EPA-450/4-90-018

EVALUATION OF DENSE GAS SIMULATION MODELS

By

James G. Zapert
Richard J. Londergan
Harold Thistle

TRC Environmental Consultants, Inc.
East Hartford, CT 06108

EPA Contract No. 68-02-4399

EPA Technical Representative: Jawad S. Touma

Office Of Air Quality Planning And Standards
Office Of Air And Radiation
U. S. Environmental Protection Agency
Research Triangle Park, NC 27711

May 1991

This report has been reviewed by the Office Of Air Quality Planning And Standards, U. S. Environmental Protection Agency, and has been approved for publication as received from the contractor. Approval does not signify that the contents necessarily reflect the views and policies of the Agency, neither does mention of trade names or commercial products constitute endorsement or recommendation for use.

EPA-450/4-90-018

ACKNOWLEDGEMENTS

This report was prepared for the U.S. Environmental Protection Agency by TRC Environmental Consultants, Inc. (TRC), East Hartford, Connecticut. Mr. Jawad S. Touma was the Technical Representative on this project for EPA. Mr. Raymond J. Topazio served as Project Manager. The principal authors were Dr. Richard J. Londergan, Mr. James G. Zapert and Dr. Harold Thistle. Technical assistance was provided by Mr. Michael A. Ratte and Mr. Sean Hayden.

TABLE OF CONTENTS

SECTION			PAGE
		DISCLAIMER	ii
		ACKNOWLEDGEMENTS	iii
		TABLE OF CONTENTS	v
		LIST OF TABLES	vii
		LIST OF FIGURES	ix
1.0		INTRODUCTION	1-1
	1.1	Description of the Study	1-2
2.0		MODEL SELECTION	2-1
	2.1	Public Domain Models	2-3
	2.1.1	DEGADIS	2-3
	2.1.2	SLAB	2-3
	2.2	Proprietary Models	2-4
	2.2.1	AIRTOX	2-4
	2.2.2	CHARM	2-4
	2.2.3	FOCUS	2-5
	2.2.4	SAFEMODE	2-5
	2.2.5	TRACE	2-6
3.0		DATA BASE DESCRIPTION AND SELECTION	3-1
	3.1	Desert Tortoise Pressurized Ammonia Releases	3-1
	3.2	Burro LNG Spill Tests	3-4
	3.3	Goldfish Anhydrous Hydrofluoric Acid Spill Experiments	3-10
4.0		EVALUATION METHODOLOGY	4-1
	4.1	Development of Test Packages	4-1
	4.2	Model Application (Input Assumptions)	4-3
	4.2.1	DEGADIS	4-5
	4.2.2	SLAB	4-6
	4.2.3	AIRTOX	4-8
	4.2.4	CHARM	4-11
	4.2.5	FOCUS	4-13
	4.2.6	SAFEMODE	4-15
	4.2.7	TRACE	4-17
	4.3	Model Application (Output Assumptions)	4-19
	4.3.1	DEGADIS	4-19
	4.3.2	SLAB	4-21
	4.3.3	AIRTOX	4-22
	4.3.4	CHARM	4-22
	4.3.5	FOCUS	4-23
	4.3.6	SAFEMODE	4-23
	4.3.7	TRACE	4-24
	4.4	Experiment Data Analysis	4-24
	4.5	Model Limitations	4-25
	4.6	Model Averaging Time	4-26
	4.7	Statistical Methods	4-26

TABLE OF CONTENTS
(Continued)

SECTION			PAGE
5.0		RESULTS OF THE MODEL EVALUATION	5-1
	5.1	Desert Tortoise Ammonia Releases	5-1
	5.1.1	Maximum Concentrations	5-1
	5.1.2	Cloud Half-Width	5-7
	5.2	Goldfish HF Releases	5-9
	5.2.1	Maximum Concentrations	5-9
	5.2.2	Cloud Half-Width	5-15
	5.3	Burro LNG Spill Experiments	5-15
	5.3.1	Maximum Concentrations	5-15
	5.3.2	Cloud Half-Width	5-22
	5.4	Inter-Model Comparison	5-24
	5.5	Summary of Model Performance	5-27
6.0		CONCLUSIONS	6-1
7.0		REFERENCES	7-1

APPENDIX

A	STATISTICAL PROTOCOL FOR EVALUATION OF AIR TOXICS MODELS

LIST OF TABLES

TABLE		PAGE
2-1	Toxics Models Considered for the Evaluation Study .	2-2
3-1	Burro Series Test Summary (1980)	3-7
4-1	Potential Significant Release Characteristics . . .	4-4
4-2	"DEGADIS" Input Parameters	4-7
4-3	"SLAB" Input Parameters	4-9
4-4	"AIRTOX" Input Parameters	4-12
4-5	"CHARM" Input Parameters	4-14
4-6	"FOCUS" Input Parameters	4-16
4-7	"SAFEMODE" Input Parameters	4-18
4-8	"TRACE" Input Parameters	4-20
4-9	Performance Measures for the Dense Gas Model Evaluation Study	4-28
5-1	Performance Statistics for Maximum Concentration Values for Desert Tortoise NH_3 Experiments . . .	5-6
5-2	Performance Statistics for Cloud Half-Width for Desert Tortoise NH_3 Experiments	5-8
5-3	Performance Statistics for Maximum Concentration Values for Goldfish HF Experiments	5-14
5-4	Performance Statistics for Cloud Half-Width for Goldfish HF Experiments	5-16
5-5	Performance Statistics for Maximum Concentration Values for Burro LNG Experiments	5-21
5-6	Performance Statistics for Cloud Half-Width for Burro LNG Experiments	5-23
5-7	Average Fractional Bias of Maximum Observed and Predicted Values by Model and by Distance Category	5-25
5-8	Summary Chart of Model Predicted and Observed Centerline Concentrations, and Fractional Bias .	5-28

LIST OF FIGURES

FIGURE		PAGE
3-1	Release Configuration for the Desert Tortoise Ammonia Experiments	3-2
3-2	Desert Tortoise Ammonia Test Grid	3-5
3-3	Site Plan of Naval Weapons Center (NWC) Spill Facility at China Lake	3-8
3-4	Network Configuration for the Burro LNG Tests . . .	3-9
3-5	HF Release Configuration	3-11
3-6	HF Sampling Grid	3-13
5-1	Observed and Predicted Maximum Concentrations versus Downwind Distance for Desert Tortoise NH_3 Test 1	5-2
5-2	Observed and Predicted Maximum Concentrations versus Downwind Distance for Desert Tortoise NH_3 Test 2	5-3
5-3	Observed and Predicted Maximum Concentrations versus Downwind Distance for Desert Tortoise NH_3 Test 4	5-4
5-4	Observed and Predicted Maximum Concentrations versus Downwind Distance for Goldfish HF Test 1	5-10
5-5	Observed and Predicted Maximum Concentrations versus Downwind Distance for Goldfish HF Test 2	5-11
5-6	Observed and Predicted Maximum Concentrations versus Downwind Distance for Goldfish HF Test 3	5-13
5-7	Observed and Predicted Maximum Concentrations versus Downwind Distance for Burro LNG Spill Test 3	5-17
5-8	Observed and Predicted Maximum Concentrations versus Downwind Distance for Burro LNG Spill Test 5	5-19
5-9	Observed and Predicted Maximum Concentrations versus Downwind Distance for Burro LNG Spill Test 8	5-20

1.0 INTRODUCTION

The U.S. Environmental Protection Agency (EPA) has an ongoing program to evaluate the performance of several categories of air quality simulation models by comparing observed and predicted concentrations using performance measures recommended by the American Meteorological Society (AMS). Rural, urban, mobile, complex terrain, and long range transport models are categories of models that have already been evaluated. Models for toxic pollutant releases represent a broad class of models for which little evaluation work has previously been performed.

Releases of toxic chemicals to the atmosphere can involve complex source dynamics and chemistry. A toxic chemical release can be modeled using the procedures developed for criteria air pollutants, provided that the gas is not dense or highly reactive and it does not rapidly deposit on surfaces. In order to select an appropriate modeling approach for a toxic chemical, the release must first be categorized in terms of physical state, release condition, and dispersive characteristics (McNaughton and Bodner, 1988).

An understanding of the process/release condition of a toxic release is required to model the atmospheric dispersion of a toxic chemical. This condition can help determine both the physical state and dispersive characteristics of the released chemical. For example, if the release is from a leak in a pressurized liquefied gas storage tank, additional source modeling is required to determine the state of the material as it enters the atmosphere, since the release may include both liquid and gaseous (aerosol) components. In addition to the source term and initial dispersive characteristics, air toxics models must also provide proper simulation of the cloud characteristics downwind of the release. The cloud characteristics are often complex when dense, highly reactive or rapidly depositing chemicals are released.

Toxic chemical releases are often of short duration and the concentrations of interest are near instantaneous averages. Typical concerns from a toxic release are the maximum instantaneous concentration and the maximum dosage. Many toxics models are designed to provide concentration predictions for unit averaging times ranging from 0.1 seconds to 1 hour. By contrast, regulatory models for most criteria pollutants have a basic averaging time of one hour for concentration estimates.

1.1 Description of the Study

This report describes the approach and presents the results of an evaluation study performed for several models capable of simulating dense gas releases. Dense gas releases represent a subset of toxic release scenarios. Models for simulating dense gas releases need to account for the source term, initial gravitational spreading of a heavy gas cloud, and the downwind dispersion of the cloud in air.

For this study two public domain models (DEGADIS and SLAB) and five proprietary models (AIRTOX, CHARM, FOCUS, SAFEMODE, and TRACE) were evaluated against the data from three experimental programs. The data bases include controlled releases of ammonia (Desert Tortoise), liquefied natural gas (Burro), and hydrofluoric acid (Goldfish).

A discussion of the model selection criteria and description of each model are presented in Section 2, while Section 3 describes each data base and how tests were selected for this evaluation. In Section 4 the methodology for the evaluation study is presented. Also, the development of model inputs, interpretation of model outputs, selection criteria for receptor locations, and statistical methods are discussed. The results of the evaluation are presented in Section 5, followed by conclusions in Section 6.

2.0 MODEL SELECTION

The models used in the evaluation study were initially selected from lists and surveys performed by McNaughton et al. (1986), Worley et al. (1986), and Hanna and Drivas (1987). Models that were considered appropriate for determining the impact of routine (non-accidental) releases of toxic pollutants are listed in Table 2-1. EPA attempted to contact the developers of these models to solicit interest for an evaluation study. Each model developer was presented with the objectives of the evaluation study and a list of candidate data bases involved. The developers were given 30 days to respond with an expression of interest.

The invitations sent to the model developers brought a number of favorable responses, but not all of the models from Table 2-1 have been included in the present study. Some models were judged not applicable to dense-gas release. For others, the developers declined to participate, or TRC and EPA were unable to identify a person or institution to provide the technical support which this study required. From the original list, 3 public domain (DEGADIS, HEGADAS, and SLAB) and 6 proprietary models (AIRTOX, CHARM, EAHAP, MESOCHEM, SAFEMODE and TRACE) were eventually selected. TRC contacted those model developers in order to initiate the model acquisition process. For each proprietary model, a confidentiality agreement was established between the developer and TRC before the model documentation and software were provided to TRC. Under this agreement, TRC is to return all the material provided after the conclusion of the study.

During preparation of test packages, several changes to the list of models were made. Seven models were evaluated using experimental data. Of the public domain models, DEGADIS 2.1 and SLAB were evaluated. HEGADAS was excluded from the evaluation after the model developers stated that HEGADAS, in its present state, should not be applied to either the Desert Tortoise or

TABLE 2-1

TOXICS MODELS CONSIDERED FOR THE EVALUATION STUDY

Model	Developer	Proprietary Rights
AFTOX	U.S. Air Force	Public
AIRTOX	ENSR Corporation	Private
ALOHA	NOAA	Public
AVACTA	Aerovironment, Inc.	Public
CHARM	Radian Corporation	Private
CHEMS-PLUS	Arthur D. Little	Private
DEGADIS	U.S. Coast Guard and Gas Research Institute	Public
EAHAP	Energy Analysts, Inc.	Private
HEGADAS	Shell Development Company	Public
MESOCHEM	Impell Corporation	Private
OME	Ontario Ministry of the Environment	Public
SAFEMODE	Technology and Management Systems, Inc.	Private
SLAB	Lawrence Livermore National Laboratory	Public
TRACE	E.I. DuPont De Nemours & Company	Private

Goldfish tests. Of the six proprietary models, four were eventually evaluated: AIRTOX, CHARM Version 5, SAFEMODE and TRACE II. In addition, a fifth proprietary model (FOCUS Version 1.0) was included. The MESOCHEM model was excluded from the evaluation at the model developer's request, since the developer indicated that MESOCHEM was not applicable for this type of study. The model EAHAP was removed from the evaluation due to a lack of support from the model developer. When EAHAP was withdrawn, Quest Consultants, Inc., whose engineers and scientists developed EAHAP, offered a new model named FOCUS.

2.1 Public Domain Models

2.1.1 DEGADIS

The Dense Gas Dispersion (DEGADIS) model was originally developed for the U.S. Coast Guard and the Gas Research Institute to simulate the dispersion of accidental or controlled releases of hazardous liquids or gases into the atmosphere.

DEGADIS Version 2.1 was used for this evaluation. It includes the Ooms module for predicting the trajectory and dilution of an elevated dense gas jet. DEGADIS currently 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 to the model 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.

2.1.2 SLAB

The SLAB model was developed by Lawrence Livermore National Laboratory to simulate the atmospheric dispersion of denser-than-air releases. SLAB models

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, averaged for a user specified time period, and information to categorize the cloud width distance. SLAB does not contain an internal chemical library, but the User's Guide provides the necessary parameters for many of the chemicals of interest.

2.2 Proprietary Models

2.2.1 AIRTOX

AIRTOX has been developed by ENSR Consulting and Engineering to calculate downwind concentrations from time dependent toxic chemical releases to the atmosphere. Chemical releases are simulated by AIRTOX in either a jet or non-jet mode. AIRTOX is a spreadsheet based model that utilizes Lotus 1-2-3 software. Chemical properties are provided automatically through an internal data base. 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. Centerline concentrations are provided 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 minute averaged dispersion coefficients.

2.2.2 CHARM

The Complex Hazardous Air Release Model (CHARM) is a Gaussian puff model created by Radian Corporation to assess the location, extent, and concentration of the cloud which results from the release of a toxic substance to the air. CHARM includes a chemical data base that provides all of the

necessary chemical parameters to the model. CHARM is a menu-driven system comprised of two parts: CHARM1 and CHARM2. CHARM1 contains all of the screens for data input, while CHARM2 performs the calculations for the evolving cloud and controls the output. CHARM can be used in a "planning mode" or as part of an "emergency response system". 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. This display provides a "snapshot" of the cloud passage with time. Options are provided with the snapshot display to produce concentration/dosage information for the release. The concentration information represents instantaneous values while dosage represents time-averaged (user specified) values.

2.2.3 FOCUS

FOCUS is a hazards analysis software package that was designed by Quest Consultants Inc. to evaluate transient hazards resulting from accidental or controlled releases of hazardous liquids or gases. 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 a logic control module which determines the sequence of programs to be executed, with periodic input from the user.

The FOCUS model provides downwind centerline concentrations as a function of time since release and the lateral distance to three user-specified concentration limits. In addition, FOCUS outputs information regarding the release profile and pool characteristics. The predicted concentrations from FOCUS represent values averaged over the release duration.

2.2.4 SAFEMODE

The Safety Assessment for Effective Management of Dangerous Events (SAFEMODE) model 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 as selected distances downwind of the release location. The user can specify concentration averaging time. SAFEMODE has an internal chemical property library for over 100 common chemicals.

2.2.5 TRACE

The TRACE model was developed by E.I. DuPont De Nemours & Company as a tool to evaluate the potential impact of toxic chemical spills into the environment. TRACE is an interactive, menu driven model that allows the user multiple options when developing a release scenario. TRACE contains an extensive chemistry 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.

3.0 DATA BASE DESCRIPTION AND SELECTION

The data bases chosen for the evaluation study were selected from a set of data bases archived into EPA's Model Evaluation Support System (MESS) (Zapert and Londergan, 1990). MESS includes six air toxics data bases: Goldfish Anhydrous Hydrofluoric Acid Spill Experiments, Burro Liquefied Natural Gas Experiments, Desert Tortoise Liquid Ammonia Experiments, Hanford Instantaneous Tracer Experiments, Thorney Island Heavy Gas Dispersion Experiments, and Washington State University Isoprene Flux Experiments. Data bases for Goldfish, Burro and Desert Tortoise were selected for this evaluation since each of these programs involved dense gas releases that have similar source scenarios (continuous releases).

3.1 Desert Tortoise Pressurized Ammonia Releases

Four large scale pressurized liquid ammonia experiments were conducted in 1983 at the Liquefied Gaseous Fuels Spill Test Facility in Nevada (Goldwire et al., 1985). The releases were conducted by Lawrence Livermore Laboratory with the sponsorship of the U.S. Coast Guard, the Fertilizer Institute, and Environment Canada. Four high volume, high pressure spill releases were made, ranging from 15 to 60 m^3 over time periods of 1 to 8 minutes. Figure 3-1 shows the release configuration for the ammonia spill tests. Ammonia tanks were used to feed a six inch pipeline leading to the spill point. Ammonia was released at elevated (storage) pressure and ambient temperature with a nitrogen system used to provide constant tank pressure. The actual release was made through an orifice plate at the end of the spill pipe at a height of 0.79 m. The jet release was directed horizontally down the grid. An analysis of time of arrival of the gas cloud as indicated by the temperature and concentrations time series on both the 100 m and 800 m arcs indicated that the

FIGURE 3-1
RELEASE CONFIGURATION FOR THE DESERT TORTOISE AMMONIA EXPERIMENTS

evidence of a momentum jet (i.e., cloud speed in excess of wind speed) at 100 m in all tests and at 800 m for tests 2 and 4.

Release sizes and rates are as follows:

Test	Amount (m^3)	Rate (m^3/min)
1	14.9	7.0
2	43.8	10.3
4	60.3	9.5

Release pressures were approximately 150-170 psia at temperatures of 20-24°C.

Additional release characteristics for three tests are available in the MESS Archive; these include flow rates, temperatures and pressures near the release point, and liquid flow rates and temperature measurements in the soil designed to evaluate pooling. Liquid volume release rate was converted to mass release rate (g/s) for archiving.

Several assumptions were made regarding the release configuration for the evaluation. The release configuration resulted in a complex two-phase, horizontal jet with substantial momentum. Documentation indicates that some liquid pooling was observed near the release point, but pool characteristics were not reported. As a result, the release configuration for the archive assumes a gaseous release of a heavy gas, with unspecified initial cloud dimensions. For the archive, emissions were estimated using mass flux estimation techniques using observed concentration data at 800 m downwind of the release point. Only 70 percent of mass could be accounted for with this technique.

The three tests were conducted for similar meteorological conditions. Tests were performed under D or E stability conditions in moderate winds of 4.5 to 7.4 m/s. Ambient temperatures ranged from 28.8 to 33.7°C. Wind measurements taken at 2 m height and averaged for test duration were included in the archive. The site is a desert location on a normally dry lake bed.

Surface roughness for the site was reported as 0.3 cm, but various portions of the sampling grid were covered with water in three of the four tests, due to unusual rainfall in the area. Humidity on the site could be assumed high due to evaporation of the lake, but humidity and pressure measurements are not available at the site.

Ammonia concentrations were sampled on a downwind grid as shown in Figure 3-2. At 100 m downwind, gas samplers were located 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 crosswind separation. Further downwind, eight portable sensors were used to sample concentrations at either 1400 m, 2800 m or 5500 m at a height of 1 m. Concentrations were averaged to obtain 30 second values for the MESS archive.

3.2 Burro LNG Spill Tests

The Burro Series of liquefied natural gas (LNG) spill experiments (Koopman et al., 1982) were performed at the Naval Weapons Center, China Lake, California in the summer of 1980. A total of eight spills of LNG onto water were made. The volume of LNG released varied from 24 to 39 m^3 at spill rates of 11 to 18 m^3/min. Concentration measuring devices were located at radii of 57, 140, 400, and 800 m from the source. The meteorological data included wind, turbulence and temperature measurements to describe the turbulent atmospheric boundary layer.

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 with a slight slope (less than 1 degree) north to south, left to right for cloud travel.

Of the eight Burro tests conducted, Nos. 3, 5 and 8 were selected for inclusion in the MESS Archive. Tests 6 and 9 were excluded due to several

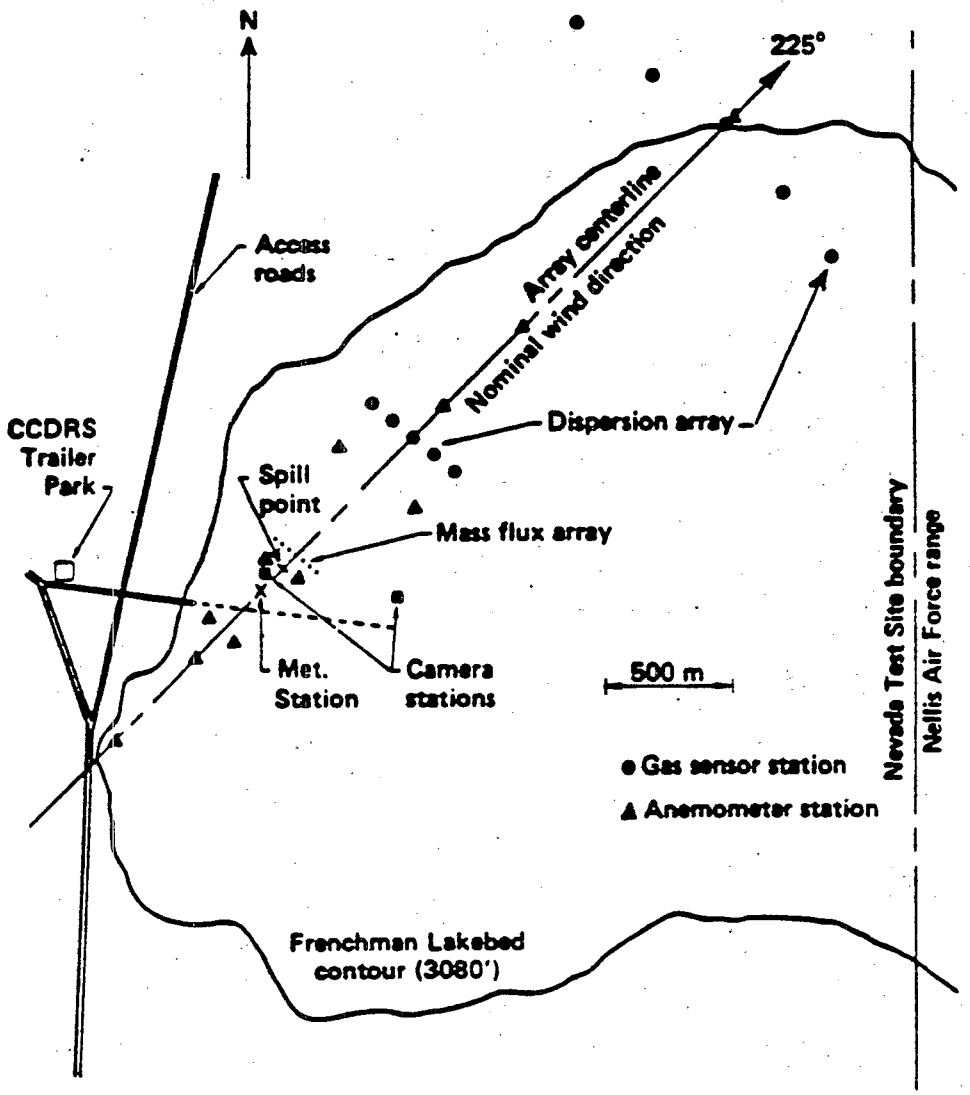


FIGURE 3-2

DESERT TORTOISE AMMONIA TEST GRID

rapid phase transitions, or small explosions, which occurred during the releases. Burro tests 4 and 7 appeared to have centerline concentrations outside the grid and were not included. Test 2 was excluded, on the advice of Lawrence Livermore Laboratory staff, because of data uncertainties.

Test conditions are summarized in Table 3-1. Wind speeds averaged from 5.4 to 7.4 m/sec for the unstable and slightly unstable atmospheric conditions in tests 3 and 5. Wind speed averaged only 1.8 m/sec for test 8, with slightly stable conditions.

The LNG was released from a cryogenic liquid storage tank. A 25 cm diameter stainless steel spill line ran from the tank to the center of a 58 m diameter spill pond filled with water to a depth of approximately 1 m. The water level was 1.5 m below the surrounding ground level. The spill pipe was directed straight down toward the water with a splash plate installed below the spill pipe outlet at a shallow depth beneath the water surface to limit penetration of the LNG into the water. Consequently, after the LNG stream encountered the water, it was directed radially outward along the surface of the water. The release configuration is shown in Figure 3-3. Little information is available to accurately define the liquid pool area, although, for modeling, total pool flux is assumed to be equal to release rate as pool spreading and vaporization reach equilibrium with the release rate. Spill rates were close to 12 m^3/min for Burro tests 3 and 5, then increased to 16.0 m^3/min for test 8.

Gas concentration data were measured at heights of 1, 3 and 8 m above the ground at distances of 57, 140, 400 and 800 m from the source. A total of 30 stations recorded gas concentrations. Network configuration is depicted in Figure 3-4. Gas concentration data were originally recorded at rates of 1 to 5 Hertz. All concentration data were averaged to 10 seconds for the MESS archive.

TABLE 3-1

BURRO SERIES TEST SUMMARY (1980)

Test Name	Date	Spill Volume (m^3)	Spill Rate (m^3/min)	Averaged Wind Speed (m/s)	Averaged Wind Direction (degrees)	Atmospheric Stability
Burro 3	2 July	34.0	12.2	5.4	224	Unstable
Burro 5	16 July	35.8	11.3	7.4	218	Slightly Unstable
Burro 8	3 Sept.	28.4	16.0	1.8	235	Slightly Stable

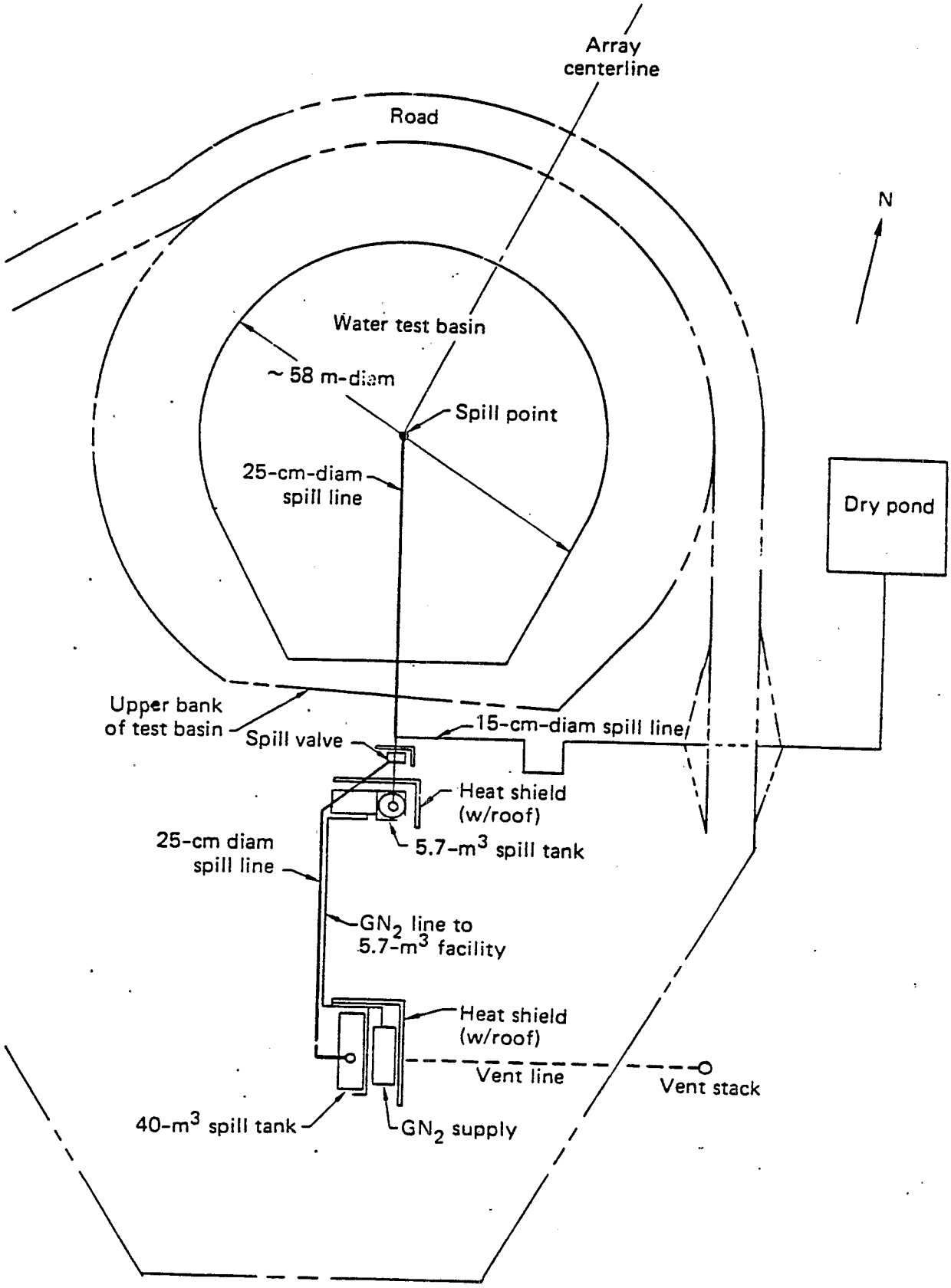


FIGURE 3-3 Site plan of Naval Weapons Center (NWC) spill facility at China Lake.

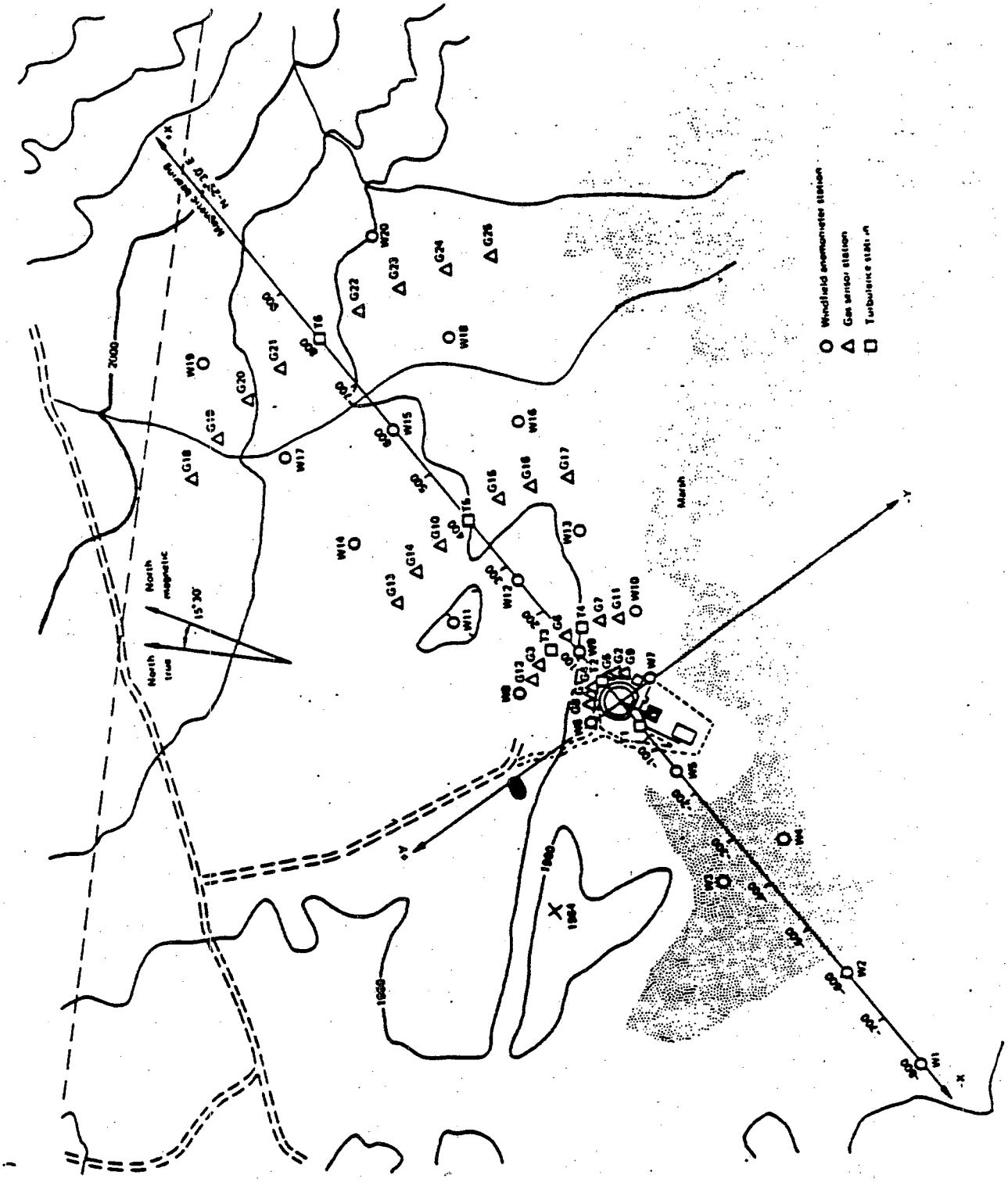


FIGURE 3-4

NETWORK CONFIGURATION FOR THE BURRO LNG TESTS

Meteorological data were collected using standard cup and vane anemometers 2 m above the ground, located at 20 stations upwind and downwind of the spill source. Wind data from the 20 cup and van anemometers were averaged to obtain single values of wind speed and direction for the MESS archive. An average value for the test duration for temperature, humidity, stability, and Monin-Obukhov length were taken from the Burro data report.

3.3 Goldfish Anhydrous Hydrofluoric Acid Spill Experiments

In 1986, AMOCO Oil Company and Lawrence Livermore Laboratory conducted six experiments (Blewitt et al., 1987) to study atmospheric releases of anhydrous hydrofluoric acid from heated, pressurized storage (40°C, 6.8 atm). Three of the six tests (Tests 1-3) were designed primarily to study vaporization and aerosol generation, cloud density and dispersion. The other three were designed to study the effect of water sprays as mitigation measures in the event of a release. Tests 1-3 are included in the MESS archive.

Releases were made as a horizontal liquid jet from a spill pipe. Concentrations were sampled at multiple vertical levels on three sampling arcs. Sampling distances were sufficient to record concentrations at trace levels. In one of the archived dispersion experiments (Test 3), additional moisture was added to the air upwind of the source using a combination of a pond and steam generators.

Liquid anhydrous hydrofluoric acid was spilled at the Liquefied Gaseous Fuels Spill Test Facility, using a release system similar to the Desert Tortoise ammonia tests. Figure 3-5 shows the release configuration for the HF spill tests. The HF releases were made from a tank truck at constant pressure maintained with a nitrogen purging system. From the tank, the HF was carried under pressure by pipe to the spill point. The HF exited the pipe

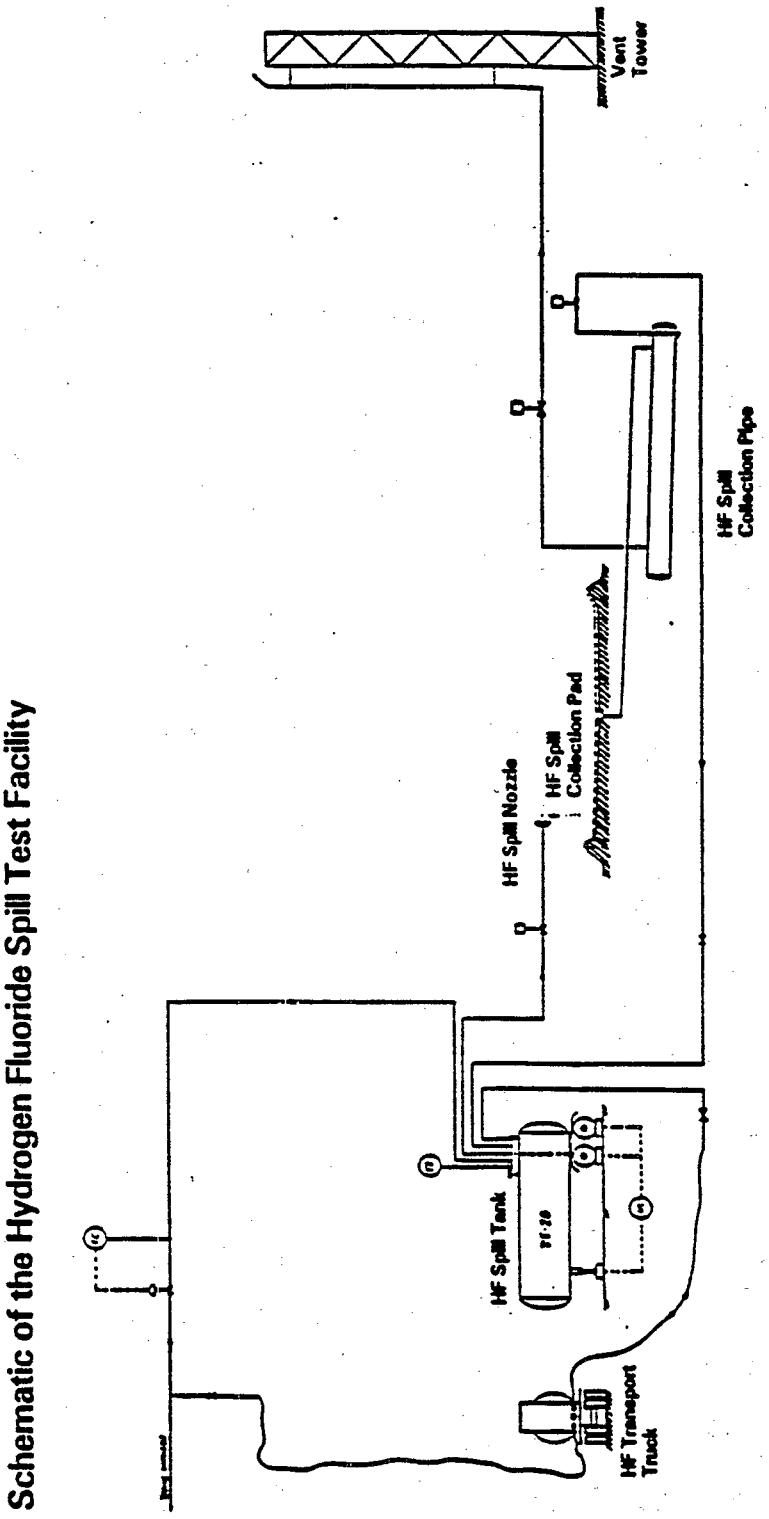


FIGURE 3-5 HF RELEASE CONFIGURATION

horizontally through an orifice plate, 1 m above the surface of an impermeable spill pad. In addition, the spill pad was designed to collect any pooling HF. In the tests conducted, pooling was not observed. The spill rate data chosen for the archive were obtained by performing a linear regression of the load cell data. The horizontal jet was directed downwind. HF release rates for the three experiments were 1.78 m^3/min, 0.65 m^3/min and 0.66 m^3/min, with release times of 2 minutes, 6 minutes, and 6 minutes, respectively.

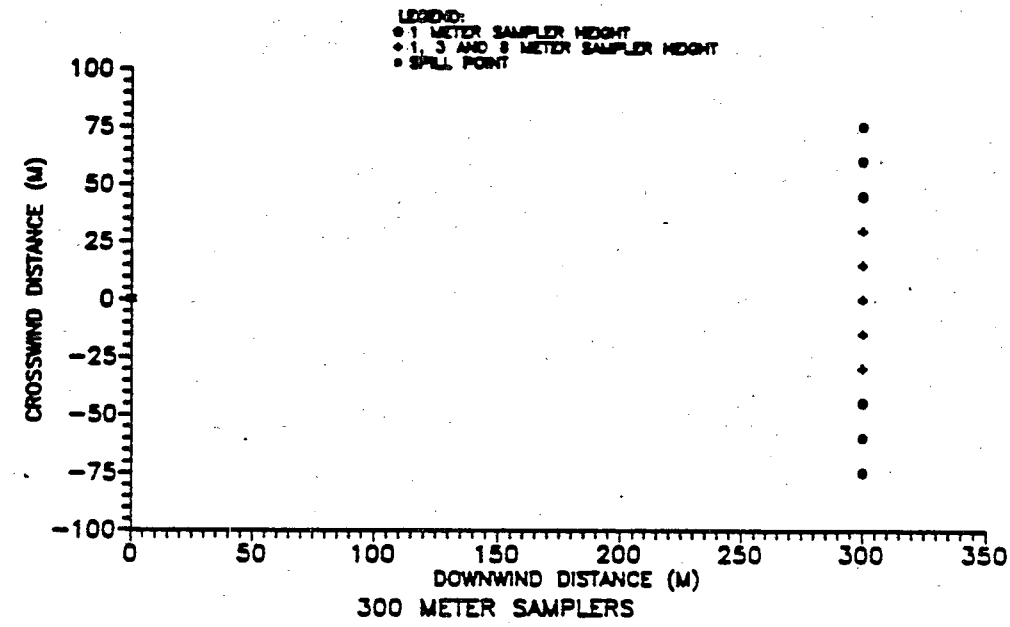
The release configuration was designed to provide reliable source estimates for dispersion testing. To this end, source definition was a major consideration. Three approaches were used to quantify emissions: 1) load cells, 2) mass flux, and 3) orifice calculations.

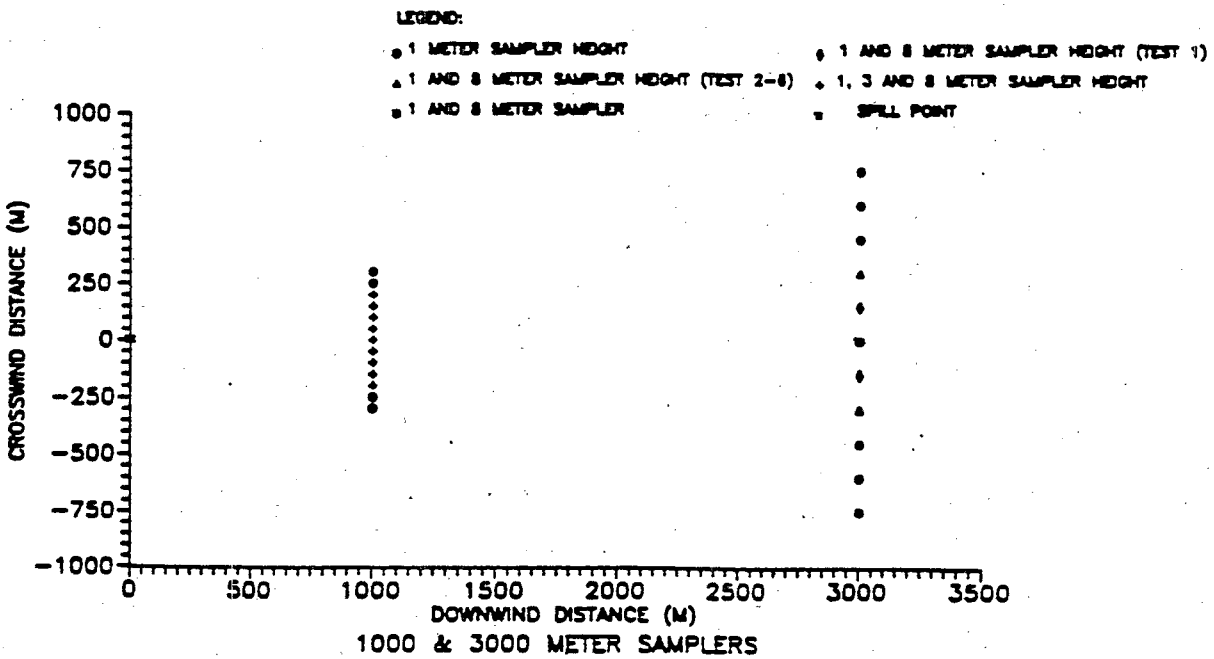
Release characteristics are as follows:

Test	HF Spill Rate (m^3/min)	Duration (sec)	HF Temperature (°C)	HF Pressure (psig)
1	1.78	125	40	111
2	0.65	360	38	115
3	0.66	360	39	117

Liquid volume release rate was converted to mass release rate (g/s) for archiving.

Concentration data collected included measurements on three sampling arcs (300, 1,000, 3,000 m) shown in Figure 3-6. Measurements were made at 1, 3, and 8 m above the ground surface. For test 1, cloud transport and dispersion was centered on the grid axis resulting in steady-state plumes at all three arc distances. In test 2, the peak concentration at 3,000 m was measured at the edge of the sampling array. This makes it impossible to determine the maximum concentration or plume width. In test 3, sensor problems led to inadequate plume coverage at the 3,000 m arc.





PLOT PLAN OF THE HF SAMPLERS

FIGURE 3-6

HF SAMPLING GRID

Averaging times for concentration measurements were 66.6 seconds, 83.3 seconds, or 100 seconds, depending on the sample's location and the test conditions.

Wind measurements at a height of 2 m were made at 18 sites ranging from 2 km upwind to 3 km downwind of the release. The archived meteorological data represent test-average values as provided by Blewitt et al. (1987). The atmospheric stability was neutral for the three tests. Wind speeds ranged from 4.2 to 5.6 m/s and ambient temperature from 26.5 to 37.0°C.

4.0 EVALUATION METHODOLOGY

The models and data bases included in this evaluation study are diverse and vary in complexity. The study involves statistical comparison of predictions from each model with the experimental data. TRC attempted to define common variables for each model and each data base. Since all of the experiments involved continuous releases, TRC decided to use the release duration as the averaging time for predictions. The release period represents the shortest averaging time of interest for estimating concentrations given the model assumptions and test-period average meteorological input.

The averaging times for measured concentrations were also set to approximate the release duration in each experiment. The values used are multiples of the basic time unit used in the experimental data archive.

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 for each arc is defined as the lateral distance between the maximum concentration and the location at which concentrations have decreased to 50 percent of the maximum. For measured concentrations, the two half-width values on either side of the maximum were averaged. A more detailed discussion of evaluation statistics is provided in Section 4.7 and in Appendix A, "Statistical Protocol for Evaluation of Air Toxics Models".

4.1 Development of Test Packages

The dense gas model evaluation study required an understanding of each model in order to properly apply that model to the experimental data bases. 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 of the models, the user interface presented a serious obstacle to realistically simulating the experimental releases. TRC relied heavily on the model developers for advice in these situations.

The model developers provided a model at the beginning of the evaluation and this model was used throughout the evaluation. The developers were not allowed to customize their models to simulate the different release scenarios. Many of the model developers do routinely customize their models based on the type of release which is being simulated.

The sophistication of the chemical libraries in the models ranged from non-existent to extensive. TRC did not evaluate the model chemical libraries. An indication of how a model handles a given chemical is given by the input streams. Whether a model treats a plume as non-reactive or reactive, or whether it allows chemical transformation after release, or whether the model ought to do these things can only be surmised from the model output and model accuracy.

After TRC received the models from the developers, test packages were developed for each model using one test from each experimental data base. In several cases model developers enclosed test cases with their model which 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 was mailed to EPA for review, and then to the respective model developer. 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.

4.2 Model Application (Input Assumptions)

Each model used in the evaluation study required a unique set of input parameters for each experimental data base. Model inputs were obtained either directly from the MESS Archive and experimental data reports, or estimated based on available information. The models involved vary in applicability, complexity and diversity, causing a situation where decisions concerning model inputs had to be made on a model by model basis. Whenever possible, values chosen for input variables were consistent between models. Model inputs for each experimental test and the physical assumptions for deriving the inputs to each model are described below.

A list of potentially significant release characteristics for each experiment is presented in Table 4-1. Table 4-1 is subdivided into two sections; the first section presents test attributes that are commonly addressed by dense gas models and the second addresses features that are simulated by only the more sophisticated models. Data to characterize each of these attributes are provided to the models either through user input, or are internally calculated in the case of the more sophisticated models. For the models involved in this evaluation, only DEGADIS attempts to simulate the surface layer mixing of LNG and water for the Burro releases, and FOCUS is the only model that allows the distinction between substrate types for the release site and the surface over which the vapor cloud is dispersed.

Meteorological inputs that were common to all models included wind velocity, relative humidity, ambient temperature and Pasquill stability class. Other meteorological variables used by one or more models include substrate or surface temperature, ambient pressure and solar radiation.

Site characteristics which have a bearing on local dispersion conditions include surface roughness (an aerodynamic surface characteristic), albedo (surface reflectivity), soil conductivity and soil thermal diffusivity.

TABLE 4-1

POTENTIAL SIGNIFICANT RELEASE CHARACTERISTICS

DT	Goldfish	Burro
A. Release Characteristics Considered in Three or More Dense Gas Models		
• Pressurized release • Horizontal jet • Aerosolization • Potential pooling	• Pressurized release • Horizontal jet • Aerosolization	• Evaporating pool • Cryogenic release
B. Release Characteristics Simulated by One or Two Dense Gas Models		
• Reaction with ambient water vapor	• Reaction with ambient water vapor • Chemical transformation	• Boiling liquid • Release onto water

4.2.1 DEGADIS

DEGADIS model inputs for Desert Tortoise were prepared following the recommendations of Spicer and Havens in their article "Development of Vapor Dispersion Models for Non-Neutrally Buoyant Gas Mixtures-Analysis of TFI/NH3 Test Data" (Spicer and Havens, 1988a). The DEGADIS User's Guide was used to prepare input specifications for Burro. For Goldfish, the articles "Conduct of Anhydrous Hydrofluoric Acid Spill experiments" (Blewitt et al., 1987) and "Modeling HF and NH3 Spill Data Using DEGADIS" (Spicer and Havens, 1988b) provided the most relevant guidance. DEGADIS simulates all of the release characteristics provided in Table 4-1A. Considerations regarding the momentum jet for the Desert Tortoise and Goldfish tests are described below. It is not apparent whether DEGADIS simulates HF chemical transformations in a vapor cloud; however, it does model surface mixing of a boiling liquid on water as stated earlier.

The Desert Tortoise and Goldfish tests were modeled by DEGADIS in a similar manner. Both experiments are characterized as jet releases. DEGADIS Version 2.1 models only vertical jet releases, but previous analyses (Spicer and Havens, 1988a, 1988b) suggested that the jet momentum had a negligible influence on predicted concentrations for these experiments.

DEGADIS requires as input ordered triples consisting of mole fraction, concentration of chemical, and air/chemical density. Spicer and Havens calculated the density of an ammonia/air/water mixture as a function of ammonia mole fraction using the TRAUMA model. They used a similar procedure for hydrogen fluoride. TRAUMA accounts for release thermodynamics by creating the ordered triplets which simulate thermodynamic effects. The Desert Tortoise and Goldfish releases were simulated as steady-state, pure chemical releases. The jet releases are modeled as "isothermal" because the TRAUMA module has already accounted for heat exchange by generating ordered triplets

which are input to DEGADIS. Source radius for the jet releases is larger than the orifice area and represents an initial puff size as suggested by the model developer.

The Burro experiments were simulated by DEGADIS using the assumption that the releases were non-isothermal, steady-state spills of LNG. Water transfer effects were included. The LNG pool size was only measured during Burro Test 9. For each Burro test included in the evaluation, pool area is assumed to have the same proportionality to release volume as existed in Test 9.

A tabulation of model inputs for DEGADIS is provided in Table 4-2.

4.2.2 SLAB

Inputs for the SLAB model were derived primarily from the MESS Archive and experiment reports, and from the SLAB User's Guide. SLAB considers evaporating pool releases, horizontal jet releases, vertical jet or stack releases and instantaneous or short duration evaporating pool releases. All the values for the chemical parameters were provided by the User's Guide, and there is no internal chemical library. SLAB considers all of the release characteristics for all three data bases specified in Table 4.1A. The Desert Tortoise and Goldfish experiments were modeled by SLAB using similar assumptions. Both cases were run as horizontal jets in order to properly characterize the aerosol effects. Aerosol effects were accounted for by specifying an initial liquid mass fraction of 0.83 for ammonia and 0.8 for hydrogen fluoride. For the Burro tests, the releases are simulated as evaporating pool releases, with the release temperature specified as the boiling point temperature of LNG. For the Burro releases, initial mass fraction is set to zero, since the releases are evaporating pools. Source areas for the Burro tests were computed using the formula:

$$A(s) = q(s)/p(s)w(s)$$

TABLE 4-2
"DEGADIS"
Input Parameters

PARAMETER	DT1	DT2	DT4	G1	G2	G3	B3	B5	B8
Wind Velocity (m/s)	7.42	5.76	4.51	5.60	4.20	5.40	5.40	7.40	1.80
Ambient Temperature (K)	302.39	303.80	305.95	310.15	309.15	299.65	306.95	313.65	306.25
Relative Humidity (%)	13.2	17.5	21.3	4.9	10.5	27.6	5.2	5.8	4.6
Emission Rate (kg/s)	81	117.3	107.9	30.2	10.4	10.4	86.4	79.4	112.6
Surface Roughness (m)	0.003	0.003	0.003	0.003	0.003	0.003	0.0002	0.0002	0.0002
Pasquill Stability Class	D	D	E	D	D	D	B	C	E
Averaging Time (sec)	120	240	360	66.6	333/353.2	333/353.2	160	180	100
Reference Height (m)	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
Monin-Obukhov Length(m)	92.7	94.7	45.2	10000.0	10000.0	10000.0	-9.1	-25.5	16.5
Pressure (atm)	0.897	0.898	0.891	0.891	0.900	0.900	0.936	0.929	0.929
Isothermal (yes,no)	Y	Y	Y	Y	Y	Y	N	N	N
Initial Chemical Temp (K)	302.39	303.80	305.95	310.15	309.15	299.65	111.70	111.70	111.70
Ordered Triples (yes,no)	Y	Y	Y	Y	Y	Y	N	N	N
Source Radius (m)	1.45	1.45	1.45	5.00	5.00	5.00	18.00	17.30	20.50
Surface Temp (K)							308.6	314.9	305.8

where:

q(s) = release rate (kg/s)
p(s) = density of LNG (kg/m^3)
w(s) = liquid regression rate (4.2E-04 for LNG)

following the advice of the model developer.

A tabulation of model inputs for SLAB is provided in Table 4-3.

After reviewing the draft final report, the SLAB model developer recommended one change to the SLAB input streams for the Goldfish and Desert Tortoise tests. In the case of a jet release, the cross sectional area of the fully expanded jet should be used as a source area, as opposed to the orifice area which was used by TRC. This modification would greatly reduce the source velocity and could significantly alter model results. This change was not implemented, since it was identified only after performance results had been obtained.

4.2.3 AIRTOX

Data inputs for AIRTOX were obtained primarily from the MESS Archive and experiment data reports, through user documentation, and from conversations with the model developer. AIRTOX can be run in jet or non-jet mode. AIRTOX simulates all of the release characteristics listed in Table 4.1A, except potential pooling for the jet releases. AIRTOX has an on-line chemical library, but it may not simulate the HF chemical effects listed in Table 4.1B. AIRTOX requires a release profile to describe chemical emissions. This profile allows emission rates to vary with time. Meteorological conditions are also allowed to vary with time. For this evaluation the emission rates and meteorology data were held constant. AIRTOX was run in the jet mode for both the Desert Tortoise and Goldfish tests, and non-jet mode for Burro.

TABLE 4-3
"SLAB"
Input Parameters

PARAMETER	DT1	DT2	DT4	G1	G2	G3	B3	B5	B8
Wind Velocity (m/s)	7.42	5.76	4.51	5.60	4.20	5.40	5.40	7.40	1.80
Ambient Temperature (K)	302.39	303.80	305.95	310.15	309.15	299.65	306.95	313.65	306.25
Relative Humidity (%)	13.2	17.5	21.3	4.9	10.5	27.6	5.2	5.8	4.6
Emission Rate (kg/s)	81	117.3	107.9	30.2	10.4	10.4	86.4	79.9	112.6
Surface Roughness (m)	0.003	0.003	0.003	0.003	0.003	0.003	0.0002	0.0002	0.0002
Pasquill Stability Class	4	4	4.5	4	4	4	2	3	5
Release Height (m)	0.79	0.79	0.79	1.00	1.00	1.00	0.00	0.00	0.00
Averaging Time (sec)	120	240	360	66.6	333/353.2	333/353.2	160	180	100
Release Duration (sec)	126.0	255.0	381.0	125.0	360.0	360.0	166.8	190.0	107.0
Reference Height (m)	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
Initial Chemical Temp (K)	294.65	293.25	297.25	313.15	311.15	312.15	111.70	111.70	111.70
Source Area (m**2)	0.0052	0.0070	0.0070	0.0081	0.0081	0.0081	485.1000	448.6000	632.2000
Init. Liquid Mass Frac.	0.83	0.83	0.83	0.8	0.8	0.8	0	0	0
Release Type	Horiz.jet	Horiz.jet	Horiz.jet	Horiz.jet	Horiz.jet	Horiz.jet	Evap.pool	Evap.pool	Evap.pool
Heat Of Vaporiz.(J/Kg)	1.3708E+06	1.3708E+06	1.3708E+06	3.7320E+05	3.7320E+05	3.7320E+05	5.0988E+05	5.0988E+05	5.0988E+05
Liquid Heat Cap. (J/kg-K)	4294	4294	4294	2528	2528	2528	3349	3349	3349
Vapor Heat Capacity (J/Kg-K)	2170	2170	2170	1450	1450	1450	2240	2240	2240
Liquid Density (kg/m**3)	682.8	682.8	682.8	957	957	957	424.1	424.1	424.1
Boiling Point (K)	239.7	239.7	239.7	292.7	292.7	292.7	111.7	111.7	111.7
Molecular Weight (kg)	0.01703	0.01703	0.01703	0.02001	0.02001	0.02001	0.01604	0.01604	0.01604
Sat. Press. Const.(spb)K	2132.52	2132.52	2132.52	3404.51	3404.51	3404.51	-1.00	-1.00	-1.00
Sat. Press. Const.(spc)K	-32.98	-32.98	-32.98	15.06	15.06	15.06	-1	-1	-1
N CALC	1	1	1	1	1	1	1	1	1

AIRTOX requires the user to specify an initial percent of liquid and percent aerosol. The non-jet mode also requires an air-to-gas ratio, in order to account for the amount of ambient air initially entrained into the release. For all experiments the portion of liquid chemical prior to the release was specified as 100 percent. In jet mode, AIRTOX does not allow pool formation and treats all of the released liquid as aerosol. For these cases, AIRTOX automatically sets the percentage aerosol to 100 percent. For Burro tests, since the releases are evaporating pool releases, the percent aerosol is set to zero. The air/gas ratio is specified as 10, which is the lowest suggested value to simulate a pure release. Minimum pool depth was set to a small value to simulate a film. Soil characteristics are taken directly from the AIRTOX on-line user's guide with soil type estimated from descriptions of the release site.

For the Desert Tortoise (AIRTOX jet mode) experiments, exit velocities and source diameters were determined using the suggestions of the model developer:

$$V_e = Q/\Gamma A_e \qquad (1)$$

where:

V_e = the exit velocity at the orifice (m/s)
Q = mass emission rate (kg/s)
Γ = liquid density (kg/m^3)
A_e = actual orifice area (m^2)

$$V'_e = V_e + (P_s - P_{atm})/(Q/A_e) \qquad (2)$$

where:

V'_e = the effective release velocity to be input to AIRTOX (m/s)
P_s = pressure immediately upstream of the orifice (newtons/m^2)
P_{atm} = atmospheric pressure (newtons/m^2) = 101325 newtons/m^2 at sea level

$$A'_e = A_e V_e/V'_e \qquad (3)$$

where:

A'_e = effective area of the release to be used as input to AIRTOX

For the Goldfish tests, equations (2) and (3) gave unrealistic values. Based on the advice of the model developer, exit velocities for Goldfish were computed using equation (1), and the actual orifice diameter was input. Model inputs for AIRTOX are summarized in Table 4-4.

After reviewing model performance results, the AIRTOX model developers made two further suggestions regarding implementation of the AIRTOX model. AIRTOX assumes an anemometer height of 10 m. Wind speed measurements during all of the release experiments were made at lower heights. The developers recommended using adjusted wind speeds based upon an assumed exponential profile. Secondly, to compared observed and predicted concentrations, the AIRTOX output in "kg/m^3" was converted to "ppm" at ambient temperature. In the case of the Burro tests, the developers recommended using cloud temperature rather than ambient temperature for this conversion. These "after the fact" suggestions were not implemented for this study.

4.2.4 CHARM

The CHARM model allows the user to select one of six different release scenarios: (1) quick loss of liquid, (2) quick loss of gas, (3) continuous liquid spill with pool contained by dikes or terrain, (4) continuous liquid spill with uncontained flow, (5) continuous gas release, and (6) user supplying complete description of puffs generated by release (liquid or gas, continuous or quick). CHARM considers all of the release characteristics described in Table 4.1A, except the partitioning of a jet into aerosol cloud and pool. CHARM contains an on-line chemical library, but it was not determined during this evaluation whether the chemical characteristics listed in Table 4.1B are considered. For these tests, option (6) was selected in order that the test release rates and durations could be specified by the user. Option (6) requires that the initial condition of the first puff be

TABLE 4-4
"AIRTOX"
Input Parameters

PARAMETER	DT1	DT2	DT4	G1	G2	G3	B3	B5	B8
Wind Velocity (m/s)	7.42	5.76	4.51	5.60	4.20	5.40	5.40	7.40	1.80
Ambient Temperature (K)	302.39	303.80	305.95	310.15	309.15	299.65	306.95	313.65	306.25
Relative Humidity (%)	13.2	17.5	21.3	4.9	10.5	27.6	5.2	5.8	4.6
Emission Rate (kg/s)	81	117.3	107.9	30.2	10.4	10.4	86.4	79.9	112.6
Surface Roughness (m)	0.003	0.003	0.003	0.003	0.003	0.003	0.0002	0.0002	0.0002
Pasquill Stability Class	D	D	E	D	D	D	B	C	E
Release Height (m)	0.79	0.79	0.79	1.00	1.00	1.00	-	-	-
Release Duration (sec)	126.0	255.0	381.0	125.0	360.0	360.0	166.8	190.0	107.0
Initial Liquid (%)	100	100	100	100	100	100	100	100	100
Initial Aerosol (%)	100	100	100	100	100	100	0	0	0
Release Angle (Degrees)	0	0	0	0	0	0	0	0	0
Orifice Area (m**2)	0.001470	0.002020	0.001710	0.000150	0.008100	0.008100	-	-	-
Exit Velocity (m/s)	81.2	85.2	92.5	3.89	1.34	1.34	-	-	-
Release Temp (K)	294.65	293.25	297.25	313.15	311.15	312.15	-	-	-
Dilution Factor	-	-	-	-	-	-	10.00	10.00	10.00
Dike Area (m**2)	0.00	0.00	0.00	0.00	0.00	0.00	2642.0	2642.0	2642.0
Min Pool Depth (m)	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Soil Conduct.(Kcal)(msK)-1	-	-	-	-	-	-	5.280E-03	5.280E-03	5.280E-03
Soil Thermal Diffusivity (m**	-	-	-	-	-	-	9.488E-07	9.488E-07	9.488E-07
Type of Release	1	1	1	1	1	1	0	0	0
Storage Temp	-	-	-	-	-	-	111.7	111.7	111.7
Jet Length	33	53	68	4	0	0	-	-	-

described in detail. This involves quantifying movement and composition for the initial puff. The releases were defined in CHARM as continuous with a constant emission rate. Source data were provided from the MESS Archive and experiment data reports, or estimated based on available data.

The evaluation tests for Desert Tortoise and Goldfish were treated as jet releases by specifying an initial puff movement calculated using formula 1 of AIRTOX. Initial puff depth was calculated internally by the model. Initial puff temperature was set at the boiling point of the released chemical. Release height was specified based on the experimental site description except in the case of the HF releases. When the actual experimental release height is used, the HF cloud lifts off the ground and simulated 1 m concentrations are zero. To prevent this, the release height is input as zero as recommended by the model developer. Internally, CHARM does not allow cloud liftoff if the release height is zero. This more realistically models HF releases since experimental evidence indicates such clouds do stay near the surface.

The jet releases for Desert Tortoise and Goldfish were modeled as continuous pure chemical spills with flash fractions specified as 0.17 and 0.20, respectively.

For the Burro tests, pool area was assumed to have the same proportionality to release volume as existed in test 9 (where pool size was measured). The diameter of the pool was calculated and used for the initial puff diameter. All puff movement components were set to zero. See Table 4-5 for a summary of all test inputs for CHARM.

4.2.5 FOCUS

The experimental data inputs for FOCUS were derived directly from the MESS archive and experimental data reports. Very few physical assumptions were required. The release characteristics listed in Table 4-1A were simulated in

TABLE 4-5
"CHARM"
Input Parameters

PARAMETER	DT1	DT2	DT4	G1	G2	G3	B3	B5	B8
Wind Velocity (m/s)	7.42	5.76	4.51	5.60	4.20	5.40	5.40	7.40	1.80
Ambient Temperature (C)	29.24	30.65	32.80	37.00	36.00	26.50	33.80	40.50	33.10
Relative Humidity (%)	13.2	17.5	21.3	4.9	10.5	27.6	5.2	5.8	4.6
Emission Rate (g/s)	81000	117300	107900	28125	10400	10400	86200	79800	113100
Pasquill Stability Class	D	D	E	D	D	D	B	C	E
Release Height (m)	0.79	0.79	0.79	0.00	0.00	0.00	0.00	0.00	0.00
Averaging Time (sec)	120	240	360	66.6	333/353.2	333/353.2	160	180	100
Release Duration (sec)	126.0	255.0	381.0	125.0	360.0	360.0	166.8	190.0	107.0
Pressure (atm)	0.897	0.898	0.891	0.891	0.900	0.900	0.936	0.929	0.929
Release Type (note)	6	6	6	6	6	6	6	6	6
Vertical Release Velocity	0	0	0	0	0	0	0	0	0
Horiz. Release Velocity (m/s)	22.90	24.51	22.60	3.66/3.89	1.34	1.34	0	0	0
Initial Puff Depth (m)	blank	blank	blank	blank	blank	blank	blank	blank	blank
Initial Puff Diameter (m)	0.0810	0.0945	0.0945	0.1016	0.1016	0.1016	36.00	34.60	40.00
Initial Puff Temperature (C)	-33.35	-33.35	-33.35	20.00	20.00	20.00	-161.45	-161.45	-161.45
Frac. Emitted as Droplets	0.83	0.83	0.83	0.8	0.8	0.8	0	0	0
Molar Air Fraction	0	0	0	0	0	0	0	0	0
Molar Water Vapor	0	0	0	0	0	0	0	0	0

FOCUS by selecting the options for regulated, continuous, vapor dispersion with no spill confinements specified. It is not apparent whether FOCUS simulates any chemical transformations. The required input variables were all readily available in the experimental data reports. However, the wind speed data required by FOCUS is at the 10 m level. Wind speed data were scaled to the 10 m level using the power-law wind equation (Panofsky and Dutton, 1984):

$$u = u_1 \left(\frac{z}{z_1} \right)^p \text{ and } p = \left(\ln \left(\frac{z}{z_0} \right) \right)^{-1}$$

where:

u = wind speed at 10 m
u_1 = wind speed at lower level (MESS Archive value)
z = 10 m
z_1 = height of wind measurement at lower level (MESS Archive value)
z_0 = roughness length

The Burro simulations were run with the spill surface roughness described as "calm open seas" to simulate an over water release. See Table 4-6 for a summary of all test inputs for FOCUS.

4.2.6 SAFEMODE

Release scenarios are specified in SAFEMODE as either continuous or instantaneous. SAFEMODE does not explicitly treat many of the characteristics listed in Table 4.1A. However, they may be addressed internally in the model. SAFEMODE was provided to TRC with a limited, on-line chemical library which contained only the chemical being evaluated in the test. It is likely that SAFEMODE accounts for reactions with ambient water vapor, but the documentation provided does not address this issue. For this evaluation, all releases are treated as continuous. SAFEMODE allows the user to select one of four different continuous release scenarios: (1) Hole in Tank, (2) Short Pipe from Tank, (3) Hole in Pipe, and (4) Severed Pipeline.

TABLE 4-6
"FOCUS"
Input Parameters

PARAMETER	DT1	DT2	DT4	G1	G2	G3	B3	B5	B8
Wind Velocity 10m (m/s)	9.05	7.02	5.50	6.80	5.10	6.60	6.30	8.60	2.10
Ambient Temperature (C)	29.24	30.65	32.80	37.00	36.00	26.50	33.80	40.50	33.10
Relative Humidity (%)	13.2	17.5	21.3	4.9	10.5	27.6	5.2	5.8	4.6
Emission Rate (kg/s)	81	117.3	107.9	30.2	10.4	10.4	86.2	79.8	113.1
Surface Roughness (m)	0.003	0.003	0.003	0.003	0.003	0.003	0.0002	0.0002	0.0002
Atmospheric Stability Class	D	D	E	D	D	D	B	C	E
Release Height (m)	0.79	0.79	0.79	1.00	1.00	1.00	0.00	0.00	0.00
Release Duration (sec)	126.0	255.0	381.0	125.0	360.0	360.0	166.8	190.0	107.0
Material Composition (%)	100 NH3	100 NH3	100 NH3	100 HF	100 HF	100 HF	92.5 Meth 6.2 Ethan 1.3 Propa	93.6 Meth 5.3 Ethan 1.1 Propa	87.4 Meth 10.3 Ethan 2.3 Propa
Release Temperature (C)	21.5	20.1	24.1	40.0	38.0	39.0	-163.0	-163.0	-163.0
Release Pressure (kPa)	1013.80	1117.20	1179.30	866.90	894.50	908.30	94.90	94.20	94.20
Source Diameter (m)	0.0810	0.0945	0.0945	0.1016	0.1016	0.1016	0.2500	0.2500	0.2500
Release Area (m**2)	0.0052	0.0070	0.0070	0.0081	0.0081	0.0081	0.0491	0.0491	0.0491
Substrate Temperature (C)	31.6	30.6	30.8	37.0	36.0	26.5	30.0	30.0	30.0
Release Angle (Degrees)	0	0	0	0	0	0	270	270	270
Spill Surface Roughness		Mud Flats			Mud Flats		Calm Open Seas		
Surrounding Area Roughness		Mud Flats			Mud Flats			Mud Flats	

The evaluation tests were simulated by SAFEMODE by allowing the model to internally generate source/release information based on user input tank pressure, storage temperature, and orifice diameter. The input orifice diameter was adjusted until release rates compared with the actual test measurements. SAFEMODE also calculates a flash fraction for the released liquid. For these tests, if the flash fraction was available in the experimental literature, the SAFEMODE computed value was changed to the documented value; otherwise the value computed by the model was used. SAFEMODE calculates a continuous stability class (4.5 etc.) using sigma theta. Sigma theta stabilities were used when available (DT and Burro), but these values were overridden if they disagreed with the stability information in the experimental literature.

SAFEMODE inputs and options were similar for Desert Tortoise and Goldfish tests. Each test was simulated as a continuous jet release from a short pipe on a tank, with no spill confinements. The Burro tests were also run by specifying a release of LNG from a short pipe on a tank, but spill confinement was specified for these tests. The diameter of the water basin was used for the dike diameter. The SAFEMODE model as provided to TRC does not perform the calculation for near field receptors in certain cases.

SAFEMODE inputs are summarized in Table 4-7.

4.2.7 TRACE

The TRACE inputs were prepared from the MESS Archive and experimental data reports, and from sample test cases provided by the model developer. TRACE considers all of the release characteristics listed in Table 4.1A. However, both the horizontal jetting and the jet partitioning into pool and aerosol were not used at the suggestion of the model developer. TRACE contains an extensive on-line chemical library, and it is likely that reactions of NH_3 and HF with ambient water are considered. TRACE requires a number of parameters

TABLE 4-7
"SAFEMODE"
Input Parameters

PARAMETER	DT1	DT2	DT4	G1	G2	G3	B3	B5	B8
Wind Velocity (m/s)	7.42	5.76	4.51	5.60	4.20	5.40	5.40	7.40	1.80
Ambient Temperature (K)	302.39	303.80	305.95	310.15	309.15	299.65	306.95	313.65	306.25
Relative Humidity (%)	13.2	17.5	21.3	4.9	10.5	27.6	5.2	5.8	4.6
Emission Rate (kg/s)	81	117.3	107.9	30.2	10.4	10.4	86.4	79.9	112.6
Surface Roughness (m)	0.003	0.003	0.003	0.003	0.003	0.003	0.0002	0.0002	0.0002
Pasquill Stability Class	D	D	E	D	D	D	B	C	E
Release Height (m)	0.79	0.79	0.79	1.00	1.00	1.00	0.00	0.00	0.00
Averaging Time (sec)	120	240	360	66.6	333/353.2	333/353.2	160	180	100
Substrate Temperature (K)	304.75	303.75	303.95	310.15	309.15	299.65	308.60	314.90	305.80
Pressure (Pa)	90866	90967	90258	90258	91170	91170	94817	94817	94108
Sigma Theta (degrees)	5.73	7.54	5.02	-	-	-	13.3	11.1	5.57
Initial Chemical Temp (K)	294.65	293.25	297.25	313.15	311.15	312.15	111.70	111.70	111.70
Storage Pressure (atm)	13.40	13.90	13.80	8.55	8.82	8.96	2.40	2.40	2.40
Storage Vessel Vol.(m**3)	14.9	43.8	60.3	4.0	3.9	3.9	34.0	35.8	28.4
Tank Diameter (m)	10	10	10	10	10	10	10	10	10
Orifice by iteration (m)	0.09	0.1075	0.1024	0.041	0.0238	0.0137	0.181	0.174	0.2065
Velocity Measurement Ht. (m)	2	2	2	2	2	2	2	2	2
Dike diameter (m)							57	57	57

which characterize the physical attributes of the source cloud, and others which control the modeling calculations. Through conversations with the model developer it was decided that TRACE default values would be used in order to compute cloud characteristics. These values include aerosol/flash and air entrainment parameters.

Desert Tortoise and Goldfish tests were run without specifying any initial momentum. The model developer, from experience with the data bases, determined that jetting effects had negligible influence on concentration levels at the experimental receptor arcs. The model was allowed to perform the source term simulations. See Table 4-8 for a summary of all test inputs for TRACE.

4.3 Model Application (Output Assumptions)

The maximum centerline concentrations and half width values for the evaluation were obtained either directly from model outputs or required some interpolative procedure. Described below are the procedures used to determine the variables for the comparison.

4.3.1 DEGADIS

The test-averaged concentrations required for the evaluation are provided in the DEGADIS model output. DEGADIS assumes that the central portion of the cloud has a uniform cross-wind concentration distribution. Outside of this region, the plume is assumed to have a Gaussian concentration distribution. For monitoring locations within the uniform region, the predicted maximum (centerline) concentration at selected downwind distances is output by the model. At intermediate distances, concentration values were obtained by interpolation. For monitors outside this distance, steady-state maximum concentrations can be calculated using:

TABLE 4-8
"TRACE"
Input Parameters

PARAMETER	DT1	DT2	DT4	G1	G2	G3	B3	B5	B8
Wind Velocity (m/s)	7.42	5.76	4.51	5.60	4.20	5.40	5.40	7.40	1.80
Ambient Temperature (K)	302.39	303.80	305.95	310.15	309.15	299.65	306.95	313.65	306.25
Relative Humidity (%)	13.2	17.5	21.3	4.9	10.5	27.6	5.2	5.8	4.6
Emission Rate (kg/s)	81.00	117.25	107.90	30.20	10.40	10.40	86.40	79.90	112.60
Surface Roughness (m)	0.003	0.003	0.003	0.003	0.003	0.003	0.0002	0.0002	0.0002
Pasquill Stability Class	D	D	E	D	D	D	B	C	E
Release Height (m)	0.79	0.79	0.79	1.00	1.00	1.00	0.00	0.00	0.00
Averaging Time (sec)	120	240	360	66.6	333/353.2	333/353.2	160	180	100
Release Duration (sec)	126.0	255.0	381.0	125.0	360.0	360.0	166.8	190.0	107.0
Monin-Obukhov Length(m)	92.7	94.7	45.2	10000.0	10000.0	10000.0	-9.1	-25.5	16.5
Initial Cloud Radius	0	0	0	0	0	0	0	0	0
Max Pool Area (m2)	10000	10000	10000	10000	10000	10000	2642	2642	2642
Min Pool Depth (m)	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Substrate (sandy soil)	wet	wet	dry	dry	dry	dry	dry	dry	dry
Aerosol/Flash Mass Ratio	5.09/TRACE	5.24/TRACE	4.77	16.76	19.6/TRACE	16.3/TRACE	0	0	0
Air Entrainment	default	default	default	default	default	default	default	default	default
Solar Radiation (W/m**2)	1000	1000	1000	1000	1000	1000	1000	1000	400
Substrate Temperature (K)	304.75	303.75	303.95	310.15	309.15	299.65	308.60	314.90	305.80
Initial Chemical Temp (K)	294.50	293.24	297.24	313.15	311.15	312.15	111.70	111.70	111.70
Initail Cloud Velocity(X)	0	0	0	0	0	0	0	0	0
Initail Cloud Velocity(Y)	0	0	0	0	0	0	0	0	0
Release Type	1	1	1	1	1	1	1	1	1
Phase of Chemical	1	1	1	1	1	1	1	1	1
Albedo	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
Initial Dilution	0	0	0	0	0	0	0	0	0
Reference Height (m)	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0

$$C(x,y,z) = C_c(x) \exp\left(- \left(\frac{|y| - b(x)}{S_y(x)} \right)^2 - \left(\frac{z}{S_z(x)} \right)^{1+\alpha} \right) \quad (1)$$

(from page 38, DEGADIS Version 2.1 User's Manual) where:

$C(x,y,z)$	=	concentration at any point (x,y,z) (kg m^{-3})
$C_c(x)$	=	centerline concentration (kg m^{-3})
y	=	lateral distance from plume centerline (m)
$b(x)$	=	half-width of the horizontally homogeneous central section of gas plume (m)
z	=	height (m)
α	=	constant in power law wind profile
$S_y(x)$	=	horizontal concentration scaling parameter (m)
$S_z(x)$	=	vertical concentration scaling parameter (m)

All of the necessary variables are provided in the DEGADIS model output. For monitors within b(x) of the centerline:

$$C(x,y,z) = C_c(x) \exp\left(- \left(\frac{z}{S_z(x)} \right)^{1+\alpha} \right) \quad (2)$$

can be used to calculate off-centerline concentration.

To calculate the characteristic plume width for the evaluation, Equation 1 (assuming z=0) is reduced to:

$$y = S_y(\ln(2))^{1/2} + b(x) \quad (3)$$

4.3.2 SLAB

The test averaged concentrations required for the evaluation are provided in the SLAB model output as a function of effective half-width distance. SLAB outputs averaged volume concentrations downwind at points where the ratio of the lateral distance (y) to effective half-width (bbc) is 0.0, 0.5, 1.0, 1.5, 2.0, and 2.5. SLAB will provide predicted concentrations at user selected heights.

The maximum centerline concentration for each arc distance is determined by linearly interpolating to the given arc distance, the concentration at the

point where the ratio of the lateral distance (y) to the effective half-width (bbc) equals zero. Similarly, effective half-width (bbc) is interpolated to the given arc distance. The half-width value for the evaluation will then be determined by linearly interpolating the ratio of y/bbc to 50 percent of maximum concentration and solving for y.

4.3.3 AIRTOX

The maximum averaged concentrations required for the evaluation are provided in the output summary of AIRTOX. AIRTOX predictions represent ground level concentrations. The centerline receptor is used to obtain the maximum concentration at a given arc distance. An average plume half-width is obtained from AIRTOX by making additional model runs, specifying receptors crosswind to the centerline at a given arc distance. The half-width distance is then determined by linearly interpolating to 50 percent of the maximum concentration value.

4.3.4 CHARM

The test-averaged concentrations required for the evaluation are provided by the CHARM2 section of the CHARM model system. Model results of interest for the evaluation are provided by CHARM2 in a graphical display, produced in an interactive mode by the user. By selecting the concentration option, a cross-hair (cursor) is provided on the graphical display to be positioned to the arc distance of interest. Through the options available in the concentration mode, graph scale, time since release, and receptor height are set to represent the test event. The scale is selected to provide detailed resolution for the given arc distance. Time since release is set at a value to allow complete cloud passage for the event. Once the cross-hair is positioned at the given arc distance, the dosage option is selected. In the

dosage mode, averaging time is set to the test-averaged value. The highest concentration at the given arc is then produced. Linear interpolation is required in some instances to obtain a concentration at the exact arc location.

To obtain the half-width distance, the concentration/dosage options are again utilized. The cross-hair is positioned along the arc in the concentration mode, and the concentration value is determined from the dosage screen. The iterative manipulation between concentration/dosage modes continues until the point along the arc with 50 percent of the maximum concentration is found. CHARM2 provides lateral distance in the form of angular degrees and radial distance. The half-width is then determined by converting the angular position to a lateral distance in meters. Linear interpolation is required to provide the exact distance.

4.3.5 FOCUS

The FOCUS output summary provides the maximum averaged concentrations required for the evaluation. FOCUS predictions are for a model default, 1 m receptor height. Linear interpolation is required in many cases to obtain the maximum concentration at a given arc distance. An average plume half-width can be obtained from FOCUS by making an additional model run, specifying the desired concentration value as one of the input concentration limits. The half-width value is then obtained by linearly interpolating to the given arc distance.

4.3.6 SAFEMODE

Maximum averaged concentrations are provided as output from SAFEMODE. The values represent ground level concentrations. Linear interpolation is required in many cases to obtain the maximum concentration at a given arc distance. An average plume half-width is obtained from SAFEMODE by performing

an additional model run, specifying the desired concentration value as the user input concentration contour.

4.3.7 TRACE

The TRACE receptor impact summary provided the maximum concentration values required for the evaluation. Maximum concentrations are provided for user specified receptor heights. A test-average plume half-width can be obtained from TRACE by making additional model runs, specifying four receptors, positioned crosswind to the centerline at a given arc distance in the vicinity of estimated half-width. The half-width value is then determined by interpolating to 50 percent of the maximum concentration value.

4.4 Experiment Data Analysis

The measured concentration data for the Goldfish, Burro, and Desert Tortoise experiments were obtained directly from the MESS Archive. For this evaluation study, averaged concentrations were calculated from the archived values in order to create averages from the measured data that best represent the release period for each test. For example: the release duration for Desert Tortoise 4 was 381 seconds. The raw data for this test represents concentration data averaged over 30 seconds. For this case a 360 second averaged concentration was created (12-30 second values).

Data for each test were averaged for a time period that approximated the release duration for the test. The maximum averaged concentrations were computed for each receptor. The concentration measurements on each receptor arc were then examined to determine whether the maximum concentration value was contained within the sampling arc. Cases where the measured maximum occurred at the end of an arc were excluded from the study because the plume

centerline may not have intersected the receptor arc making exact determination of a maximum concentration impossible.

For those receptor arcs selected, observed half-widths were determined. Half-width for this evaluation is defined as the cross-wind distance from the maximum to the point at which concentration level is 50 percent of the maximum value. This distance was computed using linear interpolation between the receptors along the arc.

4.5 Model Limitations

The models used for the evaluation study are not all designed to produce predictions which correspond exactly to measured concentrations. Many of the models provide predictions only at locations determined internally by the model. Several provide only ground-level concentration estimates. 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 varying receptor heights.

The DEGADIS, AIRTOX and SAFEMODE models only produce predictions at ground level. For these models the ground level predictions were compared to the 1 m measurements. The DEGADIS model user's guide provides a method for calculating the predicted concentrations for elevated receptors, using model output information, but the model developer recommended against this method for estimating concentrations at 1 m. Thus, ground level concentrations were used.

For CHARM, FOCUS, and SAFEMODE, difficulties were encountered in obtaining meaningful concentration predictions for the near-field arcs in certain cases. CHARM and SAFEMODE would not produce predictions for the Desert Tortoise 100 m arcs, since the 100 m arc falls within a "jetting region" predicted by the models. The FOCUS model produced erratic concentration

predictions for the Burro 57 m arc; no FOCUS predictions at this distance were included in comparisons with observed values. FOCUS treats the Burro tests as transient releases of LNG, and when simulating a transient release FOCUS produces concentration "snapshots" at points in time. These snapshot times can be varied by changing the lowest concentration of interest value in the model. Several different values were selected but no reasonable concentrations could be deduced from the output.

4.6 Model Averaging Time

Most of the models accept a user-specified averaging time for concentration prediction. For this study, all of these models were run with the averaging time specified as the release duration. This choice was made to provide a reasonable degree of consistency among the models. Measured concentrations were also averaged over the release duration, as discussed in Section 4.4.

This averaging time treatment represents a technical compromise which sacrifices potentially useful information concerning model performance for estimating quasi-instantaneous concentrations, but provides a more convenient basis for testing a large number of models in a consistent manner. (At least one of the models does not provide quasi-instantaneous predictions.) The choice of averaging time for concentration predictions should not (in principal) influence a model's simulation of the release scenarios. The temporal and spatial resolution used by each model to simulate the evolving cloud/plume are chosen internally based upon physical considerations, independent of averaging time.

4.7 Statistical Methods

The statistical methodology for evaluating dense gas models is designed to provide a straightforward assessment of model performance, using simple

measures appropriate to small data sets. For each experimental program, maximum (centerline) concentrations predicted by each model will be compared to the observed maximum value at distances where measurements were taken. Observed and predicted cloud half-width values are compared at the same distances. The number of data points provided in each data set ranges from six for Desert Tortoise (three tests, two distances) to ten for Burro (three tests, four distances, minus two arcs with inadequate data).

The observed maximum concentration at each distance will underestimate the "absolute" maximum value, since measurements are available only for a finite number of sampler locations. Comparison of predicted centerline versus observed maximum concentrations may therefore introduce an unintentional bias towards over prediction. While the extent of this bias cannot be quantified readily for all of the models, inspection of point concentration predictions for selected models and experiments suggests that this effect is generally small. The highest predicted point concentration was generally within 5 percent of the centerline prediction.

Statistical measures proposed for the dense gas evaluation were defined in advance in a statistical protocol, which is provided in Appendix A. The measures which have been used in this evaluation are summarized in Table 4-9.

For bias, measures include the average difference between observed and predicted maximum concentration values, fractional bias for maximum values, and average difference of half-width values. Each measure is computed for each distance, and for all distances combined. Confidence intervals for average differences are based on the Student's t-test.

For scatter, the root-mean-square (RMS) error is calculated for maximum concentrations, and the number of data points for which observed and predicted maximum values agree within a factor of 2 is tabulated. Correlation coefficients are calculated only for all distances combined.

TABLE 4-9

PERFORMANCE MEASURES FOR THE DENSE GAS
MODEL EVALUATION STUDY

Measure (for each experimental program - all tests combined)

1. Bias

   a. Average Difference of Maximum Values (obs-pred)

      > Each distance (with confidence interval)
      > All distances combined (with confidence interval)

   b. Fractional Bias - Average of Maximum Values

      > Each distance
      > All distances combined

   c. Average Difference of Half - Width Values (obs-pred)

      > Each distance
      > All distances combined

2. Scatter

   a. RMS Error - Maximum values

      > Each distance
      > All distances combined

   b. Factor of 2 Agreement - Maximum Values

      > Each distance
      > All distances combined

3. Correlation

   a. Pearson Correlation Coefficient

      > All distances combined

Several measures which were proposed in the protocol have been deleted, and the "factor of 2 agreement" tabulation has been added. Numerical comparisons of maximum concentrations for individual data points were replaced by graphs and tables displaying the same information. Statistical measures calculated for individual experiments (all distances combined) were deleted, because results for maximum concentrations were consistently dominated by a single data point at the closest measured distance. Correlation coefficients for each distance were dropped, because correlation statistics for data sets with at most three points are highly unreliable.

The statistical measures which have been used are also subject to serious limitations. Confidence intervals for small data sets depend heavily on the calculated standard deviation, which is itself subject to large uncertainty. Calculation of "average differences" and confidence intervals using data from different distances is also a dubious undertaking, since observed and predicted concentrations often decrease by one or more orders of magnitude between the closest and furthest measurement distance. To the greatest practical extent, model performance has been examined test-by-test and distance-by-distance, to look for consistent patterns, as an independent check of the calculated statistical measures.

5.0 RESULTS OF THE MODEL EVALUATION

Seven dense gas air toxics models have been evaluated using three experimental data sets involving heavier-than-air releases. Model performance has been analyzed separately for each experimental program, followed by a summary discussion comparing the results for each model from all three data sets.

5.1 Desert Tortoise Ammonia Releases

For Desert Tortoise, predicted and observed concentrations are compared at two distances, 100 m and 800 m downwind of the release point, for three experiments. The maximum concentrations observed and predicted at each distance are illustrated for each experiment in Figures 5-1, 5-2 and 5-3. For two models, CHARM and SAFEMODE, predictions were not obtained for the 100 m distance for reasons discussed in Section 4.5.

5.1.1 Maximum Concentrations

Maximum observed and predicted concentrations were compared at each downwind distance. For Desert Tortoise Test 1 (DT-1), the results in Figure 5-1 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. At 800 m, four of the models (TRACE, CHARM, DEGADIS and FOCUS) predicted maximum concentrations close to the observed value, but at 100 m only TRACE predicted within a factor of 2. Three out of five models produced large overpredictions at 100 m, but none of the models overpredicted significantly at 800 m.

For DT-2 and DT-4, the results in Figures 5-2 and 5-3 show a pattern similar to DT-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,

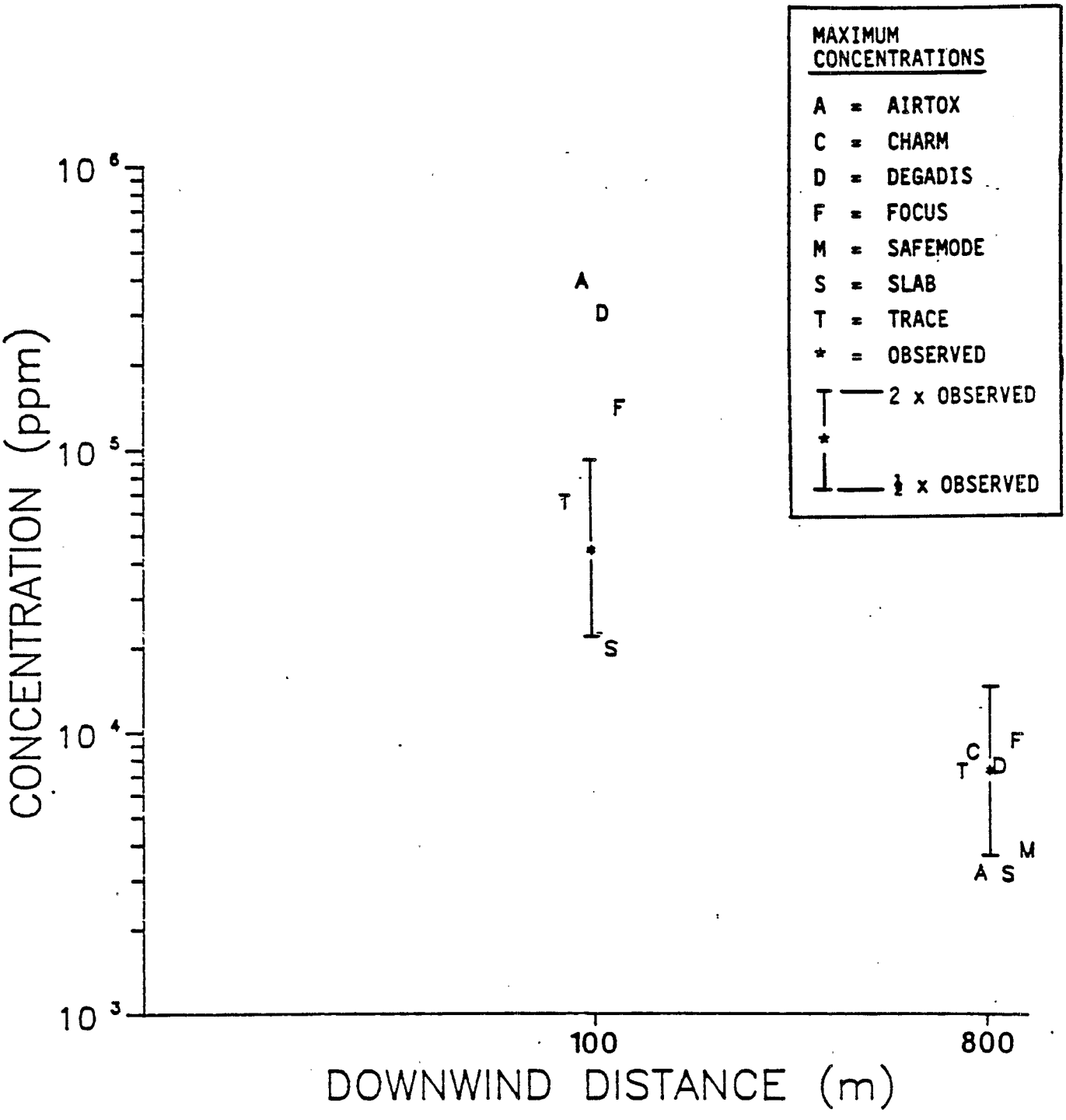


Figure 5-1. Observed and Predicted Maximum Concentrations versus Downwind Distance for Desert Tortoise NH_3 Test 1.

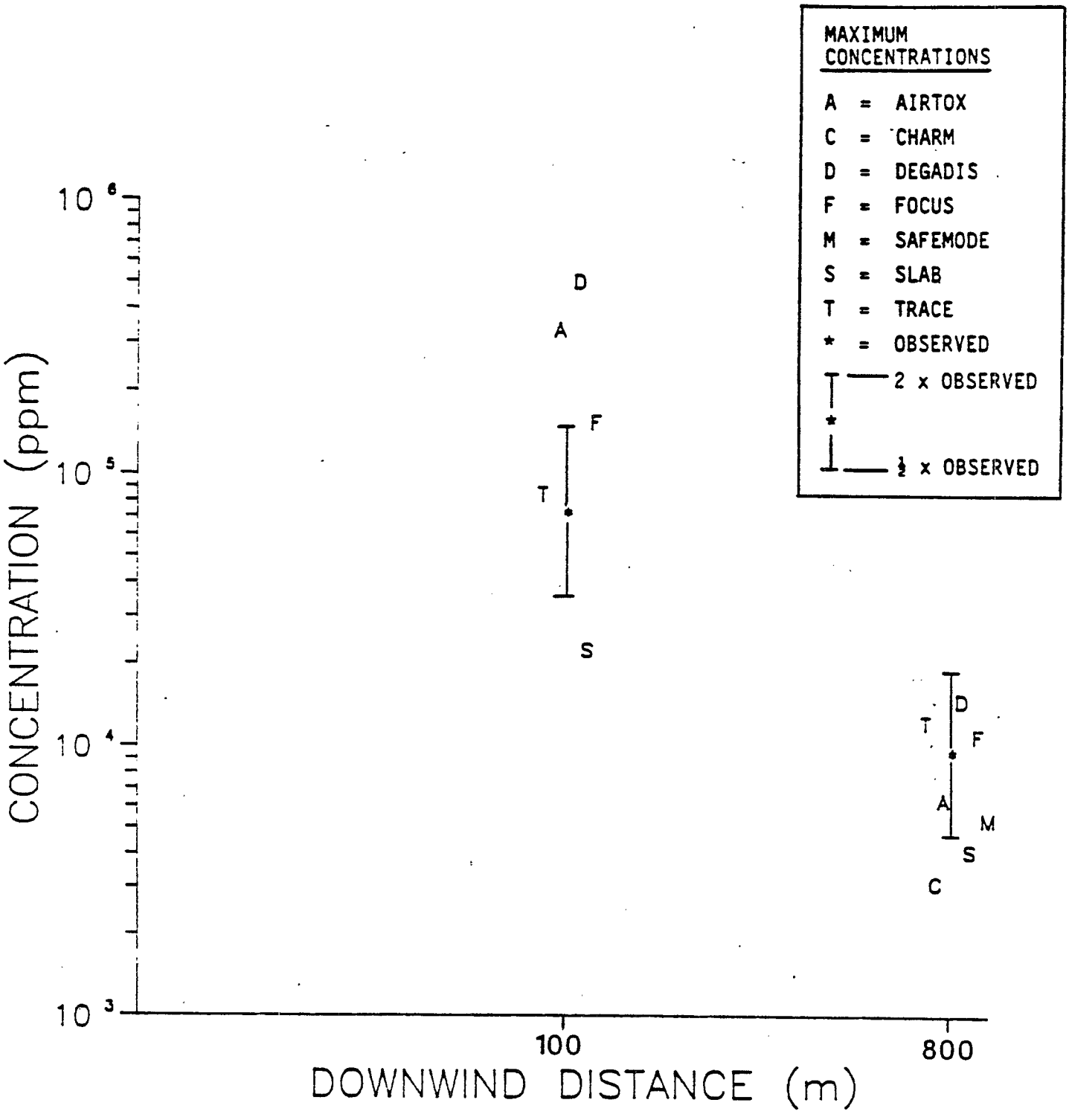


Figure 5-2. Observed and Predicted Maximum Concentrations versus Downwind Distance for Desert Tortoise NH_3 Test 2.

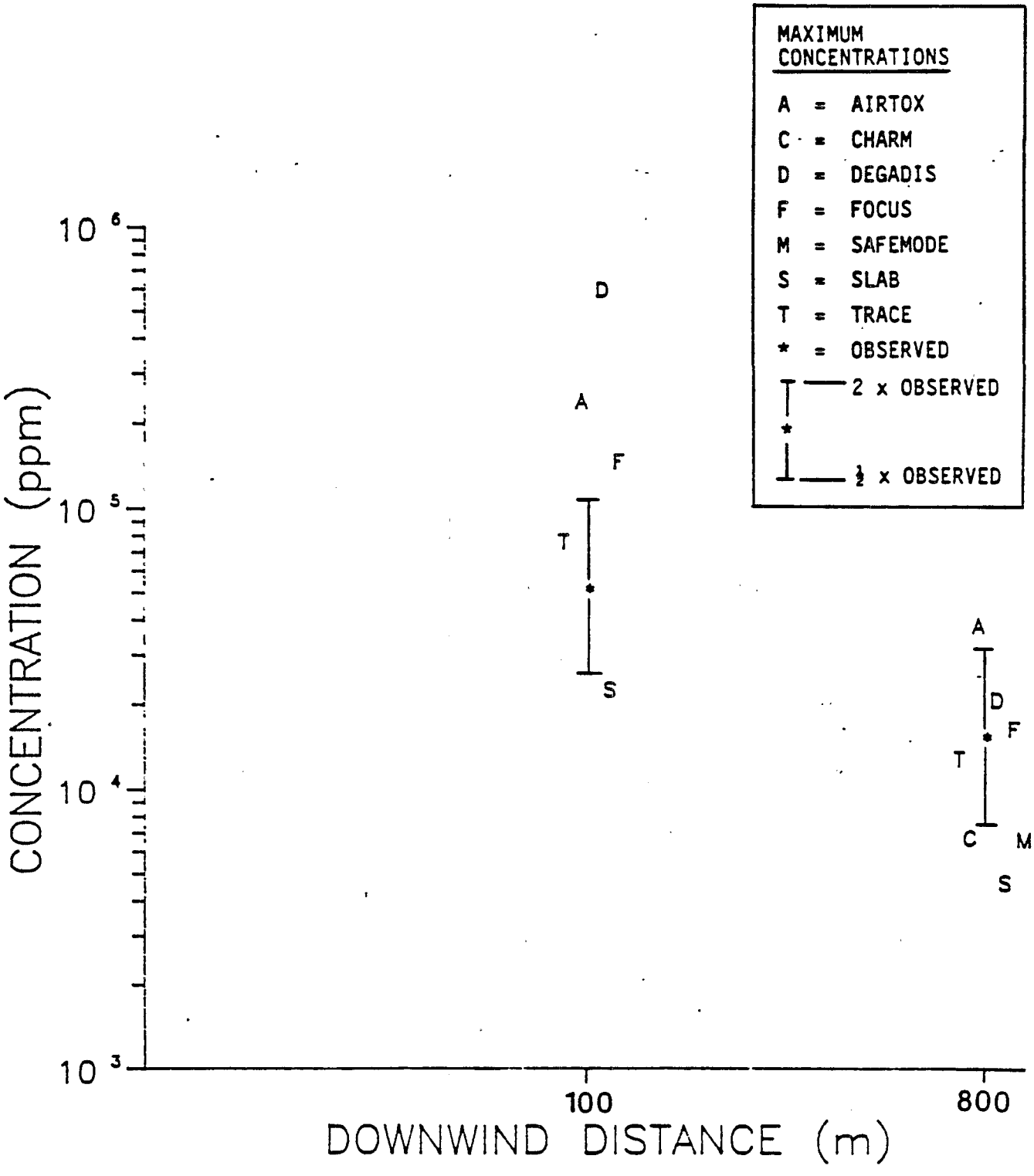


Figure 5-3. Observed and Predicted Maximum Concentrations versus Downwind Distance for Desert Tortoise NH_3 Test 4.

TRACE, FOCUS and DEGADIS consistently predicted within a factor of 2, while SLAB and SAFEMODE consistently underpredicted. AIRTOX and CHARM showed no consistent bias at 800 m.

The performance statistics for maximum concentrations for the Desert Tortoise experiments are summarized in Table 5-1. The results for the 100 m distance confirm that TRACE provided the best agreement with observed maximum values. The fractional bias of -0.29 for TRACE indicates overprediction by a factor of 1.34; the difference between observed and predicted values is significant at a 95 percent confidence level. TRACE produced the smallest RMS error, and its predicted maximum values were within a factor of 2 of observed for all three experiments. Fractional bias results indicate that SLAB underpredicted the average maximum value at 100 m by a factor of 2.6, while FOCUS overpredicted by the same factor. The RMS error is larger for FOCUS than for SLAB, since overprediction by a given factor produces errors of larger magnitude than underprediction by the same factor.

At 800 m, three of the models (TRACE, DEGADIS and FOCUS) predicted maximum values within a factor of 2 of observed for all three experiments. TRACE showed no prediction bias, while FOCUS produced the smallest RMS error and overpredicted by a factor of only 1.13. DEGADIS overpredicted by a factor of 1.33.

CHARM, SLAB and SAFEMODE all showed underprediction bias at 800 m, although differences were generally not significant at a 95 percent confidence level. SAFEMODE predictions agreed with observed maximum values within a factor of 2 for two of the three tests, and SAFEMODE produced the lowest RMS error of these three models. The fractional bias was lower for CHARM, but RMS error was higher. SLAB underpredicted by more than a factor of 2 for all three tests; the fractional bias indicates underprediction by a factor of 2.74.

Table 5-1. Performance Statistics for Maximum Concentration Values for Desert Tortoise NH3 Experiments

	TRACE	CHARM	AIRTOX	DEGADIS	SLAB	FOCUS	SAFEMODE
100m (N=3)							
average observed max value (ppm)	53,485		53,485	53,485	53,485	53,485	
fractional bias	-0.29		-1.40	-1.57	0.89	-0.89	
average diff. (obs-pred)	-17,822		-251,555	-387,762	33,051	-85,192	
95 percent confidence int.	(12,688)		(158,738)	(275,990)	(25,849)	(17,222)	
RMSE	18,539		259,545	403,366	34,651	85,473	
number of max values within 2x	3/3		0/3	0/3	0/3	0/3	
800m (N=3)							
average observed max value (ppm)	10,176	10,176	10,176	10,176	10,176	10,176	10,176
fractional bias	0.00	0.55	-0.38	-0.28	0.93	-0.12	0.69
average diff. (obs-pred)	41	4,405	-4,782	-3,310	6,445	-1,282	5,201
95 percent confidence int.	(5,000)	(4,190)	(29,170)	(5,425)	(6,679)	(1,076)	(5,202)
RMSE	2,014	5,944	12,680	3,966	6,983	1,353	5,653
number of max values within 2x	3/3	1/3	1/3	3/3	0/3	3/3	2/3
Combined(N=6)							
average observed max value (ppm)	31,830		31,830	31,830	31,830	31,830	
fractional bias	-0.25		-1.34	-1.51	0.90	-0.81	
average diff. (obs-pred)	-8,890		-128,168	-195,536	19,748	-43,237	
95 percent confidence int.	(10,226)		(138,245)	(218,050)	(16,087)	(44,353)	
RMSE	13,186		183,745	285,236	25,408	60,446	
number of max values within 2x	6/6		1/6	3/6	0/6	3/6	
Correlation coefficient	0.977		0.886	0.926	0.955	0.953	

At 800 m, AIRTOX produced the largest RMS error. The fractional bias for AIRTOX indicates overprediction by a factor of 1.47. For the three experiments, however, AIRTOX at 800 m produced underprediction by a factor of 2.3 for DT-1, agreement within a factor of 2 for DT-2, and overprediction by a factor of 3 for DT-3. The combined statistical results for both distances indicate clearly that TRACE performed best for predicting maximum concentrations for Desert Tortoise. TRACE produced the least bias, the smallest RMS error, and the highest correlation coefficient, and predicted 6 of 6 maximum values within a factor of 2. The other four models which provided predictions at both 100 m and 800 m all produced fractional bias values greater than 0.67 (exceeding a factor of 2 difference between observed and predicted maximum values) and also produced larger RMS error. The "combined" statistics tend to be dominated by the results at 100 m, because concentration values at this distance are larger by roughly a factor of 5. The correlation coefficients are relatively high for all five models, since the dominant feature in both the observed and predicted values is the large decrease in maximum values between 100 m and 800 m.

5.1.2 Cloud Half-Width

The results for cloud half-width, summarized in Table 5-2, are generally consistent among the three Desert Tortoise experiments. AIRTOX predicted half-width values closest to observed. SLAB and CHARM underpredicted cloud half-width by a moderate degree, while FOCUS underpredicted by more than a factor of 2 at both 100 m and 800 m. TRACE and SAFEMODE overpredicted cloud widths at 800 m by about 50 percent, while DEGADIS overpredicted at both 100 m and 800 m by more than a factor of 2. No direct relationship between prediction biases for half-width and maximum concentration values is evident when results in Tables 5-1 and 5-2 are compared.

Table 5-2. Performance Statistics for Cloud Half-Width for Desert Tortoise NH3 Experiments

	TRACE	CHARM	AIRTOX	DEGADIS	SLAB	FOCUS	SAFEMODE
100m (N=3)							
average observed value	22		22	22	22	22	
average diff. (obs-pred)	-34		1	-37	9	14	
800m (N=3)							
average observed value (m)	120	120	120	120	120	120	120
average diff. (obs-pred)	-65	43	-13	-134	26	74	-54
Combined(N=6)							
average observed value (m)	71		71	71	71	71	
average diff. (obs-pred)	-49		-6	-86	17	44	

5.2 Goldfish HF Releases

For the Goldfish HF experiments, concentration measurements were made at three downwind distances: 300 m, 1,000 m and 3,000 m. For two of the three tests selected for this evaluation, measured concentrations at 3,000 m did not provide a reliable basis for estimating the observed maximum concentration or cloud half-width.

5.2.1 Maximum Concentrations

Figure 5-4 illustrates the observed and predicted maximum concentrations at each distance for Goldfish Test 1 (G-1). At 300 m, the predicted maximum values for G-1 span a factor of 10. At 1,000 m and 3,000 m, model predictions are clustered within a factor of 4. At all three distances, the majority of predicted values are lower than observed; at 1,000 m, all seven models underpredict for G-1. CHARM and DEGADIS achieved relatively good agreement with observed maximum values at all three distances. TRACE also achieved relatively good agreement at 1,000 m and 3,000 m, but underpredicted by a factor of 2 at 300 m. FOCUS predicted very close to the observed maximum at 300 m, but underpredicted by a factor of 2 at 1,000 m and by a factor of 3 at 3,000 m. SAFEMODE results for G-1 improved with distance, starting with underprediction by a factor of 10 at 300 m. SLAB and AIRTOX underpredicted by more than a factor of 2 at all three distances.

Results for G-2 are illustrated in Figure 5-5. Many similarities to G-1 are evident; all of the predicted maximum values are less than or equal to the observed maximum. AIRTOX and SLAB again show large underprediction at both distances. CHARM and DEGADIS give the best agreement at 300 m and 1,000 m, respectively. TRACE results again improved with distance, while FOCUS results worsened. SAFEMODE again gave large underprediction at 300 m.

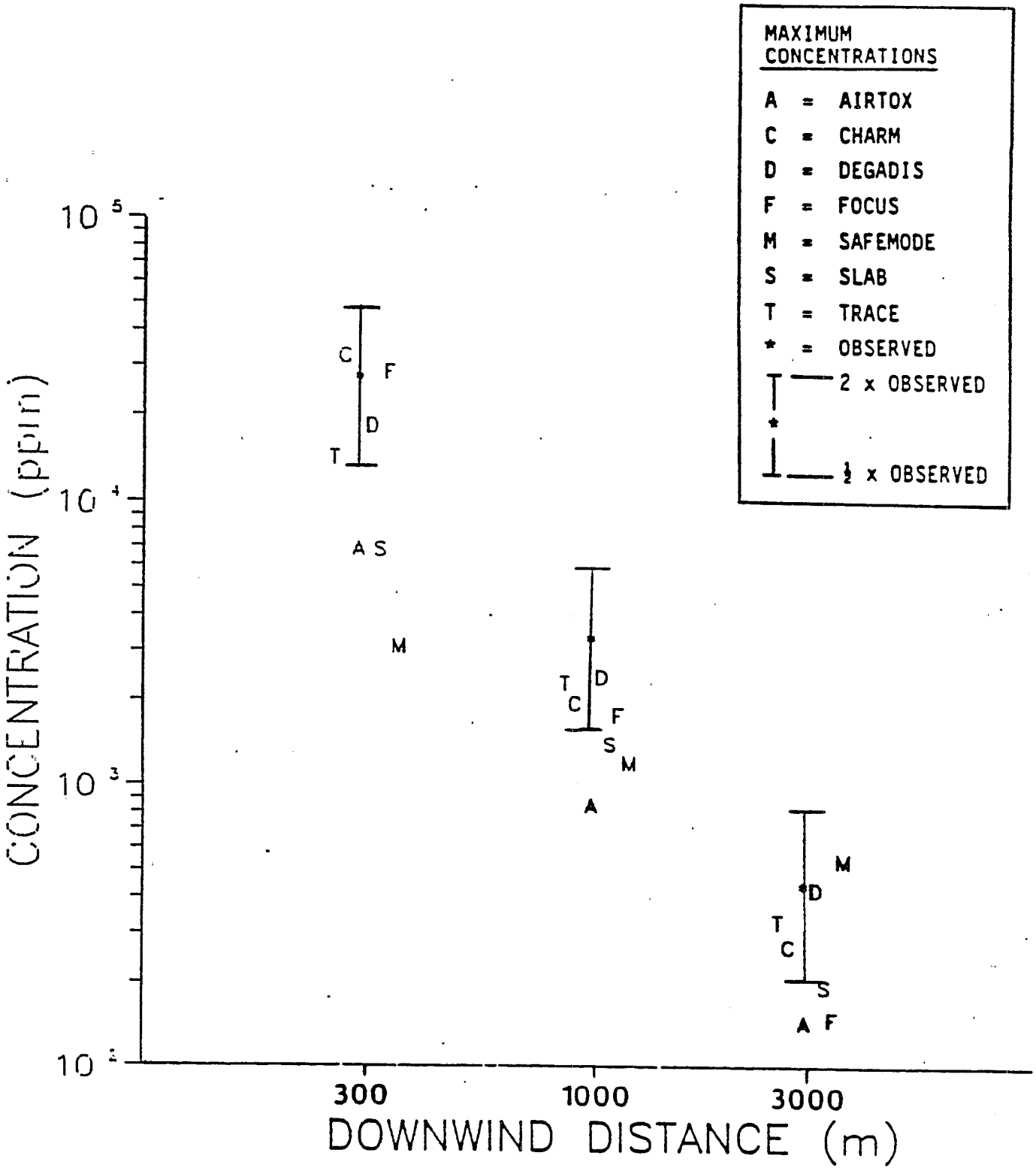


Figure 5-4. Observed and Predicted Maximum Concentrations versus Downwind Distance for Goldfish HF Test 1.

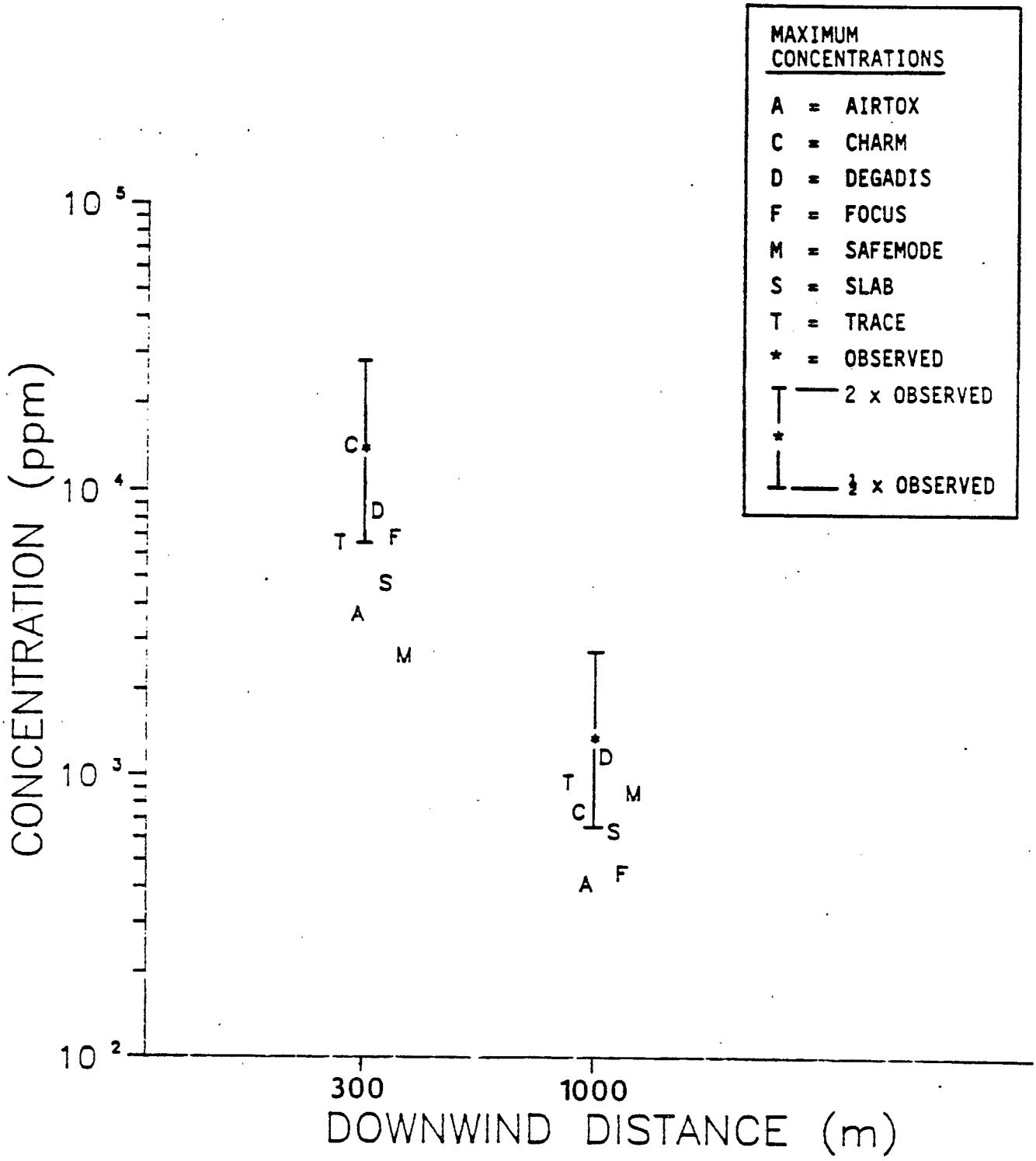


Figure 5-5. Observed and Predicted Maximum Concentrations versus Downwind Distance for Goldfish HF Test 2.

For G-3 (Figure 5-6), the model predictions at both distances and the observed maximum value at 300 m are very similar to G-2, but the observed maximum value at 1,000 m for G-3 is higher than G-2 by a factor of 2. All seven models underpredicted at 1,000 m for G-3 by more than a factor of 2.

Performance statistics for Goldfish for maximum concentration values are summarized in Table 5-3. At 300 m, CHARM and FOCUS achieved the least prediction bias (smallest fractional bias) and smallest RMS error. DEGADIS underpredicted (on average) at 300 m by a factor of 1.57, and gave agreement within a factor of 2 for all three maximum values. TRACE underpredicted (on average) by a factor of 2; SLAB and AIRTOX underpredicted by factors of 3.5 and 3.8, respectively; while SAFEMODE underpredicted by more than a factor of 5.

At 1,000 m, DEGADIS and TRACE underpredicted by factors of 1.50 and 1.67, respectively, while CHARM underpredicted by a factor of 2. Three models (SLAB, FOCUS, SAFEMODE) underpredicted by factors between 2 and 3. AIRTOX underpredicted by a factor of 4 and gave the largest RMS error.

At 3,000 m, comparisons are based only on results from one test. Biases for this case are similar to those found at 1,000 m, except for a moderate overprediction by SAFEMODE.

Combined statistics over all distances for Goldfish are dominated by results from the 300 m distance, where concentrations are largest. This influence can be seen most clearly in the fractional bias values. The "combined" statistics for bias and RMS error indicate that CHARM and FOCUS achieved the best overall performance, although DEGADIS and TRACE both performed better than FOCUS at 1,000 and 3,000 m. These same measures also suggest that AIRTOX and SAFEMODE were the two poorest performing models. All of the models produced relatively high correlation coefficients. The number of cases for which maximum observed and predicted values agree within a factor

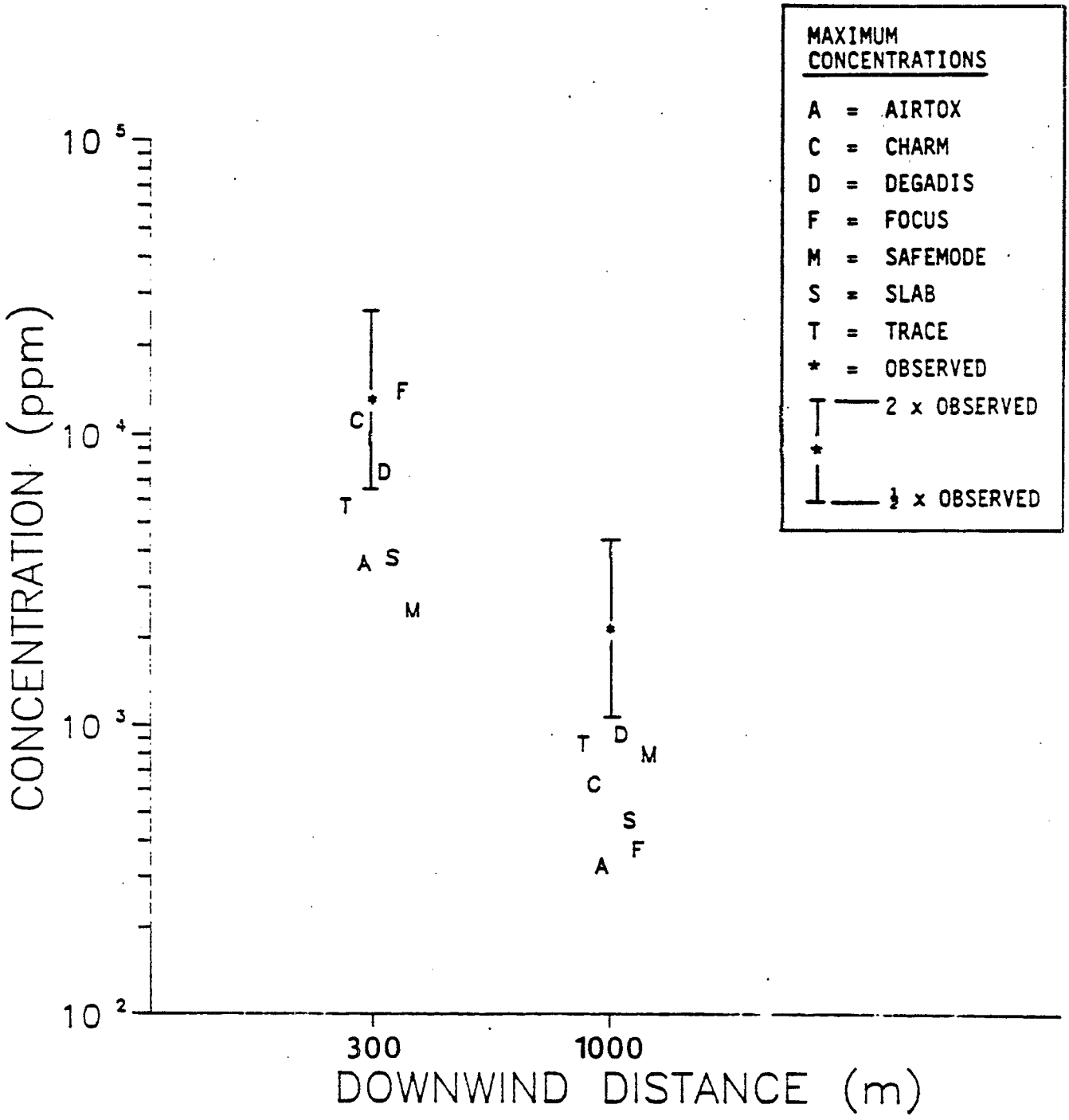


Figure 5-6. Observed and Predicted Maximum Concentrations versus Downwind Distance for Goldfish HF Test 3.

Table 5-3. Performance Statistics for Maximum Concentration Values for Goldfish HF Experiments

	TRACE	CHARM	AIRTOX	DEGADIS	SLAB	FOCUS	SAFEMODE
300m (N=3)							
average observed max value (ppm)	17,112	17,112	17,112	17,112	17,112	17,112	17,112
fractional bias	0.68	-0.07	1.17	0.45	1.11	0.09	1.47
average diff. (obs-pred)	8,721	-1,218	12,606	6,242	12,245	1,478	14,501
95 percent confidence int.	(5,557)	(7,738)	(11,068)	(3,045)	(11,782)	(9,499)	(14,077)
RMSE	9,003	3,345	13,370	6,361	13,132	4,100	15,569
number of max values within 2x	1/3	3/3	0/3	3/3	0/3	2/3	0/3
1000m (N=3)							
average observed max value (ppm)	2,142	2,142	2,142	2,142	2,142	2,142	2,142
fractional bias	0.50	0.68	1.24		0.92	0.89	0.81
average diff. (obs-pred)	835	1,088	1,639	707	1,347	1,317	1,236
95 percent confidence int.	(850)	(912)	(1,416)	(999)	(1,170)	(852)	(1,510)
RMSE	903	1,149	1,735	814	1,428	1,361	1,377
number of max values within 2x	2/3	2/3	0/3	2/3	0/3	1/3	1/3
3000m (N=1)							
average observed max value (ppm)	411	411	411	411	411	411	411
fractional bias	0.26	0.46	0.99	0.01	0.76	0.98	-0.23
average diff. (obs-pred)	96	153	273	3	226	270	-109
RMSE	96	153	273	3	226	270	109
number of max values within 2x	1/1	1/1	0/1	1/1	0/1	0/1	1/1
Combined(N=7)							
average observed max value (ppm)	8,311	8,311	8,311	8,311	8,311	8,311	8,311
fractional bias	0.66	0.00	1.17	0.44	1.09	0.16	1.36
average diff. (obs-pred)	4,109	-34	6,144	2,979	5,858	1,237	6,729
95 percent confidence int.	(3,947)	(2,142)	(5,861)	(2,736)	(5,885)	(2,354)	(7,130)
RMSE	5,924	2,316	8,827	4,198	8,648	2,830	10,232
number of max values within 2x	4/7	6/7	0/7	6/7	0/7	3/7	2/7
Correlation coefficient	0.992	0.990	0.999	0.994	0.987	0.963	0.954

of 2 provides a different and more balanced indicator of performance, which suggests that CHARM and DEGADIS produced the most consistent agreement with observations, while AIRTOX and SLAB were the poorest performers.

5.2.2 Cloud Half-Width

The results for cloud half-width for Goldfish are summarized in Table 5-4. Average differences indicate overprediction by all seven models for the one case at 3,000 m, but mixed results at 300 m and 1,000 m.

SAFEMODE shows consistent overprediction bias, with particularly large bias at 300 m. TRACE and DEGADIS also overpredict cloud half-widths, while the remaining four models tend to underpredict. No direct relation between half-width and maximum concentration is evident. CHARM and DEGADIS provide relatively good agreement with observed half-widths and also performed well for maximum values. AIRTOX performed comparatively well for half-widths, despite relatively poor results for maximum concentration.

5.3 Burro LNG Spill Experiments

For the three Burro experiments selected for this evaluation study, concentrat on measurements were taken at four distances from the release location: 57 m, 140 m, 400 m and 800 m. For Burro Test 5 (B-5), measurements at 400 m were not adequate to determine the maximum observed concentration and cloud half-width. For B-8, observed values were not determined at 140 m. Model predictions were not obtained for FOCUS at 57 m nor for CHARM at 57 m and 140 m for reasons discussed in Section 4.5.

5.3.1 Maximum Concentrations

The observed and predicted maximum concentrations at each distance for test B-3 are illustrated in Figure 5-7. Over the range of distances

Table 5-4. Performance Statistics for Cloud Half-Width for Goldfish HF Experiments

	TRACE	CHARM	AIRTOX	BEGADIS	SLAB	FOCUS	SAFEMODE
300m (N=3)							
average observed value (m)	42	42	42	42	42	42	42
average diff. (obs-pred)	-24	8	18	-16	16	25	-92
1000m (N=3)							
average observed value (m)	95	95	95	95	95	95	95
average diff. (obs-pred)	-21	10	14	-7	20	30	-39
3000m (N=1)							
average observed value (m)	112	112	112	112	112	112	112
average diff. (obs-pred)	-101	-98	-128	-83	-47	-62	-75
Combined(N=7)							
average observed value (ppm)	75	75	75	75	75	75	75
average diff. (obs-pred)	-34	-6	-4	-22	9	15	-67

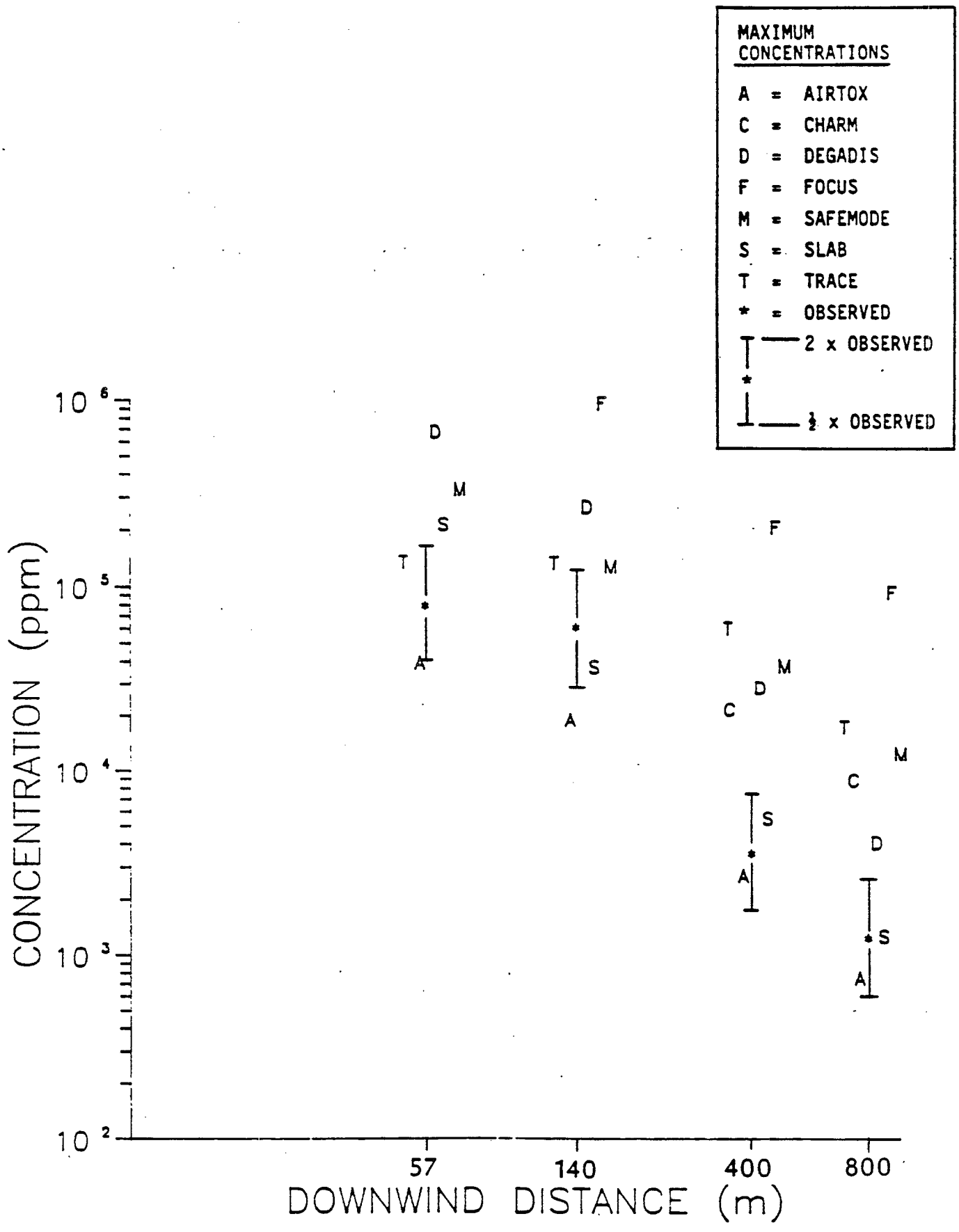


Figure 5-7. Observed and Predicted Maximum Concentrations versus Downwind Distance for Burro LNG Spill Test 3.

considered, predictions from two models, AIRTOX and SLAB, gave maximum concentrations comparable to observed values. AIRTOX underpredicted the maximum observed value at all four distances, while SLAB showed no consistent bias. TRACE, CHARM, DEGADIS and SAFEMODE all overpredicted substantially, while FOCUS overpredicted by more than an order of magnitude.

Results for experiment B-5 are illustrated in Figure 5-8. The observed maximum concentrations at 57 m and 140 m for B-5 are quite similar to B-3, but the observed maximum at 800 m is a factor of 2 higher for B-5. The same general pattern of model performance which was described for B-3 is evident for B-5, with overprediction by all of the models except AIRTOX and SLAB.

For test B-8, observed maximum concentrations are substantially higher, and model performance was quite different, as illustrated in Figure 5-9. SLAB again predicted maximum concentrations comparable to observed at all distances, while AIRTOX and SAFEMODE produced large overprediction. TRACE predicted the maximum concentration reasonably well at 57 m, but overpredicted by more than a factor of 2 at 400 m and 800 m. DEGADIS overpredicted substantially at 57 m, but gave better agreement at 400 m and 800 m. FOCUS overpredicted by about a factor of 2 at 400 m but gave good agreement at 800 m, while CHARM gave good agreement at 400 m and underpredicted at 800 m.

Performance statistics for maximum concentrations for Burro are summarized in Table 5-5. At 57 m, the fractional bias is negative for all five models, indicating overprediction. At this distance, TRACE achieved the lowest bias, smallest RMS error, and factor of 2 agreement for two of the three tests. DEGADIS produced the largest fractional bias, with overprediction by a factor of 4.7, and the largest RMS error. SAFEMODE also showed large overprediction at 57 m, while SLAB and AIRTOX gave intermediate results.

At 140 m, SLAB achieved the smallest fractional bias and RMS error. AIRTOX underpredicted by a factor of 2.2, while TRACE and SAFEMODE

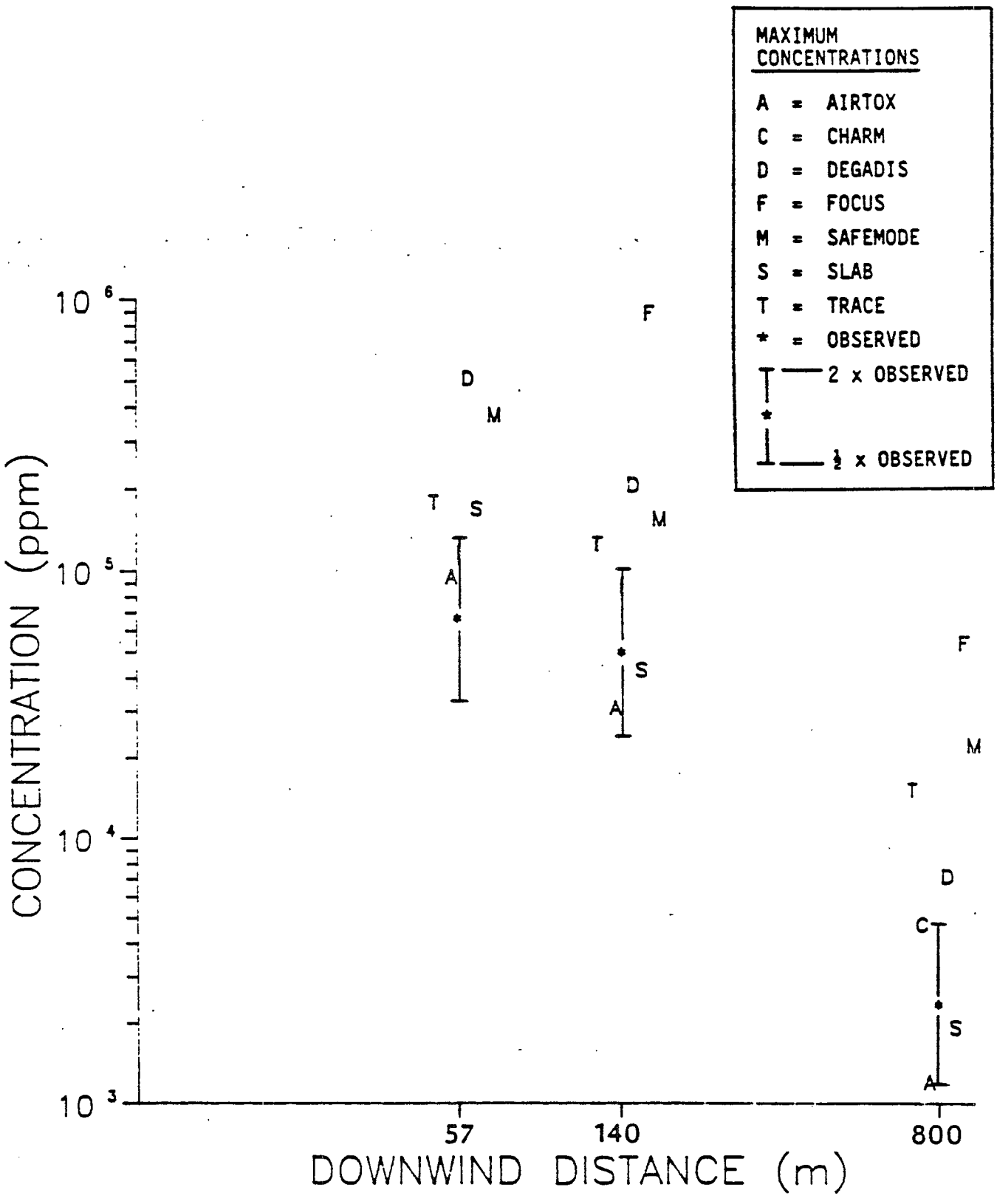


Figure 5-8. Observed and Predicted Maximum Concentrations versus Downwind Distance for Burro LNG Spill Test 5.

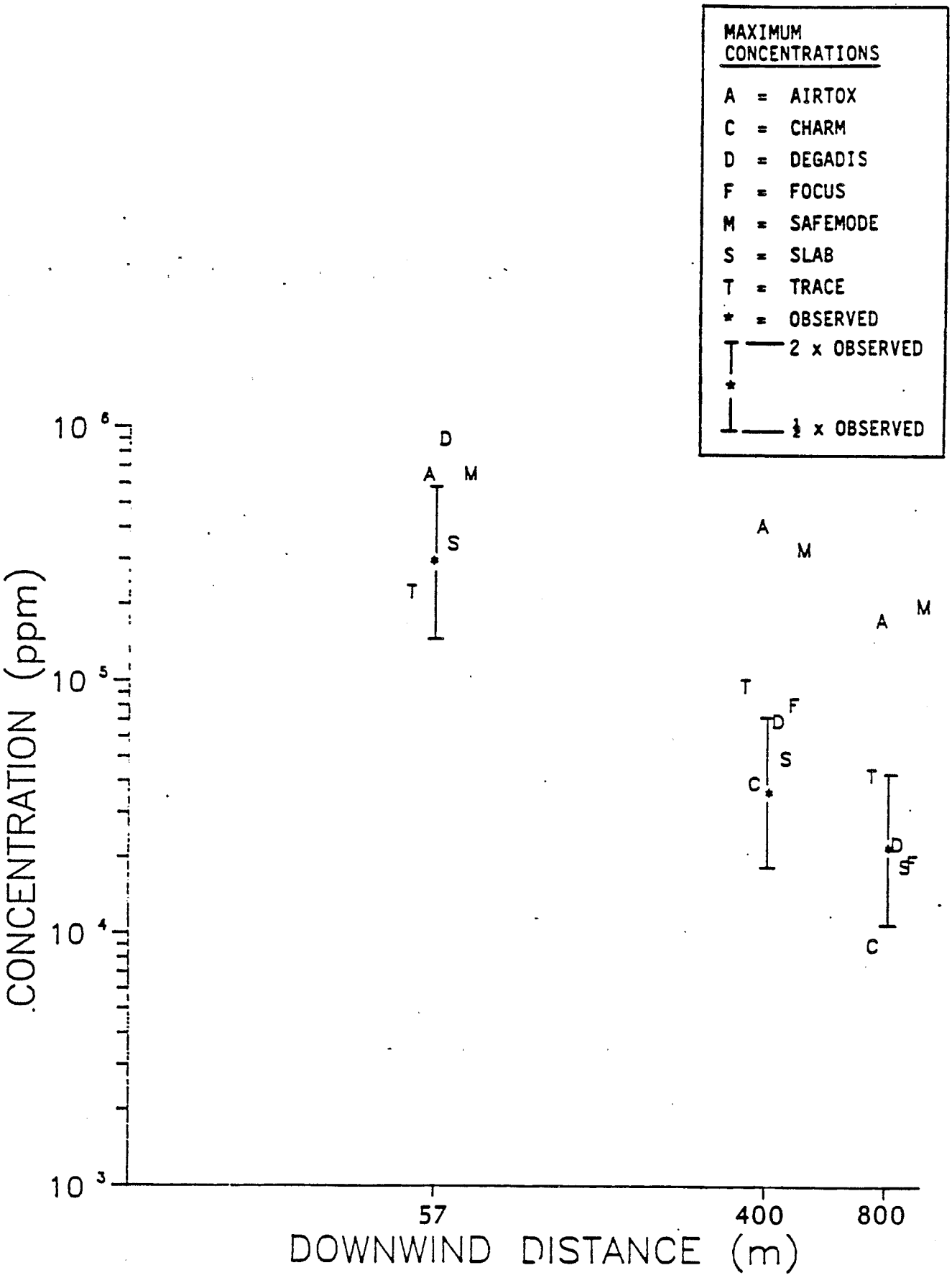


Figure 5-9. Observed and Predicted Maximum Concentrations versus Downwind Distance for Burro LNG Spill Test 8.

Table 5-5. Performance Statistics for Maximum Concentration Values for Burro LNG Experiments

	TRACE	CHARM	AIRTOX	HEGADIS	SLAB	FOCUS	SAFEMODE
57m (N=3)							
average observed max value (ppm)	140,289		140,289	140,289	140,289		140,289
fractional bias	-0.18		-0.55	-1.30	-0.48		-1.01
average diff. (obs-pred)	-28,297		-107,354	-516,044	-88,711		-288,437
95 percent confidence int.	(184,206)		(401,467)	(166,515)	(82,670)		(105,736)
RMSE	79,373		194,026	520,380	94,748		291,561
number of max values within 2x	2/3		1/3	0/3	1/3		0/3
140m (N=2)							
average observed max value (ppm)	51,088		51,088	51,088	51,088	51,088	51,088
fractional bias	-0.81		0.75	-1.25	0.31	-1.78	-0.89
average diff. (obs-pred)	-69,708		27,783	-169,413	13,688	-815,543	-81,183
RMSE	69,713		29,314	170,828	15,175	815,790	83,017
number of max values within 2x	0/2		1/2	0/2	2/2	0/2	0/2
400m (N=2)							
average observed max value (ppm)	18,914	18,914	18,914	18,914	18,914	18,914	18,914
fractional bias	-1.18	-0.41	-1.65	-0.84	-0.32	-1.50	-1.61
average diff. (obs-pred)	-54,104	-9,741	-175,305	-27,221	-7,086	-114,936	-154,621
RMSE	54,139	11,959	248,506	27,534	8,853	136,021	197,561
number of max values within 2x	0/2	1/2	1/2	1/2	2/2	0/2	0/2
800m (N=3)							
average observed max value (ppm)	8,071	8,071	8,071	8,071	8,071	8,071	8,071
fractional bias	-0.98	0.14	-1.49	-0.28	0.17	-1.46	-1.60
average diff. (obs-pred)	-15,400	1,025	-47,379	-2,590	1,238	-43,446	-65,194
95 percent confidence int.	(7,504)	(20,361)	(169,203)	(3,850)	(3,666)	(88,778)	(178,264)
RMSE	15,693	8,261	82,974	3,018	1,926	56,257	96,956
number of max values within 2x	1/3	1/3	2/3	1/3	3/3	1/3	0/3
Combined(N=10)							
average observed max value (ppm)	58,508		58,508	58,508	58,508		58,508
fractional bias	-0.49		-0.79	-1.25	-0.35		-1.13
average diff. (obs-pred)	-37,871		-75,924	-194,917	-24,922		-153,250
95 percent confidence int.	(32,670)		(101,413)	(158,654)	(33,037)		(84,665)
RMSE	59,347		160,879	295,346	52,498		193,667
number of max values within 2x	3/10		5/10	2/10	8/10		0/10
Correlation coefficient	0.817		0.796	0.867	0.909		0.878

overpredicted by factors of 2.36 and 2.60, respectively. DEGADIS overpredicted by a factor of 4, and FOCUS by more than a factor of 10.

At 400 m, all seven models overpredicted. SLAB again achieved the best performance, while CHARM produced slightly higher fractional bias and RMS error. DEGADIS overpredicted by a factor of 2.45 at 400 m, and TRACE by a factor of 3.9, while AIRTOX, FOCUS and SAFEMODE overpredicted by more than a factor of 7.

At 800 m, SLAB again produced the smallest RMS error, underpredicted by a factor of 1.19, and matched the observed maximum value within a factor of 2 for three of three tests. DEGADIS and CHARM also produced relatively small fractional bias and RMS error. TRACE overpredicted by a factor of 2.9 at 800 m, while AIRTOX, FOCUS and SAFEMODE overpredicted by more than a factor of 6.

The combined statistics indicate that SLAB achieved the best overall performance for estimating maximum concentrations for the Burro tests: the smallest fractional bias and RMS error, highest correlation, and factor of 2 agreement for eight of ten data points. SLAB provided the best performance at three of the four distances, while TRACE performed best at 57 m. CHARM gave relatively good performance at 400 m and 800 m, but did not provide useful predictions at closer distances.

5.3.2 Cloud Half-Width

Results for Burro for cloud half-width are summarized in Table 5-6. At 57 m distance, TRACE underpredicts the cloud width by a factor of 2.8, but the four other models produced only small average differences. At 140 m, all of the models gave reasonable agreement with measured half-widths. At 400 m, AIRTOX produced relatively large overpredictions, while SAFEMODE greatly underestimated cloud widths. At 800 m, all of the models except SAFEMODE

Table 5-6. Performance Statistics for Cloud Half-Width for Burro LNG Experiments

	TRACE	CHARM	AIRTOX	DEGADIS	SLAB	FOCUS	SAFEMODE
57m (N=3)							
average observed value (ppm)	34		34	34	34		34
average diff. (obs-pred)	22		-7	4	-4		9
140m (N=2)							
average observed value (ppm)	36		36	36	36	36	36
average diff. (obs-pred)	9		7	2	10	1	13
400m (N=2)							
average observed value (ppm)	98	98	98	98	98	98	98
average diff. (obs-pred)	44	-36	-61	27	0	-13	70
800m (N=3)							
average observed value (ppm)	52	52	52	52	52	52	52
average diff. (obs-pred)	-25	-139	-153	-42	-57	-134	20
Combined (N=10)							
average observed value (ppm)	53		53	53	53		53
average diff. (obs-pred)	10		-59	-5	-16		25

overpredicted the half-widths. No direct relationship is evident between model performance for maximum concentration and half-width measures.

5.4 Inter-Model Comparison

An exploratory analysis by EPA (Cox, 1990) was undertaken to assess whether the apparent differences in performance results between models are statistically significant. For this analysis, the performance measure of concern is the fractional bias of maximum values. Model performance results were grouped by source-receptor distance (two distance categories) and by experimental data base (three data bases). A distance cut-off of 300 m was used to separate results into "near-field" and "far-field" groups. Three models (CHARM, FOCUS, and SAFEMODE) did not provide predictions for both distance categories for all tests; these models were excluded from this exploratory analysis to avoid the complications posed by unequal numbers of data points.

The statistical technique chosen for this analysis is a multivariate analysis of variance (MANOVA). In this context, the data base consists of eight "dependent" variables (four models at two distances) and three "independent" variables (data bases). MANOVA can be used to test hypotheses involving the influence of independent variables on dependent variables, or relationships between dependent variables. The exploratory analysis addressed three hypotheses:

H1: Does model performance (i.e., fractional bias) vary between experimental data bases?

H2: Does model performance vary with distance category?

H3: Does performance vary between models?

In Table 5-7, the arithmetic average fractional bias for each model and each distance category are listed. A quick inspection of Table 5-7 indicates

TABLE 5-7

AVERAGE FRACTIONAL BIAS OF MAXIMUM OBSERVED AND PREDICTED VALUES BY MODEL AND BY DISTANCE CATEGORY

Distance Category	Model	Experimental Data Base: Burro AFB*	Experimental Data Base: Tortoise AFB*	Experimental Data Base: Goldfish AFB*	Combined (All Experiments) AFB*
Near-Field (≤300 m)	AIRTOX	0.06	-1.39	1.16	-0.06
	DEGADIS	-1.27	-1.55	0.47	-0.78
	SLAB	-0.24	0.87	1.10	0.58
	TRACE	-0.41	-0.30	0.71	0.00
Far-Field (>300 m)	AIRTOX	-0.19	0.12	1.22	0.38
	DEGADIS	-0.88	-0.25	0.37	-0.25
	SLAB	-0.03	0.89	0.93	0.60
	TRACE	-1.32	-0.01	0.51	-0.28
Combined (all distances)	AIRTOX	-0.06	-0.63	1.19	0.16
	DEGADIS	-1.07	-0.90	0.42	-0.52
	SLAB	-0.13	0.88	1.01	0.59
	TRACE	-0.87	-0.16	0.61	-0.14

* AFB = average fractional bias

several patterns within the results. The models tended to overpredict (negative AFB) for Burro and to underpredict for Goldfish HF, while DT results were mixed. The MANOVA test results for H1 (difference in performance between data sets) confirms this visual result, indicating that fractional bias (across all models) is significantly different among the three data bases. Model performance varies more significantly between data sets for the near-field distance category.

The second hypothesis (H2) tests whether fractional bias varies by distance category. MANOVA results indicate that differences between near-field and far-field fractional bias, across all experiments, are significant for DEGADIS and TRACE, but not for AIRTOX and SLAB.

The third hypothesis (H3) tests whether there are systematic differences in fractional bias between models. To reduce the number of model pairs for this exploratory analysis, TRACE was arbitrarily compared with each of the other three models. Results indicate significant differences in fractional bias for DEGADIS and SLAB, compared to TRACE, but no significant (systematic) difference between AIRTOX and TRACE, over all data bases. For all three data bases, the difference in average fractional bias between TRACE and DEGADIS is positive, indicating that maximum concentrations predicted by TRACE are generally lower than DEGADIS predictions. The differences between TRACE and DEGADIS are largest for Desert Tortoise. Differences between TRACE and SLAB, by contrast, are generally negative, indicating that TRACE predicted higher maximum concentrations than SLAB. Differences are again largest for Desert Tortoise.

This exploratory analysis indicates that multivariate analysis of variance is a potentially useful technique for comparing model performance among multiple data bases. Because the size of the data set is extremely limited (three experiments per data base), results should be viewed with caution.

5.5 Summary of Model Performance

The principal features of statistical results from each experimental program (Desert Tortoise ($DT-NH_3$), Goldfish (G-HF), and Burro (B-LNG)) are summarized below for each dense gas model. A tabulation of observed and predicted maximum concentrations and fractional bias values for each experiment is given in Table 5-8.

Some of the models evaluated in this study have been developed or tested previously using one or more of the data bases. Many of the model developers provided detailed example input streams to TRC for certain experiments, indicating a prior working knowledge of the data bases being used in the evaluation. Any experimental data which has been used during the development of a model does not provide an independent test of model performance. The data sets and experimental reports used in this evaluation have been available to the public for many years.

When the performance statistics for maximum concentration values at different distances are combined (Tables 5-1 through 5-3), the fractional bias results are generally dominated by those values at the near-field distences, where concentrations are lagrest. At these distances, predictions are strongly dependent on source characterization. It is important to examine model performance as a function of distance, and not to rely soley on combined statistics.

The dense gas models evaluated in this study vary widely in their design and technical complexity. Some models provide more rigorous treatment of physical and chemical processes associated with source characterization and dispersive behavior, while others incorporate many simplifying assumptions. The performance of a given model for a given data base depends as much or more

TABLE 5-8

SUMMARY CHART OF MODEL PREDICTED AND OBSERVED
CENTERLINE CONCENTRATIONS, AND FRACTIONAL BIAS

EXPERIMENT TEST #	HEIGHT (m)	CENTERLINE DISTANCE (m)	C (MAX) PPM OBSERVED	C (MAX) PPM TRACE	FRACTIONAL BIAS TRACE	C (MAX) PPM CHARM	FRACTIONAL BIAS CHARM	C (MAX) PPM AIRTOX	FRACTIONAL BIAS AIRTOX	C (MAX) PPM DEGADIS	FRACTIONAL BIAS DEGADIS	C (MAX) PPM SLAB	FRACTIONAL BIAS SLAB	C (MAX) PPM FOCUS	FRACTIONAL BIAS FOCUS	C (MAX) PPM SAFEMOD	FRACTIONA BIAS SAFEMODE
DT1	1	100	42483	62250	-0.38			375519	-1.59	288000	-1.49	18800	0.77	133400	-1.03		
DT1	1	800	6975	6882	0.01	8074	-0.15	3001	0.80	7198	-0.03	2980	0.80	8824	-0.23	3638	0.63
DT2	1	100	68664	79488	-0.15			313333	-1.28	469740	-1.49	21100	1.06	144100	-0.71		
DT2	1	800	8991	11441	-0.24	2917	1.02	5931	0.41	13760	-0.42	3840	0.80	10190	-0.13	5037	0.56
DT4	1	100	49307	72181	-0.38			226266	-1.28	566000	-1.68	21400	0.79	138530	-0.95		
DT4	1	800	14562	12081	0.19	6321	0.79	35943	0.85	19500	-0.29	4374	1.08	15360	-0.05	6249	0.80
G1	1	300	25473	13588	0.61	30920	-0.19	6582	1.17	17500	0.37	6520	1.18	27170	-0.06	2968	1.58
G1	1	1000	3098	2211	0.33	1868	0.50	813	1.24	2321	0.29	1340	0.79	1699	0.58	1145	0.92
G1	1	3000	411	315	0.26	258	0.46	138	0.99	408	0.01	185	0.76	141	0.98	520	-0.23
G2	1	300	13347	6226	0.73	13520	-0.01	3501	1.17	8050	0.50	4490	0.99	6490	0.69	2499	1.37
G2	1	1000	1287	894	0.36	702	0.59	390	1.07	1103	0.15	600	0.73	425	1.01	822	0.44
G3	1	300	12515	5359	0.80	10550	0.17	3435	1.14	7060	0.56	3590	1.11	13240	-0.06	2366	1.36
G3	1	1000	2042	816	0.86	592	1.10	308	1.48	881	0.79	445	1.28	352	1.41	753	0.92
B3	1	57	74000	125401	-0.52			36465	0.68	631000	-1.58	199000	-0.92			308703	-1.23
B3	1	140	54640	123472	-0.77			17507	1.03	246000	-1.27	34400	0.45	890260	-1.77	116471	-0.74
B3	1	400	3321	55474	-1.77	20000	-1.43	2490	0.29	26400	-1.55	5100	-0.42	191000	-1.93	34967	-1.65
B3	1	800	1167	16043	-1.73	8180	-1.50	698	0.50	3760	-1.05	1180	-0.01	85700	-1.95	11485	1.63
B5	1	57	63478	168815	-0.91			90184	-0.35	485000	-1.54	160000	-0.86			355297	-1.39
B5	1	140	47535	118118	-0.85			29103	0.48	195000	-1.22	40400	0.16	843000	-1.79	146070	-1.02
B5	1	800	2236	14226	-1.46	4425	-0.66	1133	0.65	6723	-1.00	1820	0.21	50630	1.83	20929	-1.61
B8	1	57	283390	211543	0.29			616281	-0.74	853000	-1.00	328000	-0.15			622197	-0.75
B8	1	400	34507	90562	-0.90	37310	-0.08	385947	-1.67	65870	-0.62	46900	-0.30	76700	-0.76	312103	-1.60
B8	1	800	20810	40144	-0.63	8532	0.84	164520	-1.55	21500	-0.03	17500	0.17	18220	0.13	187380	-1.60

upon that model's ability to simulate a specific release scenario as it does on the model's treatment of dispersive behavior. Design features which may influence model performance for each data base are noted below as results for each model are summarized.

TRACE

Relative to the other models, TRACE provided the best performance for DT and intermediate performance for G-HF and B-LNG. For DT-NH_3 experiments, TRACE performed well for predicting maximum concentrations at both the 100 m and 800 m distances, with little bias and small RMS error. For G-HF releases, TRACE underpredicted maximum concentrations at 300 m by a factor of 2, but showed less bias at 1,000 m and 3,000 m. For B-LNG spill tests, TRACE performed relatively well for estimating maximum concentrations at 57 m, but consistently overpredicted by more than a factor of 2 at 140 m, 400 m and 800 m. TRACE overpredicted cloud half-width values for DT-NH_3 and G-HF, but underpredicted for B-LNG.

TRACE requires a relatively extensive set of inputs as evidenced by Table 4-9. The model design allows the flexibility necessary to model adequately a variety of chemicals and release scenarios. The model developer's comments and letters suggest a thorough understanding and familiarity with the Desert Tortoise data base and some familiarity with Goldfish and Burro. Extensive user documentation provided with TRACE, in conjunction with technical support, allowed effective interpretation of the TRACE inputs.

CHARM

CHARM did not provide near-field predictions for either the DT-NH_3 or B-LNG tests. Relative to the other models, CHARM was among the best

performers for DT-NH_3 and G-HF, and intermediate for B-LNG. For DT-NH_3, CHARM underpredicted maximum concentrations at 800 m. For G-HF, CHARM performed relatively well for estimating maximum concentrations at all distances, with some underprediction bias at 1,000 m and 3,000 m. For B-LNG, CHARM overpredicted maximum concentration by a factor of 1.5 at 400 m and showed minimal bias but significant scatter at 800 m. CHARM underpredicted cloud half-width values for DT-NH_3, showed minimal bias for G-HF, and overpredicted for B-LNG.

CHARM is a puff model. Problems were encountered with this model in the near-field, due to a combination of the "puff" algorithm and the momentum jet simulation. Concentration predictions at near-field receptors were intermittent and erratic, symptomatic of gaps between successive puffs. According to the model developer, these problems have been resolved in a version of CHARM released after this evaluation began.

AIRTOX

Relative to the other models, AIRTOX was among the poorest performers for DT-NH_3 and G-HF, and intermediate for B-LNG. For DT-NH_3, AIRTOX produced substantial overprediction of maximum concentrations at 100 m and large scatter, but less bias, at 800 m. For G-HF, AIRTOX gave substantial underprediction of maximum concentrations at all distances. For B-LNG, AIRTOX results for maximum concentration showed a mixed pattern, with overprediction by a factor of 10 for one test, but agreement within a factor of 2 for 5 of 7 data points for the other two tests. AIRTOX predicted cloud half-widths with little bias for DT-NH_3 and G-HF, but overpredicted for B-LNG.

According to the model developer, near-field predictions by AIRTOX for Desert Tortoise are sensitive to the momentum jet simulation. The large overpredictions here suggest that the initial cloud size was underestimated

for these tests. Conversely, for Goldfish, AIRTOX systematically overestimated the rate of dispersal of the HF clouds (and thereby underestimated concentrations).

For the Burro simulations, the non-jet mode was used and AIRTOX showed less systematic bias. The performance in the Burro simulations was accomplished without explicit water effects specified and the only pool descriptors being dike area and minimum pool depth.

DEGADIS

DEGADIS was the best performing model for G-HF and gave intermediate results for DT-NH_3 and B-LNG. For DT-NH_3, DEGADIS gave large overprediction of maximum concentrations at 100 m, but relatively good agreement at 800 m. For G-HF, DEGADIS underpredicted maximum concentrations by roughly a factor of 1.5 at all distances. For B-LNG, DEGADIS again gave large overprediction at 57 m and 140 m, but showed smaller overprediction bias at 400 m and 800 m. DEGADIS overpredicted cloud half-widths for both DT-NH_3 and G-HF, but showed little bias for B-LNG.

The DEGADIS model developer provided example run set-ups for the Burro and DT data bases with the model literature and indicated that the model had been tested on these experiments. The developers have also published test results simulating the Goldfish releases. There is comprehensive documentation available describing DEGADIS. Theory, source code and operational direction are all readily available.

For Desert Tortoise, the near-field overprediction indicates that DEGADIS has systematically underestimated the initial cloud dispersion associated with these pressurized releases. DEGADIS performed relatively well for DT at 800 m, and performance for the Goldfish HF releases was also comparatively good.

For the Burro LNG tests, DEGADIS is the only model which accounts for the potential entrainment of water from the pool into the vapor cloud, due to the vigorous boiling of the LNG spill. Despite this relatively sophisticated treatment, DEGADIS overpredicted maximum concentrations at 57 m and 140 m by more than a factor of 4.

SLAB

Relative to the other models, SLAB provided the best performance for B-LNG, but was one of the poorest performers for both DT-NH_3 and G-HF. For DT-NH_3, SLAB consistently underpredicted maximum concentrations by a factor of 2.5 at both 100 m and 800 m. For G-HF, SLAB underpredicted maximum concentrations consistently by a factor of 3 at all distances. For B-LNG, SLAB provided relatively good agreement with observed maximum concentrations at 140 m, 400 m and 800 m, with overprediction by a factor of 1.6 at 57 m. For cloud half-width, SLAB underpredicted for DT-NH_3 and G-HF, but showed little bias for B-LNG.

SLAB is a public domain model and the source code is available for examination. This model has probably been tested against the data bases used in this evaluation. SLAB requires chemical properties to be input but suggested chemical properties for NH_3, HF and LNG (Methane) are listed in the SLAB Users Guide.

Systematic bias is seen in SLAB simulations of the DT and HF releases while the Burro simulations appear relatively successful. The primary difference in the input streams between these simulations is the release type. DT and HF tests were simulated as horizontal jets, and the predictions systematically underestimated observed values. Apparently, the SLAB jetting algorithm overestimates initial cloud dispersion.

FOCUS

Relative to the other models, FOCUS provided intermediate performance for all three experimental programs. For DT-NH_3, FOCUS overpredicted maximum concentrations at 100 m but provided good agreement at 800 m. For G-HF, FOCUS predicted maximum concentrations with little bias at 300 m, but underpredicted at both 1,000 and 3,000 m. For B-LNG, FOCUS produced overpredictions at 140 m, 400 m and 800 m. (FOCUS predictions were not analyzed at 57 m). Cloud half-width values predicted by FOCUS were lower than observed for DT-NH_3 and G-HF, but larger than observed for B-LNG, particularly at 800 m.

The FOCUS model developers provided test runs for all three data bases used in the evaluation. The model has been tested using all these data bases by the developers. FOCUS is a relatively new model in its current form, so little descriptive literature was available at the time of the evaluation. However, FOCUS utilizes a sophisticated input stream. Frequent contact with the model developer was necessary to run the model correctly.

The model yielded one million ppm at the 57 m arc for Burro indicating that the model simulated that receptor as in the pool of LNG. Overprediction for all of the Burro tests indicates that the predicted LNG cloud was not adequately dispersed.

SAFEMODE

Relative to the other models, SAFEMODE was an intermediate performer for DT-NH_3 and G-HF, but gave the poorest performance for B-LNG. For DT-NH_3, SAFEMODE underpredicted maximum concentrations at 800 m by roughly a factor of 2. For G-HF, SAFEMODE produced underpredictions at 300 m and 1,000 m, but gave reasonable agreement (for 1 test) at 3,000 m. For B-LNG, SAFEMODE overpredicted maximum concentrations consistently at all distances, by as much

as a factor of 10. Cloud half-width values predicted by SAFEMODE were larger than observed for DT-NH_3 and for G-HF, but smaller than observed for B-LNG.

SAFEMODE was a difficult model to apply for DT and GF because none of the model release scenarios adequately describes the type of experimental release. The model developer has stated that this type of release (controlled release with constant pressure and flowrate) is not typical of accidental scenarios.

The release scenario chosen is a "short pipe from tank." The input orifice size is calculated to generate the correct emission rate. This may introduce inaccuracy in the simulation because the artificial orifice size forces an artificial release velocity. Also, the model internally generates a flash fraction. In several instances, especially in the DT simulations, the internally generated values were altered to match information available from the experimental data. The resulting release scenario is highly artificial and may not be physically consistent.

SAFEMODE does not calculate concentrations within the jet region thus eliminating the 100 m DT arc from analysis. The model developers explain that the jet region is typically well inside the IDLH (immediate danger to life and health) concentration contour; therefore, it is computationally inefficient to compute concentrations in this region.

6.0 CONCLUSIONS

The evaluation of dense gas models has provided a basis for judging the performance of seven models as applied to three different experimental release scenarios. The findings from this study include conclusions drawn from the performance results obtained for each experimental program, plus insights gained during the preparation of model inputs and test packages.

Five conclusions summarize the principal findings from this study:

- The initial characterization of the dense gas release is critical for estimating concentrations over distances up to 1,000 m from the source. Models which contain very similar treatment of atmospheric dispersion and utilize the same meteorological inputs produce concentration predictions which differ by more than an order of magnitude, as a consequence of such "initial conditions". While source characterization is critical for good model performance, however, accurate estimation of near-field concentrations is not sufficient to guarantee good performance at greater distances.

- Among the models evaluated in this study, none demonstrated good performance consistently for all three experimental programs. Different models performed more effectively for different release scenarios, reflecting the advantages and disadvantages of the various design features which characterize each model. Given the complexities of dense gas dispersion, and of the models, it was not feasible to attribute model performance to any specific algorithms or design features. Over all three programs, two models, TRACE and CHARM, provided agreement within a factor of 2 for more than half of the observed and predicted maximum values, while DEGADIS and SLAB each provided the best performance for one experimental program.

- An equitable, "hands-off" evaluation of air toxics models was very difficult to achieve as a practical objective. Many of the proprietary models, in particular, 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. The data bases used in this study are publicly available and have been used previously to develop and test dense gas models.

- Testing with the selected experimental programs provides valuable insights regarding the performance of dense gas models, but these programs may not represent a realistic test of the models for their intended application to accidental releases of toxic air pollutants. For several models, design features which enhance and simplify their use for accidental releases imposed limitations

which interfere with the simulation of "controlled" releases. Given the differences in model performance obtained for the three experimental programs, and the small number of releases considered, it is highly uncertain how well any of these models might perform for any other dense gas release scenarios.

- Responses to this project were generally favorable. Of seven model developers, most agreed with the technical approach and application of their model within the evaluation study. Two had no comments. Four had suggestions for improving the performance of their models and offered a number of constructive comments. One model developer expressed serious reservations concerning the method, results and conclusions of the study.

7.0 REFERENCES

Blewitt, D.N., Yohn, J.F., Koopman, R.P., and Brown, T.C., 1987: Conduct of Anhydrous Hydrofluoric Acid Spill Experiments, International Conference on Vapor Cloud Modeling, John Woodward, ed., American Institute of Chemical Engineers.

Cox, William, Personal Communication, Letter dated October, 1990.

Goldwire, H.C., McRae, T.G., Johnson, G.W., Hipple, D.L., Koopman, R.P., McClure, J.W., Morris, L.K., and Cederwall, R.T., 1985: Desert Tortoise Series Data Report -- 1983 Pressurized Ammonia Spills, Lawrence Livermore National Laboratory Report UCID-20562, Livermore, California.

Hanna, S. and Drivas, P., 1987: Guidelines for Use of Vapor Cloud Dispersion Models, Center for Chemical Process Safety, AIChE, New York, New York.

Koopman, R., Baker, J., Cederwall, R., Goldwire, H., Hogan, W., Kamppinen, L., Kiefer, R., McClure, J., McRae, T., Morgan, D., Morris, L., Spann, M., and Lind, C., 1982: Burro Series Data Report LLNL/NWC 1980 LNG Spill Tests. Lawrence Livermore National Laboratory Report UCID-19075, Vol. 1, Livermore, California.

McNaughton, D.J., 1987: Data Archiving Protocol Preparation for Toxic Model Evaluation, EPA Internal Report, EPA Contract 68-02-3886, Work Assignment No. 62.

McNaughton, D., Atwater, M., Bodner, P. and Worley, G., 1986: Evaluation and Assessment of Models for Emergency Response Planning, Prepared for the Chemical Manufacturers Association, Washington, DC.

McNaughton, D., and Bodner, P., 1988: A Workbook of Screening Techniques for Assessing Impacts of Toxic Air Pollutants, EPA Contract No. 68-02-3886, Work Assignment No. 7.

Panofsky, H.A. and Dutton, J.A., 1984: Atmospheric Turbulence: Models and Methods for Engineering Applications, John Wiley and Sons, Inc., New York, New York.

Spicer, T.O. and Havens, J. 1988a: Development of Vapor Dispersion Models for Non-Neutrally Buoyant Gas Mixtures - Analysis of TFI/NH_3 Test Data, USAF Engineering and Services Laboratory, Final Report.

Spicer, T.O. and Havens, J., 1988b: Modeling HF and NH_3 Spill Test Data Using DEGADIS, Paper 87b, Summer National Meeting of the American Institute of Chemical Engineers.

Worley, G., Bodner, P. and McNaughton, D., 1986: Technical Memorandum: Air Toxics Models, Identification and Characterization, Prepared for EPA SRAB under Contract No. 68-02-3886, Work Assignment No. 45.

Zapert, J.G. and Londergan, R.J., 1990: Toxic Model Evaluation Data Archiving Report, EPA Final Report, EPA Contract 68-02-4399, Work Assignment No. 62.

APPENDIX A

STATISTICAL PROTOCOL FOR EVALUATION OF AIR TOXICS MODELS

1.0 INTRODUCTION

A set of dispersion models designed for application to releases of toxic gases will be evaluated by comparing observed and predicted concentrations for a number of field experimental programs. These field programs represent a variety of source characteristics, including both heavier-than-air and neutrally buoyant gases, evaporating pools, jet releases, and both instantaneous (puff) and continuous releases. Each model may only be applicable to some of the experimental programs. The model performance results obtained with each experimental data base will emphasize specific model features and components. This protocol describes the general approach to performance testing, and then defines specific procedures to be used for the initial evaluation of dense gas models.

1.1 Sampling and Averaging Times

All of the experimental data bases which have been selected and archived for the air toxics evaluation have involved near-ground releases, with concentration measurements collected for an array of samplers deployed along lines or arcs at selected distances downwind of the source. The sampling times associated with concentration measurements and meteorological data collected during the experiments set a lower limit to the averaging time for which model performance can be tested. The duration of the release and the source dimensions also influence the time scale of each experiment. Pertinent information for the experimental programs in the data archive is summarized in Table 1-1.

All of the experimental programs except the WSU area source simulations involve instantaneous or time-varying source characteristics, and

TABLE 1-1

SUMMARY OF DATA BASE CHARACTERISTICS FOR AIR TOXICS EVALUATION

Experimental Program	Number of Tests[a]	Source Dimensions or Characteristics	Duration of Release	Averaging Times[a]		Number of Samplers
				Concentration Measurements	Meteorological Measurements	
Hanford	6	Point	Instantaneous	9.6s	60s	40 (2 arcs)
WSU	14	150m x 200m (forest canopy)	1 hour (continuous)	1 hour	1 hour	20 (2 arcs)
Thorney Island						
(a)	9	15m high 14m diameter (volume)	Instantaneous	60s	Test average	37 (5 arcs)
(b)	2	Point	Continuous	60s	Test average	37 (5 arcs)
Desert Tortoise	3	Horizontal Jet	2-7 min.	60s	Test average	12 (2 arcs)
Burro LNG	4	Spill (LNG onto water) Pond diameter 58m	2-3 min.	10s	Test average	30 (4 arcs)
Goldfish HF	3	Horizontal Jet	2-6 min.	60-100s	Test average	35 (3 arcs)

[a] As archived

concentration measurements provide time resolution of 2 minutes or less. These short-duration concentration values are referred to as "instantaneous" concentrations in the discussions which follow. The models which are applicable to these experimental programs will also predict time-varying concentration fields. For the time scales and transport distances (800m or less) involved in these experiments, steady state (test average) meteorological conditions will generally be used as model inputs.

To facilitate comparisons between tests and experimental programs, all concentration values will be normalized to the mass emission rate of each tracer gas.

2.0 PERFORMANCE MEASURES AND STATISTICS

The performance measures which will be used to compare observed and predicted concentrations for the air toxics models have been chosen to obtain a thorough characterization of model performance for each experimental program, plus basic measures for each individual experiment. The small number of tests (14 or less) in each program makes it important to examine model performance for individual tests, since combined results may be dominated by one or two tests. Measures have been selected to characterize bias, scatter, range, and correlation based on comparisons of observed and predicted values for each line or arc of samplers. (For near-ground releases, concentrations will consistently decrease with distance from the source. Results from samplers closest to the source would dominate if data from all arcs were combined.) Measures will be computed for both peak instantaneous and test average concentrations. Averages will be computed for the period when significant impacts were observed on each arc. This period ranges between one and ten minutes in duration, depending upon the experiment, except for the 1-hour WSU tests. In addition, predicted and measured plume widths, vertical

concentration profiles and times of maximum impact will be tabulated (where appropriate) as diagnostic tools relating to model performance. The proposed performance measures are summarized in Table 2-1 and are discussed below.

Bias. Observed and predicted maximum, mean, and median values will be compared for each sampling arc.

Scatter. Measures of scatter are based on observed and predicted values paired in time and location. Measures will include the standard deviation of residuals (σ_d), the RMS error and average absolute difference.

Range. The comparison of the standard deviations of observed and predicted concentration values provides a measure of prediction range. Comparisons of observed and predicted frequency distributions of "instantaneous" (e.g., 60s) concentrations will also be made, including 90, 75, 50 (median), 25 and 10 percent values, plus the maximum frequency difference.

Correlation. The Pearson correlation coefficient will provide a measure of the spatial and temporal correlation of observed and predicted concentration values.

For all instantaneous values, performance measures will test for prediction bias and range, but not for scatter or correlation. It is unreasonable to expect models to predict temporal and spatial variations in concentration values on a time scale of 10 to 60 seconds using test-average meteorological inputs.

2.1 Individual Tests

A simplified set of performance measures will be computed for individual tests. These measures will include the comparison of maximum observed and predicted concentrations; average concentration differences over all locations; and the cumulative frequency distributions of observed and predicted values.

TABLE 2-1

PERFORMANCE MEASURES FOR EACH EXPERIMENTAL PROGRAM
IN AIR TOXICS MODEL EVALUATION

MEASURE (each arc)	Peak Instantaneous Concentration (one per location, per test)	Test-Average Concentration (one per location, per test)	All Instantaneous Concentrations (many per location, each test)
a. Bias			
Difference of characteristic values (obs. - pred.)			
Maximum	X	X	
Mean (average)	X (C.I.)	X (C.I.)	
Median	X	X	X
b. Scatter			
RMS error	X	X	
σ_D	X (C.I.)	X (C.I.)	
AAD	X	X	
c. Range			
Variance comparison	X (C.I.)	X (C.I.)	
Frequency Distribution Percentiles	X	X	X
Maximum Frequency Difference			X (C.I.)
d. Correlation			
Pearson correlation coefficient	X (C.I.)	X (C.I.)	

This small set of composite measures will illustrate model performance for individual tests, to aid in interpreting the combined results for each experimental program. Graphical and tabular displays of performance results by test will also be prepared to summarize the performance of each model.

2.2 Confidence Intervals

Standard statistical methods will be used to compute confidence intervals wherever practical. Measures for which confidence intervals will be computed are shown in Tables 2-1 and 2-2. For concentration differences, confidence intervals will be based on a two-sample T test. For the standard deviation of residuals, the confidence interval is given by a Chi-Square test. For the variance comparison, the F test will be used. The K-S test provides a confidence level for the maximum frequency difference, and the Fisher Z test provides a confidence level for the correlation coefficient.

2.3 Dense Gas Model Evaluation

The dense gas model evaluation is planned as the initial phase of the overall effort to evaluate the performance of air toxics models. For this evaluation, the final three experimental data sets listed in Table 1-1 (Desert Tortoise, Burro LNG, and Goldfish HF) will be used.

The performance measures of interest for these experiments relate to test-period average concentrations. Most of the models being evaluated provide as output predicted centerline concentrations at distances selected internally by the model, plus parameters to describe the width of the concentration "footprint" downwind of the release point. The models generally do not give predictions at a large number of user-specified receptor locations. The list of performance measures has been reduced to include only

TABLE 2-2

PERFORMANCE MEASURES FOR INDIVIDUAL TESTS FOR AIR TOXICS MODEL EVALUATION

1. Comparison of Maximum Concentrations (unpaired) - each arc

   - peak instantaneous values
   - test average values

2. Difference of Mean Concentrations for All Samplers - each arc (obs.- pred.)

   - peak instantaneous values (C.I.)
   - test average values (C.I.)

3. Frequency Distribution Comparison (based on instantaneous concentration values)

   - 90, 75, 50, 25, 10 percentiles

measures which can be computed readily from both observed concentrations and model outputs. The proposed measures are summarized in Table 2-3.

The maximum or centerline concentration value (for each sampling arc) can be identified directly from the measures concentrations. It will generally be necessary to calculate the predicted centerline value at the arc distance by interpolating between centerline values at distances selected by the model.

The "half-width" distance is defined as the crosswind distance on a sampling arc from the centerline/maximum concentration to the point where the concentration drops to 50 percent of the maximum value. For measured concentrations, this distance will be determined by interpolating between sampling points on each arc, and averaging for the two sides of the "plume". For predicted concentrations, the model outputs will be used to compute an equivalent half-width value.

2.4 Comparisons Between Models

For each experimental program, the performance of different models will be compared using statistical results. While these results are useful for judging the relative performance of different models, primary emphasis will be placed on identifying strengths and weaknesses of each model, as indicated by comparisons with observed concentrations, and relating those results (where practical) to specific model features. This "diagnostic" approach is appropriate, given the variety of models and the different scenarios represented by the experimental programs. It does not appear practical to develop composite performance scores based on results obtained for one model with two or more experimental programs, given the differences between these programs.

TABLE 2-3

PERFORMANCE MEASURES FOR DENSE GAS MODEL EVALUATION

Measure	Individual Test	Each Experimental Program* (All Tests Combined)
a. Bias		
Difference of maximum values (obs-pred)		
>Each arc	X	X (C.I.)
>Average - all arcs	X	X (C.I.)
Fractional bias of maximum values		
>Each arc	X	X
>Average - all arcs	X	X
Difference of half-width values (obs-pred)		
>Each arc	X	X (C.I.)
>Average - all arcs	X	X (C.I.)
b. Scatter		
RMS Error - maximum values		
>Each arc		X
>All arcs combined		X
Standard deviation (σ_d) of maximum values		
>Each arc		X (C.I.)
>All arcs combined		X (C.I.)
c. Range		
Variance comparison - maximum values		
>Each arc		X (C.I.)
>All arcs combined		X (C.I.)
d. Correlation		
Pearson correlation coefficient - maximum values		
>Each arc		X
>All arcs combined		X

* C.I. indicates that a confidence interval will be computed for this measure.

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO.	2.	3. RECIPIENT'S ACCESSION NO.
EPA-450/4-90-018		

4. TITLE AND SUBTITLE
Evaluation of Dense Gas Simulation Models

5. REPORT DATE
May 1991

6. PERFORMING ORGANIZATION CODE

7. AUTHOR(S)
James G. Zapert, Richard Londergan & Harold Thistle

8. PERFORMING ORGANIZATION REPORT NO.

9. PERFORMING ORGANIZATION NAME AND ADDRESS
TRC Environmental Consultants
800 Connecticut Boulevard
East Hartford, CT 06108

10. PROGRAM ELEMENT NO.

11. CONTRACT/GRANT NO.
68-02-4399

12. SPONSORING AGENCY NAME AND ADDRESS
U.S. Environmental Protection Agency
Office of Air Quality Planning and Standards
Research Triangle Park, N.C. 27711

13. TYPE OF REPORT AND PERIOD COVERED
Final Report

14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES
EPA Technical Representative: Jawad S. Touma

16. ABSTRACT

This report describes the approach and presents the results of an evaluation study of seven dense gas simulation models using data from three experimental programs. The models evaluated are two in the public domain (DEGADIS and SLAB) and five that are proprietary (AIRTOX, CHARM, FOCUS, SAFEMODE, and TRACE). The data bases used in the evaluation are the Desert Tortoise Pressurized Ammonia Releases, Burro Liquefied Natural Gas Spill Tests and the Goldfish Anhydrous Hydroflouric Acid Spill Experiments. A uniform set of performance statistics are calculated and tabulated to compare maximum observed concentrations and cloud half-width to those predicted by each model. None of the models demonstrated good performance consistently for all three experimental programs.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Air Pollution Hazardous Waste Assessment Toxic Air Pollutants Dense Gas Models Air Quality Dispersion Models	Dispersion Modeling Meteorology Air Pollution Control	13B

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS *(This Report)*	21. NO. OF PAGES
Release Unlimited	Unclassified	107
	20. SECURITY CLASS *(This page)*	22. PRICE
	Unclassified	

EPA Form 2220-1 (Rev. 4-77) PREVIOUS EDITION IS OBSOLETE