Performance Evaluation of Dense Gas Dispersion Models

JAWAD S. TOUMA*
Atmospheric Sciences Modeling Division, Air Resources Laboratory, National Oceanic and Atmospheric Administration, Research Triangle Park, North Carolina

WILLIAM M. COX
Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina

HAROLD THISTLE
Missoula Technology and Development Center, USDA Forest Service, Missoula, Montana

JAMES G. ZAPERT
TRC Environmental Corp., Austin, Texas

(Manuscript received 27 August 1993, in final form 29 March 1994)

ABSTRACT

This paper summarizes the results of a study to evaluate the performance of seven dense gas dispersion models using data from three field experiments. Two models (DEGADIS and SLAB) are in the public domain and the other five (AIRTOX, CHARM, FOCUS, SAFEMODE, and TRACE) are proprietary. The field data used are the Desert Tortoise pressurized ammonia releases, Burro liquefied natural gas spill tests, and the Goldfish anhydrous hydrofluoric acid spill experiments. Desert Tortoise and Goldfish releases were simulated as horizontal jet releases, and Burro as a liquid pool. Performance statistics were used to compare maximum observed concentrations and plume half-width to those predicted by each model. Model performance varied and no model exhibited consistently good performance across all three databases. However, when combined across the three databases, all models performed within a factor of 2. Problems encountered are discussed in order to help future investigators.

1. Introduction

The U.S. Environmental Protection Agency (EPA) has an ongoing program to evaluate the performance of several categories of air quality dispersion models by comparing observed and predicted concentrations using performance measures recommended by the American Meteorological Society. Rural, urban, complex terrain, mobile, and long-range transport models are categories of models that have already been evaluated (Londergan et al. 1982; Londergan et al. 1983; Wackter and Londergan 1984; U.S. EPA 1986a,b; 1987). Models for toxic pollutant releases represent a broad class of models for which a systematic, indepen-

dent evaluation has not previously been performed. Dense gas releases represent a subset of toxic releases. Recently, two independent model evaluation studies have been completed. The first by Zapert et al. (1991) was sponsored by the EPA, and the second by Hanna et al. (1993) was sponsored by the American Petroleum Institute. This paper presents the results of the EPA-sponsored study. A full description of the study is provided in the EPA report.

An accurate simulation of the source–release conditions is critical in the determination of both the physical state and the initial dispersion characteristics downwind of a dense gas release. For example, if the release is from a leak in a pressurized liquefied gas storage tank, additional modeling is required to determine the state of the material as it enters the atmosphere, since the release may include both liquid (aerosol) and gaseous components. Parameters such as release density, temperature, and exit velocity can then be determined. Dense gas dispersion is complex because source–release conditions, initial gravitational spreading of a dense gas, and subsequent downwind dispersion as cloud density decreases must be taken into account.
2. Model selection and description

The dense gas models used in this study were selected from initial lists and surveys (Zapert et al. 1991). The EPA contacted the model developers to solicit interest for an evaluation study. Each model developer was presented with the objectives of the evaluation study and a list of candidate databases involved. From the initial list of models, three public domain [DEGADIS 2.1 (dense gas dispersion), HEGADAS (August 1988 version), and SLAB (March 1990 version)] and six proprietary models [AIRTOX (April 1990 version), CHARM 5.0 (complex hazardous air release model), FOCUS 1.0, MESOCHEM 2C, SAFEMODE (safety assessment for effective management of dangerous event, March 1990 version), and TRACE 2.54] were eventually selected. For each proprietary model, a confidentiality agreement, whereby all the material provided was to be returned after the conclusion of the study, was established before the model software and documentation were provided.

During preparation of test packages, several changes to the list of models were made. In particular, the HEGADAS and MESOCHEM models were withdrawn from consideration at the request of the model developers.

a. AIRTOX

AIRTOX (Mills 1888) was developed by ENSR Consulting and Engineering to calculate concentrations from chemical releases in either a jet or nonjet mode. AIRTOX is a spreadsheet-based model that uses Lotus 1-2-3 software. Chemical properties are provided automatically through an internal database. AIRTOX provides “snapshots” of predicted concentrations as a function of distance for user-specified times and as a function of time for user-specified locations. Concentration values are computed as a function of downwind distance. In addition, AIRTOX outputs information regarding the release profile and pool characteristics. The predicted concentrations from AIRTOX represent instantaneous snapshot values computed using 10-min-averaged dispersion coefficients.

b. CHARM

CHARM (Radian Corporation 1991) was developed by Radian Corporation to assess the location, extent, and concentration of the cloud that results from the release of a toxic substance into the air. CHARM includes a chemical database that provides all of the necessary chemical parameters to the model. CHARM is a menu-driven system composed of two parts: CHARM1 contains all of the screens for data input, while CHARM2 performs the calculations for the evolving cloud and controls the output. For this evaluation, CHARM was run in “planning mode,” which allows all data to be entered by the user. Model results are provided by CHARM in a graphical display that provides a “snapshot” of the cloud passage with time and produces concentration–dosage information for the release. The concentration information represents instantaneous values, while dosage represents time-averaged (user specified) values.

c. DEGADIS

The DEGADIS model (Spicer and Havens 1989) was developed at the University of Arkansas. The model includes a module for predicting the trajectory and dilution of an elevated dense gas jet. DEGADIS simulates aerosol dispersion with a user-specified concentration–density relation (based on adiabatic mixing of release aerosol and ambient air). The concentration–density relation is described using ordered triplets consisting of mole fraction, concentration, and mixture density. DEGADIS contains an internal chemical library that provides the model with the physical properties for the chemical being modeled. The user has the option to change these values. DEGADIS also allows the user to vary the averaging time for predicted concentrations.

d. FOCUS

FOCUS (Quest Consultants, Inc. 1990) is a hazards analysis model designed by Quest Consultants, Inc. to evaluate transient hazards from liquid or gas releases. FOCUS predicts hazard zones resulting from fires and explosions and the vapor clouds formed from releases of toxic and/or flammable materials. The model is controlled by an interactive control module that determines the sequence of programs to be executed. The model provides centerline concentrations as a function of time since release and the lateral distance to three user-specified concentration limits. In addition, it outputs information regarding the release profile and pool characteristics. The predicted concentrations represent values averaged over the release duration.

e. SAFEMODE

The SAFEMODE model (Raj 1990) was developed by Technology and Management Systems, Inc. as a tool for assessing the potential for acute hazards arising from the accidental release of toxic chemicals into the atmosphere. The user specifies source–release conditions in detail, including container dimensions, chemical name, storage conditions, leak geometry, and environmental conditions. After the model has calculated source parameters, the user has the opportunity to review and modify these values before allowing the model to continue with the release simulation. Predicted concentrations are displayed graphically as contours for specified hazard concentrations. Centerline concentrations and cloud widths are output at selected distances downwind of the release location. The user can specify
concentration averaging time. SAFEMODE has an internal chemical property library.

f. SLAB

The SLAB (Ermak 1989) model was developed at Lawrence Livermore National Laboratory to model four categories of releases: evaporating pools, horizontal jets, vertical jet or stack releases, and instantaneous or short duration evaporating pool releases. Releases can be treated as transient, steady state, or a combination of both. SLAB predicts downwind centerline concentrations and cloud width, averaged for a user-specified time period. The SLAB model user’s guide provides the necessary parameters for many of the chemicals of interest.

g. TRACE

The TRACE model (DuPont 1989) was developed by E. I. DuPont DeNemours & Company as a tool to evaluate the potential impact of toxic chemical spills. TRACE is an interactive, menu-driven model that allows the user multiple options when developing a release scenario. It contains an extensive chemical library. The TRACE model output provides information regarding vapor cloud dynamics, “snapshots” of concentration isopleths, and receptor impacts. The cloud dynamics section displays various cloud parameters as a function of time after release. TRACE provides time-averaged (user specified) concentrations at up to four user-specified receptor locations and 14 model-generated receptor positions.

3. Description of selected field experiments

Due to testing complexity and costs, databases applicable to dense gas releases are very limited. The databases selected for this evaluation study involved continuous releases from three well-known field experiments. Data from these experiments are in the public domain. Herein each experiment is described briefly.

a. Desert Tortoise pressurized ammonia releases

Four high-volume, pressurized liquid ammonia releases were conducted in 1983 at the Liquefied Gaseous Fuels Spill Test Facility in Nevada (Goldwire et al. 1985). Ammonia tanks were used to feed a 6-in. pipeline leading to the spill point. Ammonia was released at elevated (storage) pressure and ambient temperature. The jet release was directed horizontally downwind. A momentum jet (i.e., cloud speed in excess of wind speed) was observed at 100 m in all tests and at 800 m for tests 2 and 4. Tests 1, 2, and 4 were selected for this evaluation study. Evidence indicated that there was some liquid pooling near the release point but the pool could not be characterized. Release duration was assumed as 126, 255, and 381 s, for tests 1, 2, and 4, respectively.

Ammonia concentrations were sampled on a downwind grid along several arcs. At 100 m downwind, gas samplers were located on a tower at heights of 1, 3.5, and 6 m above ground. A second gas sampling arc was located at 800 m with five 10-m towers sampling gas at 1, 3.5, and 8.5 m with 100-m cross-wind separation. Concentrations were averaged to obtain 30-s values.

These three tests were conducted for similar meteorological conditions. Ambient temperatures ranged from 28.8° to 33.7°C. Wind measurements were taken at 2-m height and averaged for test duration. The stability class and wind speed assumed for tests 1, 2, and 4 were D and 7.4 m s⁻¹, D and 5.7 m s⁻¹, and E and 4.5 m s⁻¹, respectively. The test area is flat desert and surface roughness was reported as 0.3 cm.

b. Goldfish anhydrous hydrofluoric acid spill experiments

Six experiments were conducted in 1986 at the Liquefied Gaseous Fuels Test Facility in Nevada to study atmospheric releases of anhydrous hydrofluoric acid liquid from a heated, pressurized storage tank (40°C, 6.8 atm) (Blewitt et al. 1987). Three of the six tests (tests 1–3) were designed primarily to study vaporization and aerosol generation, cloud density, and dispersion. These three tests are selected in this evaluation study.

Releases were made as a horizontal liquid jet from a spill pipe using a release system similar to the Desert Tortoise ammonia tests. The liquid exited the pipe as a horizontal jet and was directed downwind toward a collection pad. In addition, the spill pad was designed to collect any pooling material. In the tests conducted, pooling was not observed. The spill rates were obtained by linear regression. Concentrations were sampled at multiple vertical levels on three sampling arcs. In test 3, additional moisture was added to the air upwind of the source using a combination of pond and steam generators. Release duration was assumed as 125, 360, and 360 s for tests 1, 2, and 3, respectively.

Concentration data collected include measurements on three sampling arcs (300, 1000, and 3000 m). Measurements were made at 1, 3, and 8 m above the ground surface. Concentrations were averaged to obtain 10-s values.

The meteorological data used for this study represent test-average values as provided by Blewitt et al. (1987). Ambient temperature ranged from 26° to 37°C. Wind measurements were made at a height of 2 m. The atmospheric stability was neutral for the three tests. Wind speed was 3.6, 4.2, and 5.4 m s⁻¹ for tests 1, 2, and 3, respectively. Surface roughness was reported as 0.3 cm.

c. Burro liquefied natural gas (LNG) spill tests

The Burro series of LNG spill experiments were performed in 1980 at China Lake, California (Koopman
et al. 1982). Eight spills of LNG onto water were made. Concentration measurements were made at 57, 140, 400, and 800 m from the source. All tests were conducted over a desert range with a steep slope rising 7 m in elevation from the pond to 80 m downwind. Beyond 80 m the terrain was relatively level. Of the eight Burro tests conducted, tests 3, 5, and 8 were selected for this evaluation. The other tests were excluded due to reported problems.

The LNG was released from a cryogenic liquid storage tank. The spill pipe was directed straight down toward the water with a splash plate installed at a shallow depth below the spill pipe outlet. Consequently, after the LNG stream encountered the water, it was directed radially outward along the surface of the water. Although spread of LNG due to the spill plate probably resulted in sources of finite area and transient release rates, no information is reported on the size of the area source or variability of emissions for modeling. Gas concentrations were measured at 30 stations at heights of 1, 3, and 8 m above the ground. All concentration data were averaged to 10 s for the study. Release duration was assumed as 166.8, 190, and 107 s, for tests 3, 5, and 8, respectively.

Meteorological data were collected at 2 m above the ground. Average values for test duration for temperature, humidity, Pasquill stability class, and Monin–Obukhov length were taken from the Burro data report. The stability class and wind speed assumed for tests 3, 5, and 8 were B and 5.4 m s$^{-1}$, C and 7.4 m s$^{-1}$, and E and 1.8 m s$^{-1}$, respectively. The surface is a dry lake bed with a surface roughness value of about 0.02 cm.

4. Study parameters

a. Development of test packages

The evaluation study required a thorough understanding of model details in order to properly apply that model to the experimental databases. Since five of the seven models included in the evaluation study are proprietary, interaction with model developers was required in order to assure proper application. In some cases, the user interface presented a serious obstacle to realistically simulating the releases. The model developers provided their models at the beginning of the evaluation and these models were used throughout the evaluation. The models were undergoing changes, but the model developers were not allowed to customize their models for this study. After the models were received from the developers, test packages were developed for each model using one test from each experimental database. In several cases, model developers enclosed test cases with their model that represented experiments included in the evaluation. If problems were encountered or documentation accompanying a model was inadequate, the developer was consulted as necessary for resolution of technical issues. As the test packages were completed, each model developer was given the opportunity to review and comment on the proposed application for that model. Based on the comments received from the model developers, the test packages were finalized.

b. Model application (input assumptions)

A list of potentially significant release characteristics for each experiment was developed (Zapert et al. 1991). Each model required a unique set of input parameters for each experiment. The models involved varied in applicability, complexity, and level of documentation. Model user’s guides and the technical literature were first reviewed to understand how to best apply the model to the physical processes that are being simulated. When this information was not available, the model developer was consulted. Whenever possible, chosen input values were made consistent between models.

c. Model limitations

The models are not all designed to predict at locations that correspond to where concentrations are measured; many provide predictions only at locations determined internally by the model. Several models estimate concentrations only at ground level. For the evaluation, measured concentrations were taken from samplers at 1-m height. These values were compared to model predictions at ground level, or at 1 m, if the model allowed for elevated receptors.

The DEGADIS and AIRTOX models produce only predictions at ground level. For these models, the ground-level predictions were compared to the 1-m measurements. For CHARM, FOCUS, and SAFE-MODE, difficulties were encountered in obtaining meaningful concentration predictions for the near-field arcs in certain cases. CHARM and SAFE-MODE would not produce predictions for the Desert Tortoise 100-m arc, because this arc falls within a “jetting region” predicted by the models. The FOCUS model produced erratic concentration predictions for the Burro 57-m arc; no predictions were made at this distance.

These models (and databases) are limited to prediction of dispersion of a release emitted from a flat surface at ground level dispersing over flat, unobstructed terrain.

d. Model averaging time

Model averaging time was determined as a function of the measured concentration data, meteorological data, release duration, and model limitations. Measured concentration data were averaged for a time period approximately equal to the test duration. All models were run with the concentration averaging time equal to this time period determined from the measured concentration data. As an example, for Desert
Tortoise test 4, a 360-s averaging time was used. The measured concentrations were averaged over 360 s since the release lasted 381 s and the available concentration data represent 30-s averages.

This choice of averaging times was used because all of the experiments involved continuous releases and the meteorological data are representative of test duration. In addition, it provided a reasonable degree of consistency among the models.

5. Evaluation results

The statistical evaluation of model performance compares observed and predicted maximum centerline concentrations at each receptor arc and a plume half-width indicator at each arc. The plume half-width is defined as the lateral distance between the maximum concentration and the location at which concentrations have decreased to 50% of the maximum. For measured concentrations, the two half-width values on either side of the maximum were averaged.

A variety of statistical comparisons between observed and model predicted maximum concentrations were used to assess the performance of these models (Zapert et al. 1991). These statistical methods included differences between observed and predicted values, Student's t-tests, and confidence limits for the average difference. The use of these methods for testing the various hypotheses was rejected because of the limited size of the database.

Because of these limitations, model performance is displayed in several graphical formats using the observed and predicted values. Because the radial distance between receptor arcs and source release points were different among the three experiments, it was difficult to compare results directly across the three databases. This difficulty was overcome in part by illustrating the observed and predicted values as a function of distance for each database. In addition, fractional bias was calculated using observed data and model predictions.

a. Plume centerline concentrations

Model-predicted versus observed maximum concentrations for each database are shown in Figs. 1–3. For Desert Tortoise test 1, the results in Fig. 1a show that model predictions span a wide range (more than a factor of 20) at the 100-m distance but converge to a narrower range at 800 m. For tests 2 and 4, the results in Figs. 1b, c show a pattern similar to test 1, but results at 800 m span a wider range. For all three tests, results at 100 m indicate overprediction by DEGADIS, AIRTOX, and FOCUS; relatively close agreement by TRACE; and underprediction by SLAB. At 800 m, FOCUS, DEGADIS, and TRACE were essentially unbiased and consistently predicted within a factor of 2, SLAB and SAFEMODE consistently underpredicted, and AIRTOX and CHARM showed great variability.

Figure 2a illustrates the observed and predicted maximum concentrations at each distance for Goldfish test 1. At 300 m, the predicted maximum values span a factor of 10. At 1000 and 3000 m, values from six of the models are clustered within a factor of 3. At all three distances, the majority of predicted values are lower than observed; at 1000 m, all seven models underpredict. Results of tests 2 and 3 are illustrated in Figs. 2b, c, respectively. The model predictions at both distances and the observed maximum value at 300 m are very similar, but the observed maximum values at 1000 m for test 3 are lower by more than a factor of 2.

The observed and predicted maximum concentrations at each distance for Burro test 3 are illustrated in Fig. 3a. Over the range of distances considered, AIRTOX underpredicted the maximum observed value at all four distances, while SLAB showed no consistent bias. TRACE, CHARM, DEGADIS, FOCUS, and SAFEMODE all overpredicted. Results for test 5 are illustrated in Fig. 3b. The observed maximum at 800 m is a factor of 2 higher for test 5. For test 8, observed maximum concentrations are substantially higher, and model performance was quite different, as illustrated in Fig. 3c.

To facilitate graphical presentation of model performance, the maximum observed and predicted concentration data were stratified into two distance categories: near—300 m or less, and far—over 300 m. Two of the models had missing data for at least one of the distance categories. These two models (CHARM and SAFEMODE) were excluded from further analysis. Experimental results for distances under 100 m were not included in the statistical evaluation since two of the three experiments (Desert Tortoise and Goldfish) provide no data within the 100-m range. Exclusion of data below 100 m allowed one additional model (FOCUS) to be included in the analysis.

For each distance category and experimental unit, the fractional bias was used as the measure of model performance (Cox and Tikvart 1990). This flexible statistical methodology makes it easier to combine results from different databases and to determine the best performing model. In cases where more than one data pair were available for a given distance category (e.g., Goldfish for 1000- and 3000-m arcs), separate fractional bias statistics were computed and then averaged. The fractional bias (FB) was computed using the following expression:

\[
FB = \frac{OBS - PRED}{0.5(OBS + PRED)},
\]

where OBS is the maximum measured concentration and PRED is the maximum predicted concentration.

The fractional bias is convenient for comparing model performance because it is bounded, symmetric, and dimensionless. Values range from -2.0 (extreme
overprediction) to +2.0 (extreme underprediction). Values between -0.67 and +0.67 indicate model predictions that are within a factor of 2 of the measured values. A second fractional bias was also determined using expression (1), except the plume half-width was used instead of the maximum concentrations.

For each database, a combined fractional bias was determined by arithmetically averaging the results of the individual tests. Because of these data limitations, it was inappropriate to conduct a full-scale comparative model evaluation. For Desert Tortoise and Goldfish, the combined fractional bias represents the average of three tests, while for Burro the average values are computed using just two test results. Because the number of data points provided in each dataset is small, the results should be viewed cautiously.

Table 1 presents the combined fractional bias by experiment for each of the five models and two distance
categories using the maximum concentration estimates. The last column in the table is the combined (average) fractional bias across the three experiments. The last five rows of the table represent the combined (average) fractional bias for the two distance categories for each of the five models. The results are shown graphically in Fig. 4 where the y-axis represents the combined fractional bias for the "near" distance category (300 m and less) and the x-axis represents the combined fractional bias for the "far" distance category (over 300 m) using maximum concentrations.

For Desert Tortoise (Fig. 4a), the models, except SLAB, are relatively unbiased for the far distance category but show a tendency for overprediction at distances closer to the source. For Goldfish (Fig. 4b), each model tends to underpredict the maximum concentration for both distance categories. AIRTOX appears to have the largest underpredic-
tion, while DEGADIS and TRACE underpredict within a factor of 2. For Burro (Fig. 4c), FOCUS, TRACE, and DEGADIS exhibit fairly extreme overpredictions, while SLAB and AIRTOX show relatively little bias. An examination of all three figures above suggests that none of the models appears to exhibit a consistently better or worse performance across the three databases for each distance category. For example, TRACE performs well for both Desert Tortoise and Goldfish but overpredicts at Burro. Both SLAB and AIRTOX appear to perform relatively well at Burro but do poorly at both Goldfish and Desert Tortoise.

b. Plume half-width

Data on plume half-width were examined to gain additional insight into the performance of the models.
Table 1. Average fractional bias using maximum observed and predicted concentration values.

<table>
<thead>
<tr>
<th>Distance</th>
<th>Experiment</th>
<th>Model</th>
<th>Desert</th>
<th>Tortoise</th>
<th>Goldfish</th>
<th>Burro</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Far</td>
<td>AIRTOX</td>
<td>0.12</td>
<td>1.21</td>
<td>0.52</td>
<td>0.62</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DEGADIS</td>
<td>−0.25</td>
<td>0.37</td>
<td>−1.15</td>
<td>−0.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FOCUS</td>
<td>−0.14</td>
<td>1.07</td>
<td>−1.88</td>
<td>−0.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SLAB</td>
<td>0.89</td>
<td>0.93</td>
<td>−0.01</td>
<td>0.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TRACE</td>
<td>−0.01</td>
<td>0.51</td>
<td>−1.60</td>
<td>−0.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Near</td>
<td>AIRTOX</td>
<td>−1.39</td>
<td>1.16</td>
<td>0.76</td>
<td>0.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DEGADIS</td>
<td>−1.55</td>
<td>0.47</td>
<td>−1.24</td>
<td>−0.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FOCUS</td>
<td>−0.90</td>
<td>0.19</td>
<td>−1.78</td>
<td>−0.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SLAB</td>
<td>0.87</td>
<td>1.10</td>
<td>0.31</td>
<td>0.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TRACE</td>
<td>−0.30</td>
<td>0.71</td>
<td>−0.81</td>
<td>−0.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined</td>
<td>AIRTOX</td>
<td>−0.63</td>
<td>1.19</td>
<td>0.64</td>
<td>0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DEGADIS</td>
<td>−0.90</td>
<td>0.42</td>
<td>−1.20</td>
<td>−0.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FOCUS</td>
<td>−0.52</td>
<td>0.63</td>
<td>−1.83</td>
<td>−0.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SLAB</td>
<td>0.88</td>
<td>1.01</td>
<td>0.15</td>
<td>0.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TRACE</td>
<td>−0.16</td>
<td>0.61</td>
<td>−1.21</td>
<td>−0.25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 presents the combined fractional bias by experiment for each of the five models and two distance categories. The results are shown graphically in Fig. 5. For Desert Tortoise (Fig. 5a), bias for plume-width is minimal for AIRTOX but noticeable for DEGADIS and TRACE (overpredictions) and FOCUS (underprediction) at each distance category. For Goldfish (Fig. 5b), none of the models exhibit appreciable bias at the larger distances; however, three models (FOCUS, DEGADIS, and TRACE) do exhibit noticeable bias for distances closer to the source. At Burro (Fig. 5c), the fractional bias for each of the five models is very near 0 for both distance categories.

c. Summary of model performance

By comparing the results shown in Fig. 4 with those shown in Fig. 5, it seems apparent that the models as a group perform better in predicting the plume half-width than they do in predicting maximum concentrations. The results from each experimental program (Desert Tortoise, Goldfish, and Burro) are summarized herein for each model.

- AIRTOX: For Desert Tortoise, AIRTOX produced substantial overprediction of maximum concentrations at 100 m but less bias at 800 m. For Goldfish, it substantially underpredicted maximum concentrations at all distances. For Burro, results for maximum concentration showed a mixed pattern with overprediction by a factor of 2 for one test but agreement within a factor of 2 for five of seven data points for the other two tests. The model performed relatively well at close distances but showed large overestimation at the farthest distances, especially for the Burro tests.

- CHARM: For Desert Tortoise, CHARM underpredicted maximum concentrations at 800 m. For Goldfish, it performed relatively well for estimating maximum concentrations at all distances with some underprediction bias at 1000 and 3000 m. For Burro, it overpredicted maximum concentration by a factor of 1.5 at 400 m but exhibited significant scatter at 800 m. The model overestimated plume half-width at Desert Tortoise and Burro and showed relatively good agreement at Goldfish.

- DEGADIS: For Desert Tortoise, DEGADIS gave large overprediction of maximum concentrations at 100 m but relatively good agreement at 800 m. For Goldfish, it underpredicted maximum concentrations by roughly a factor of 1.5 at all distances. For Burro, it gave large overprediction at 57 and 140 m but showed smaller overprediction bias at 400 and 800 m. The model overestimated plume half-width at Desert Tortoise, provided relatively good agreement at Goldfish but underestimated at Burro, especially at the farthest distances.

- FOCUS: For Desert Tortoise, FOCUS overpredicted maximum concentrations at 100 m but provided good agreement at 800 m. For Goldfish, it predicted maximum concentrations with little bias at 300 m but underpredicted at both 1000 and 3000 m. For Burro, it produced overpredictions at 140, 400, and 800 m. FOCUS showed large underestimation of the plume half-width at Desert Tortoise, moderate underestimation at Goldfish, and large overestimation at Burro.

- SAFEMODE: For Desert Tortoise, SAFEMODE underpredicted maximum concentrations at 800 m by roughly a factor of 2. For Goldfish, it underpredicted at 300 and 1000 m but gave reasonable agreement (for one test) at 3000 m. For Burro, it overpredicted maximum concentrations consistently at all distances by as much as a factor of 10. For plume half-width, the model showed large overestimation at Goldfish and underestimation at Burro.

- SLAB: For Desert Tortoise, SLAB consistently underpredicted maximum concentrations by a factor of 2.5 at both 100 and 800 m. For Goldfish, it underpredicted maximum concentrations consistently by a factor of 3 at all distances. For Burro, it provided relatively good agreement with observed maximum concentrations at 140, 400, and 800 m with overprediction by a factor of 1.6 at 57 m. The model showed moderate underestimation at Desert Tortoise and Goldfish but relatively large overestimation at Burro, especially at the farthest distances.

- TRACE: For Desert Tortoise, TRACE performed well for predicting maximum concentrations at both the 100- and 800-m distances. For Goldfish, it underpredicted maximum concentrations at 300 m by a factor of 2 but showed less bias at 1000 and 3000 m. For
Burro, it performed relatively well for estimating maximum concentrations at 57 m but consistently overpredicted by more than a factor of 2 at 140, 400, and 800 m. The model overestimated the plume half-width at Desert Tortoise and Goldfish but underestimated at Burro.

None of the models exhibited consistently good performance (factor of 2) across the three databases and two distance categories. As a group, the models underpredict concentrations at Goldfish and tend to overpredict concentrations at Desert Tortoise and Burro. However, the combined performance across the three databases shows that all models are performing within a factor of 2. The models were noticeably more accurate in predicting plume half-width than in predicting the magnitude of centerline maximum concentrations, especially for the more distant (greater than 300 m) category.

d. Thought for future model evaluation protocols

In contrast to the earlier model evaluation studies, it is evident that dense gas models are relatively new, technically complex, and difficult to evaluate. Although relatively few comments were received at the test package finalization stage, some developers disagreed with how their model was applied after the evaluation was completed and results reviewed. These disagreements included whether to adjust the model to include cloud height above ground, use different release averaging times, respecify initial release conditions (e.g., whether the test was actually a liquid or an aerosol release, and what is the initial release velocity, etc.), include or exclude some tests in the experiments, compare maximum concentrations or hazard extent, place more emphasis on performance with close-in versus far-field sensors, adjust certain model inputs, and correctly interpret and apply the model in general.

Dense gas field experiments are complex and costly. The quantity of data obtained from an experiment is limited for purposes of model evaluation. All the model developers have access to the publicly available databases. Thus, the models could be “fit” to perform well for the available datasets. Dense gas releases are typically of short duration, while toxic
hazard assessments involve a range of time-averaged considerations. For dense gases, maximum concentrations typically occur near the ground, especially for near-field observations where shallow clouds are formed. The field observations used measurements made at approximately 1-m height. When compared with predictions that, for the DEGADIS and AIRTOX models, were made at ground level (zero elevation), the model estimate is likely to be higher than the observed value, especially for the near-field observations where shallow dense gas clouds are likely to be formed.

Dense gas models are very sensitive to the source-release characteristics. During a field experiment, an accurate detailed documentation or description of the release scenario is nearly impossible to obtain given the nature of these releases. Assumptions or judgment calls had to be made in order to properly apply each model to the test data. Each model developer had his/her own method for simulating each release. These techniques in some cases were built into the model.

There is no general agreement in the scientific community on which experiments should be included in an evaluation. This evaluation excluded field tests due to diversity of physical release conditions. For example, the Maplin Sands and Thorney Island experiments were excluded because the releases, respectively, occurred over water and were predominately “instantaneous.” These test data may prove valuable in future evaluations. All three experiments have been conducted over flat, unobstructed terrain, and thus do not include effects associated with urban conditions. There is also no general agreement on how model averaging time should be determined for these experiments.

Clearly, dense gas models are complex and require expert advice in order to accurately simulate a release. An equitable “hands-off” evaluation of dense gas models was very difficult to achieve as a practical objective. Some of the models have limited documentation and require considerable user experience for effective application. These models are typically very sensitive to the choices involved in source definition. The model developer is generally best able to understand the implications of those choices.

6. Summary and conclusions

The initial characterization of the release is critical for estimating concentrations from dense gas models. As a consequence, models that contain similar treatment of atmospheric dispersion and use the same meteorological inputs predict concentrations that can differ by several orders of magnitude in certain cases. Given the high cost of performing field experiments on dense gas releases, the databases are necessarily small. Therefore, it is uncertain how well any of these models might perform for any other dense gas release scenario.

This dense gas model evaluation study has provided a basis for objectively judging the performance of seven dense gas models as applied to three different experimental release scenarios. The Desert Tortoise and Goldfish experiments were simulated as horizontal jet releases, and Burro as a liquid pool release. All three experiments were simulated as continuous releases. Performance statistics were used to compare observed concentrations and plume half-widths to those predicted by each model. Model performance varied and no model exhibited a consistently good performance for all three databases. However, when combined across the three databases, all models performed within a factor of 2.

Disclaimer. The information in this document has been funded wholly or in part by the United States Environmental Protection Agency under an Interagency Agreement (DW47935932-01) to NOAA. It has been subjected to Agency review and approved for presentation and publication. Mention of trade names or
commercial products does not constitute endorsement or recommendation for use.

Acknowledgments. This paper is based on the results of a detailed report titled "Evaluation of Dense Gas Simulation Models" (Zapert et al. 1991). In addition to the authors of that report, all of the model developers who participated in the study provided their valuable time and resources. These are DEGADIS (J. Havens and T. Spicer), SLAB (D. Ermakers), AIRTOX (M. Mills), CHARM (M. Elgroth), FOCUS (J. Cornwell), SAFEMODE (P. Raj), and TRACE (E. Chikhliwala). Ms. Brenda Cannady (EPA) typed the manuscript.

REFERENCES


Quest Consultants, Inc., 1990: Focus Input and Reference Manuals. Quest Consultants. [Available from Quest Consultants, P.O. Box 721387, Norman, OK 73070.]


