about 34 % for ADMS, and about 52 % for AERMOD). There is a variation in the relative results with arc distance, since the observed maximum occurred at x = 7 km, while the ISC3 predicted maximum was at 2 km, the ADMS predicted maximum was at 3 km, and the AERMOD predicted maximum was at 7 km (the latter in agreement with the observation, even though the predicted concentration was a factor of two too low at that arc). When all the data are included, the performance measures in Table 1 show that ADMS has the best performance, with an average bias of only 3 %. AERMOD underpredicts by a factor of two and ISC3 underpredicts by a factor of five, on average. The fraction of predictions within a factor of two of observations is 12 % for ISC3, 44 % for ADMS, and 29 % for AERMOD. However, there is more variation in model performance when the results for the individual arc distances are considered (see the report by Hanna et al., 1999, for details). ADMS has more consistent performance on most of the arcs. AERMOD's performance is better only on the more distant arcs (x = 30, 40, and 50 km). ISC3 consistently underpredicts at all arc-distances.

Indianapolis – For the Indianapolis data, Table 1 shows that all three models do a fairly good job of matching the highest concentration, with minor overpredictions of only 27 % for ISC3, 7 % for ADMS, and 13 % for AERMOD. The maximum observed concentration occurs at a downwind distance of 6 km, while ISC3, ADMS, and AERMOD predict that the maximum will occur at 4 km, 0.5 km, and 1 km, respectively. As downwind distance increases to 8 or 12 km, ISC has a tendency to overpredict by a factor of four and ADMS has a tendency to underpredict these highest values by a factor of three or four. However, AERMOD's performance show little bias at these distances with respect to its predictions of the very highest concentrations. The performance measure, MG, in Table 1 shows a mean overprediction of about 15 % for ISC3, a mean underprediction of about 14 % for ADMS, and a mean underprediction of about 50 % for AERMOD. AERMOD tends to underpredict by more than ISC3 or ADMS and has a larger scatter. 39 to 49 % of the models' predictions are within a factor of two of the observations. The residual plots (not included here but presented in the report by Hanna et al., 1999) suggest that, on average, ISC3 and AERMOD tend to underpredict on the closest arcs and overpredict on the farthest arcs, while ADMS overpredicts on the closest arcs and underpredicts on the intermediate arcs.

Lovett – For the Lovett data, Table 1 suggests that ISC3 overpredicts the peak or maximum concentration by a factor of ten, ADMS underpredicts this value by about 40 %, and AERMOD is within 1 % of the highest concentrations. However, this site was used by the AERMOD developers for calibration of the model parameters for complex terrain (Paine et al., 1998). Also, the ADMS predictions shown here are preliminary since they were made using an earlier version of the model that did not account for the terrain. Furthermore, the EPA's ISC3 model defaults to the COMPLEX-1 model if the terrain heights are above the stack top, as they are at Lovett. Therefore, the large ISC3 overpredictions would be associated with the default COMPLEX-1 model. Despite the clear indication of excellent AERMOD performance from the highest concentration comparisons, there are biases that appear when all the data are considered in Table 1. While ISC3 continues to overpredict the average concentration by a factor of ten, ADMS now has a reasonable mean bias of 14 % (towards underprediction), and AERMOD now has a mean overprediction of about 37 %. ADMS and AERMOD predictions are within a factor of two of observations about 30 and 25 % of the time, respectively.

COMPREHENSIVE EVALUATION OVER FIVE SITES AND GENERAL CONCLUSIONS

The previous section discussed the model evaluation results individually for each site. In this section, the performance measures are compared for all field sites and an attempt is made to

reach general conclusions. Table 1 contains a one-page summary of the major performance measures for all five sites, including the two release types (Tank and Matrix) for OPTEX. To save space and aid understanding, the results are given only for the "ALL" data category, and the results are not broken down into arc distance or stability class. A quick view of the table reveals that sometimes one model does better and sometimes another model does better. The only clear trend that is seen is that ISC3 often performs more poorly than ADMS or AERMOD. It is difficult to decide how to give a single "score" for each model based on the many (24) performance measures listed in the table. In the past, various model evaluation exercises have assigned weights to each result based on some arbitrary scheme decided upon beforehand. To simplify the scoring procedure it is assumed that the performance measures should be equally weighted. Then a ranking (best, middle, or worst) is assigned for each site and each performance measure (if the performance measures are close, they receive an equal ranking). The following result is obtained by adding up the arbitrary scores assigned to the performance measures in Table 1.

Table 2. Summary Scores for Models for Performance Measures in Table 1.

	ISC3	ADMS	AERMOD
Best	5	19	6
Middle	2	5	11
Worst	17	0	7

The "best, middle, and worst" numbers do not necessarily add up to 24 because of ties. However, the scores suggest that the models should be ranked with ADMS on top, AERMOD in the middle, and ISC3 on the bottom.

The ranking method does not reveal how close two scores may be. Consequently, the median performance measure was calculated for each performance measure and each model in Table 1. Because there are six field experiments (counting the OPTEX Tank and Matrix releases separately), the median was calculated as the average of the 3rd and 4th ranked values. The medians are given in Table 3 below:

Table 3. Median Performance Measures over All Field Experiments for Models.

	ISC3	ADMS	AERMOD
Max C_p/Max C_o	6.7	0.80	0.77
MG	0.70	1.22	1.7
VG	7.7	2.4	2.9
FAC2	0.33	0.53	0.46

The conclusions from Table 3 are similar to the conclusions from Table 2: 1) ISC3 generally has poorer performance than ADMS or AERMOD, and 2) ADMS and AERMOD have similar performance, with ADMS slightly better in each of the four performance measures. There are exceptions, such as the geometric mean, MG, of 0.70 for ISC3, which suggests a mean ISC3 overprediction of about 30 %. However, the relatively large scatter (VG = 7.7) for ISC3 suggests that the relatively good geometric mean, MG, is the result of cancellation of relatively large over and underpredictions for specific field data bases. Note that both ADMS and AERMOD have a general trend towards slight underpredictions of the overall Max C and of the mean of all concentrations.

For all field data sets except Lovett, the data represent fairly ideal conditions with persistent winds. However, the model applications for making regulatory decisions will extend beyond those limited conditions. Field data are generally not available from periods with rain, light and variable winds, dawn and dusk, and extreme stabilities.

ACKNOWLEDGEMENTS – This research was supported by the American Petroleum Institute. Because the analysis involved five field sites and three air quality models, much help was received from scientists and engineers in various agencies and companies. The authors appreciate the help of John Irwin, Warren Peters, and Steven Perry of the EPA, Roger Brode of PES, Russell Lee of Trinity Consultants, Robert Paine and Francis Kew of ENSR, John Lague of Ogden Environmental, and Christine McHugh, David Carruthers and Rex Britter of CERC.

REFERENCES

Bowne N.E., Londergan R.J., Murray D.R. and Borenstein H.S. (1983), *Overview, Results and Conclusions for the EPRI Plume Model Validation Project: Plains Site.* EPRI Report EA-3074, EPRI, Palo Alto, CA 94304.

Cambridge Environmental Research Consultants, Ltd. (1998), *ADMS Technical Specification*, CERC, 3 Kings Parade, Cambridge, U.K. CB2 1SJ (This is a compendium of 20 separate papers on various components of ADMS).

Carruthers D.J., Dyster S. and McHugh C.A. (1998), *Contrasting methods for validating ADMS using the Indianapolis dataset.* Proc. 5th Int. Conf. On Harmonization within Dispersion Modeling for Regulatory purposes, p. 104-110.

Carruthers D.J., Edmunds H.A., Ellis K.L., McHugh C.A., Davies B.M. and Thomson D.J. (1995), The Atmospheric Dispersion Modeling System (ADMS): comparisons with data from the Kincaid experiment. J. Envir. and Poll., Vol. 5, p. 111-120.

Carruthers D.J., Hunt J.C.R. and Weng W.S. (1988), *A computational model of stratified turbulent airflow over hills – FLOWSTAR I*. Computer Techniques in Environ. Studies (P. Zannetti, ed.), Springer-Verlag, p. 481-492.

Cimorelli A.J., Perry S.G., Venkatram A., Weil J.C., Paine R.J., Wilson R.J., Lee R.F. and Peters W.D. (1998), *AERMOD —Description of Model Formulation (Version 98314 (AERMOD and AERMET) and 98022 (AERMAP).* USEPA, RTP, NC 27711, 113 pages.

ENSR (1997), *Project OPTEX Field Study Report*, July 1997, ENSR Document # 0300-033-400, prepared for API, 1220 L St. NW, Washington, DC 20005.

EPA (1995), *User's Guide for the Industrial Source Complex (ISC3) Dispersion Model (Revised).* Volume II - Description of Model Algorithms. EPA-454/b-95-0036.

Hanna S.R. (1989), *Confidence limits for air quality model evaluations, as estimated by bootstrap and jackknife resampling methods,* Atmos. Environ., Vol. 23, p. 1385-1398.

Hanna S.R., Egan B.A., Purdum J. and Wagler J. (1999), *Evaluation of ISC3, AERMOD, and ADMS Dispersion Models with Observations from Five Field Sites*. HC Report P020, API, 1220 L St. NW, Washington, DC 20005-4070, 1999

Ogden Environ. and Energy Services, Inc. (1996), *Analysis of the Remote Sensing and Dispersion Data Generated by the 1995 EPA/API Coordinated Field Study at Duke Forest.* Prepared for API, 1220 L Street NW, Washington, DC 20005

Paine R.J., Lee R.F., Brode R., Wilson R.B., Cimorelli A.J., Perry S.G., Weil J.C., Venkatram A., and Peters W.D. (1998), *Model Evaluation Results for AERMOD*. USEPA, RTP. NC 27711.

TRC (1986), *Urban Power Plant Plume Studies*. EPRI Report EA-5468, Research Project 2736-1, Prepared by TRC, Hartford, CT, for EPRI, Palo Alto, CA.