

Hazard Response Modeling Uncertainty (A Quantitative Method)

Volume II Evaluation of Commonly Used Hazardous Gas Dispersion Models

HEALTH AND ENVIRONMENTAL SCIENCES
API PUBLICATION NUMBER 4546
OCTOBER 1992

American Petroleum Institute
1220 L Street, Northwest
Washington, D.C. 20005



Hazardous Response Modeling Uncertainty (A Quantitative Method)

Volume II Evaluation of Commonly Used Hazardous Gas Dispersion Models

Prepared for:

**American Petroleum Institute
Health and Environmental Sciences Department
and
Air Force Engineering and Services Center
Tyndall Air Force Base**

**PUBLICATION NUMBER 4546
OCTOBER 1992**

**PREPARED UNDER CONTRACT BY:
SIGMA RESEARCH CORPORATION
196 BAKER AVENUE
CONCORD, MASSACHUSETTS**

**American
Petroleum
Institute**



FOREWORD

API PUBLICATIONS NECESSARILY ADDRESS PROBLEMS OF A GENERAL NATURE. WITH RESPECT TO PARTICULAR CIRCUMSTANCES, LOCAL, STATE, AND FEDERAL LAWS AND REGULATIONS SHOULD BE REVIEWED.

API IS NOT UNDERTAKING TO MEET THE DUTIES OF EMPLOYERS, MANUFACTURERS, OR SUPPLIERS TO WARN AND PROPERLY TRAIN AND EQUIP THEIR EMPLOYEES, AND OTHERS EXPOSED, CONCERNING HEALTH AND SAFETY RISKS AND PRECAUTIONS, NOR UNDERTAKING THEIR OBLIGATIONS UNDER LOCAL, STATE, OR FEDERAL LAWS.

NOTHING CONTAINED IN ANY API PUBLICATION IS TO BE CONSTRUED AS GRANTING ANY RIGHT, BY IMPLICATION OR OTHERWISE, FOR THE MANUFACTURE, SALE, OR USE OF ANY METHOD, APPARATUS, OR PRODUCT COVERED BY LETTERS PATENT. NEITHER SHOULD ANYTHING CONTAINED IN THE PUBLICATION BE CONSTRUED AS INSURING ANYONE AGAINST LIABILITY FOR INFRINGEMENT OF LETTERS PATENT.

ACKNOWLEDGMENTS

THE FOLLOWING PEOPLE ARE RECOGNIZED FOR THEIR CONTRIBUTIONS OF TIME AND EXPERTISE DURING THIS STUDY AND IN THE PREPARATION OF THIS REPORT:

API STAFF CONTACT:

Howard Feldman, Health and Environmental Sciences

MEMBERS OF THE AIR MODELING TASK FORCE

Kenneth Steinberg, Exxon Research and Engineering Company

Thomas Baker, ARCO Oil and Gas Company

Douglas Blewitt, Amoco Corporation

Richard Carney, Phillips Petroleum Company

David Fontaine, Chevron Research and Technology Company

Lee Gilmer, Texaco Research

Marvin Hertz, Shell Development Company

Gilbert Jersey, Mobil Research and Development Company

George Lauer, ARCO

Robert Peace, Unocal

WE ARE INDEBTED TO CAPTAIN MICHAEL MOSS, UNITED STATES AIR FORCE, FOR HIS CONSIDERABLE EFFORTS DURING THE DEVELOPMENT OF THIS PUBLICATION.

ABSTRACT

This volume of the final report provides documentation of some of the results of a two year project entitled Hazard Response Modeling Uncertainty (A Quantitative Method). Work that has been accomplished on the technical work tasks related to evaluating the performance of commonly used hazardous gas dispersion models is summarized.

Eight datasets are used in the evaluation. Those field experiments that involve the release of dense-gas clouds are Burro, Coyote, Desert Tortoise, Goldfish, Maplin Sands, and Thorney Island. Those field experiments that involve the release of passive clouds are Hanford (Kr⁸⁵ tracer studies) and Prairie Grass. Data from these experiments are placed in a common format as a Modelers' Data Archive (MDA), and an extensive set of software was developed to prepare data-files for each model evaluated.

Fourteen dispersion models are evaluated, including six publicly-available computer models (AFTOX, DEGADIS, HEGADAS, INPUFF, OB/DG, and SLAB) and six proprietary computer models (AIRTOX, CHARM, FOCUS, GASTAR, PHAST, and TRACE). A simple Gaussian plume formula and a set of nomograms (Britter and McQuaid) are also evaluated for comparative purposes.

The statistical evaluation indicates that there are a few models that can successfully predict concentrations with a mean bias of 20 percent or less, a relative mean square error of 50 percent or less, and little variability of the residual errors with the input parameters. These models are identified in Section VII. It is also clear that model performance is not dependent upon model complexity.

It is necessary to point out that this evaluation exercise has been by no means independent, since all of the models have been previously tested by the developers with at least one of the datasets. Furthermore, some of the results may be fortuitous, since, in a few cases, certain models have been applied to source scenarios for which they were not originally intended.

TABLE OF CONTENTS

<u>Section</u>	<u>Title</u>	<u>Page</u>
	EXECUTIVE SUMMARY	ES-1
I.	INTRODUCTION	1
A.	OBJECTIVES	1
B.	BACKGROUND	3
	1. EPA Model Evaluation Program	3
	2. Model Sensitivity Studies	5
	3. Summary of Field Data	5
	4. A Methodology for Evaluating Heavy Gas Dispersion Models	6
	5. Comprehensive Model Evaluation Studies	7
	6. CMA Model Evaluation Studies	8
C.	SCOPE	8
II.	DATASETS	11
A.	CRITERIA FOR CHOOSING DATASETS	11
B.	DESCRIPTION OF INDIVIDUAL STUDIES	14
	1. Burro and Coyote	14
	2. Desert Tortoise and Goldfish	16
	3. Hanford Kr ⁸⁵	22
	4. Maplin Sands	25
	5. Prairie Grass	28
	6. Thorney Island	32
C.	CREATION OF A MODELERS' DATA ARCHIVE	36
D.	METHODS FOR CALCULATING REQUIRED VARIABLES	39
	1. Burro	40
	2. Coyote	41
	3. Desert Tortoise	41
	4. Goldfish	42
	5. Hanford Kr ⁸⁵	43
	6. Maplin Sands	46
	7. Prairie Grass	47
	8. Thorney Island	48
E.	SUMMARY OF DATASETS	49
III.	MODELS	53
A.	CRITERIA FOR CHOOSING MODELS	53
B.	DESCRIPTION OF MODELS EVALUATED	57
	1. AFTOX 3.1 (Air Force Toxic Chemical Model	58
	2. AIRTOX	60
	3. Britter & McQuaid Model	62
	4. CHARM 6.1 (Complex and Hazardous Air Release Model	65
	5. DEGADIS 2.1 (Dense Gas Dispersion Model)	67

III. B. (CONTINUED)

6.	FOCUS 2.1	69
7.	GASTAR 2.2 (Gaseous Transport from Accidental Releases)	71
8.	GPM (Gaussian Plume Model)	73
9.	HEGADAS (NTIS)	73
10.	INPUFF 2.3	75
11.	OB/DG (Ocean Breeze/Dry Gulch)	76
12.	PHAST 2.01 (Process Hazard Analysis Software Tool)	76
13.	SLAB (February, 1990)	77
14.	TRACE III	79
C.	APPLICATION OF MODELS TO DATASETS	81
1.	MDA Interface	81
2.	Initializing Individual Models	89
D.	SUMMARY OF MODELS	114
IV.	STATISTICAL MODEL EVALUATION	119
A.	PERFORMANCE MEASURES AND CONFIDENCE LIMITS	119
B.	RESULTS OF EVALUATION	121
1.	Evaluation of Concentration Predictions	125
2.	Cloud Widths (σ_y)	141
V.	SCIENTIFIC EVALUATION BY MEANS OF RESIDUAL PLOTS	145
A.	PROCEDURES	145
B.	RESULTS	145
VI.	SENSITIVITY ANALYSIS USING MONTE CARLO PROCEDURES	163
A.	OVERVIEW	163
B.	CHOICE OF MODELS AND INPUT PARAMETERS	163
C.	IMPLEMENTATION	165
D.	RESULTS	166
VII.	SUMMARY OF EVALUATION	175
A.	CONCENTRATION PREDICTIONS	175
B.	WIDTHS	177
C.	SCREENING MODEL RECOMMENDATIONS	179
REFERENCES		181

APPENDICES

<u>Appendix</u>	<u>Title</u>	<u>Page</u>
A	MODELERS' DATA ARCHIVE FILES	187
B	MODELAPPLICATION INFORMATION	205
C	TABULATION OF THE OBSERVED AND PREDICTED CLOUD-WIDTHS AND CONCENTRATIONS	293
D	TABULATION OF THE PERFORMANCE MEASURES AND THE RESULTS OF CONFIDENCE LIMITS ANALYSIS . .	315

**Documents**

Document Sharing Hub

LIST OF FIGURES

<u>Figure</u>	<u>Title</u>	<u>Page</u>
1.	Instrumentation Array for the 1980 LNG Dispersion Tests at NWC, China Lake (Reference 14)	15
2.	Instrumentation Array for the Coyote Series at NWC, China Lake. "Old" Locations Mark those used in the Burro Series (Reference 15)	17
3.	Sensor Array for the Desert Tortoise Series Experiments	20
4.	Configuration of Meteorological Towers and Kr ⁸⁵ Detectors for the Hanford Kr ⁸⁵ Trials (Reference 18)	24
5.	Initial Configuration of the Maplin Sands Site	27
6.	Revised Configuration of the Maplin Sands Site (After Trail 35)	27
7.	Configuration of Instrumentation Used for the Phase I Trials at Thorney Island (Reference 25) . .	34
8.	Correlation for continuous releases from Britter and McQuaid (Reference 42)	64
9.	Correlation for instantaneous release from Britter and McQuaid (Reference 42)	64
10.	Model performance measures, Geometric Mean Bias $MG = \exp(\ln C_o - \ln C_p)$ and geometric variance $VG = \exp[(\ln C_o - \ln C_p)^2]$ for concentration predictions and observations for the continuous dense gas group of datasets (Burro, Coyote, Desert Tortoise, Goldfish, Maplin Sands, Thorney Island). Ninety-five percent confidence intervals on MG are indicated. The solid line is the "minimum VG" curve, from Equation (33). The dashed lines represent "factor of two" agreement between mean predictions and observations	127
11.	Model performance measures, Geometric Mean Bias $MG = \exp(\ln C_o - \ln C_p)$ and geometric variance $VG = \exp[(\ln C_o - \ln C_p)^2]$ for concentration predictions and observations for the instantaneous dense gas data from Thorney Island. Ninety-five percent confidence intervals on MG are indicated. The solid line is the "minimum VG" curve, from Equation (33). The dashed lines represent "factor of	



Documents

Document Sharing Hub

two" agreement between mean predictions and observations 130

12. Model performance measures, Geometric Mean

Bias MG = $\exp(\overline{\ln C_o - \ln C_p})$ and geometric variance

VG = $\exp[(\overline{\ln C_o - \ln C_p})^2]$ for concentration predictions and observations at distances greater than or equal to 200 m. a) Continuous dense gas group of datasets (Burro, Coyote, Desert Tortoise, Goldfish, Maplin Sands, Thorney Island). b) Instantaneous dense gas data from Thorney Island. Ninety-five percent confidence intervals on MG are indicated. The solid line is the "minimum VG" curve, from Equation (33). The dashed lines represent "factor of two" agreement between mean predictions and

observations 132

13. Model performance measures, Geometric Mean

Bias MG = $\exp(\overline{\ln C_o - \ln C_p})$ and geometric variance

VG = $\exp[(\overline{\ln C_o - \ln C_p})^2]$ for concentration predictions and observations at distances less than 200 m. a) Continuous dense gas group of datasets (Burro, Coyote, Desert Tortoise, Goldfish, Maplin Sands, Thorney Island). b) Instantaneous dense gas data from Thorney Island. Ninety-five percent confidence intervals on MG are indicated. The solid line is the "minimum VG" curve, from Equation (33). The dashed lines represent "factor of two" agreement between mean predictions and observations 135

14. Model performance measures, Geometric Mean

Bias MG = $\exp(\overline{\ln C_o - \ln C_p})$ and geometric variance

VG = $\exp[(\overline{\ln C_o - \ln C_p})^2]$ for concentration predictions and observations. a) Continuous passive gas group of datasets (Prairie Grass and Hanford-continuous). b) Instantaneous passive gas dataset from Hanford. Ninety-five percent confidence intervals on MG are indicated. The solid line is the "minimum VG" curve, from Equation (33). The dashed lines represent "factor of two" agreement between mean predictions and observations

15. Model performance measures, Geometric Mean

Bias MG = $\exp(\overline{\ln C_o - \ln C_p})$ and geometric variance

VG = $\exp[(\overline{\ln C_o - \ln C_p})^2]$ for concentration

- predictions and observations at distances greater than or equal to 200 m for STABLE (class E, F) conditions a) Continuous dense gas group of datasets (Burro, Coyote, Desert Tortoise, Goldfish, Maplin Sands, Thorney Island). b) Instantaneous dense gas data from Thorney Island. Ninety-five percent confidence intervals on MG are indicated. The solid line is the "minimum VG" curve, from Equation (33). The dashed lines represent "factor of two" agreement between mean predictions and observations 140
16. Model performance measures, Geometric Mean
- Bias MG = $\exp(\ln C_o - \ln C_p)$ and geometric variance VG = $\exp[(\ln C_o - \ln C_p)^2]$ for plume width predictions and observations. a) Continuous dense gas group of datasets (Burro, Coyote, Desert Tortoise, Goldfish, Maplin Sands, Thorney Island). b) Continuous passive gas group of datasets (Hanford, Prairie Grass). Ninety-five percent confidence intervals on MG are indicated. There is no significant difference in part a) between the MG and VG values for the three better models (GPM, AIRTOX, and SLAB). The solid line is the "minimum VG" curve, from Equation (33). The dashed lines represent "factor of two" agreement between mean predictions and observations 142
17. Distributions of C_p/C_o for the datasets containing continuous releases of dense-gas clouds (Burro⁽¹⁾, Coyote⁽²⁾, Desert Tortoise⁽³⁾, Goldfish⁽⁴⁾, Mapling Sands LNG⁽⁷⁾, Maplin Sands LPG⁽⁸⁾, Thorney Island⁽¹⁰⁾). The "box plots" indicate the 2nd, 16th, 40th, 84th, and 98th percentiles of the cumulative distribution function of the N points in the class 146
18. Distributions of C_p/C_o for Thorney Island instantaneous releases of dense-gas clouds. The "box plots" indicate the 2nd, 16th, 50th, 84th, and 98th percentiles of the cumulative distribution function of the N points in the class 152
19. Distributions of C_p/C_o for Prairie Grass continuous releases of passive-gas clouds 157
20. The probability density functions (pdf) of the concentrtrions and widths (sigma-y) at 200 and 800 m downwind based on 500 Monte Carlo simulations fo the SLAB model for the Desert Tortoise 3 experiment 172



LIST OF TABLES

<u>Table</u>	<u>Title</u>	<u>Page</u>
1.	LIST OF EXPERIMENTS THAT WERE CONSIDERED FOR THE MODEL EVALUATION DATA ARCHIVE	12
2.	SUMMARY OF THE BURRO AND COYOTE TRIALS	18
3.	SUMMARY OF DESERT TORTOISE AND GOLDFISH EXPERIMENTS	23
4.	SUMMARY OF HANFORD KRYPTON-85 TRACER RELEASES	26
5.	SUMMARY OF THE MAPLIN SANDS EXPERIMENT	29
6.	SUMMARY OF SELECTED METEOROLOGICAL DATA FROM PRAIRIE GRASS TRIALS	31
7.	SUMMARY DESCRIPTION OF PHASE I TRIALS AND CONTINUOUS RELEASE TRIALS	35
8.	LIST OF INFORMATION CONTAINED IN AN MDA FILE	37
9.	SUMMARY OF CHARACTERISTICS OF THE DATASETS	50
10.	ATTRIBUTES OF MODELS	115
11.	MATRIX OF MODELS AND DATASETS FOR WHICH LATERAL CLOUD WIDTHS ARE EVALUATED	123
12.	MATRIX OF MODELS AND DATASETS FOR WHICH MAXIMUM CONCENTRATIONS ARE EVALUATED	124
13.	SUMMARY OF SUMMARY OF FB AND NMSE RESULTS FOR PREDICTING PEAK CONCENTRATIONS, AT DISTANCES > 200M	139
14.	PROBLEM REVEALED BY RESIDUAL PLOTS FOR CONTINUOUS RELEASES OF DENSE-GAS CLOUDS	151
15.	PROBLEM REVEALED BY RESIDUAL PLOTS FOR INSTANTANEOUS DENSE-GAS CLOUDS (THORNEY ISLAND)	155
16.	PROBLEM REVEALED BY RESIDUAL PLOTS FOR CONTINUOUS PASSIVE-GAS CLOUDS (PRAIRIE GRASS)	161
17.	AN EXAMPLE OF THE OUTPUT FILE GENERATED BY THE MDAMC PACKAGE, WHERE 20 MONTE CARLO SIMULATIONS OF THE SLAB MODEL FOR THE DESERT TORTOISE 3 EXPERIMENT WERE PERFORMED	167
18.	MODEL UNCERTAINTIES FOR THE SLAB MODEL WHEN ALL SEVEN PRIMARY INPUT PARAMETERS (SEE TEXT) WERE PERTURBED SIMULTANEOUSLY IN 500 MONTE CARLO SIMULATIONS FOR THE DESERT TORTOISE 3 EXPERIMENT. (S.D.: STANDARD DEVIATION)	168

LIST OF TABLES
(CONTINUED)

<u>Table</u>	<u>Title</u>	<u>Page</u>
19.	MODEL UNCERTAINTIES FOR THE SLAB MODEL WHEN ONLY THE DOMAIN-AVERAGED WIND SPEED WAS PERTURBED IN 500 MONTE CARLO SIMULATIONS FOR THE DESERT TORTOISE 3 EXPERIMENT	168
20.	MODEL UNCERTAINTIES FOR THE SLAB MODEL WHEN ONLY THE DIFFERENCE IN WIND SPEED BETWEEN DOAMIN-AVERAGE AND A TOWER WAS PERTURBED IN 500 MONTE CARLO SIMULATIONS FOR THE DESERT TORTOISE 3 EXPERIMENT	169
21.	MODEL UNCERTAINTIES FOR THE SLAB MODEL WHEN ONLY THE DIFFERENCE IN TEMPERATURE BETWEEN TWO LEVELS ON A TOWER WAS PERTURBED IN 500 MONTE CARLO SIMULATIONS FOR THE DESERT TORTOISE 3 EXPERIMENT. (S.D.: STANDARD DEVIATION)	169
22.	MODEL UNCERTAINTIES FOR THE SLAB MODEL WHEN ONLY THE RELATIVE HUMIDITY WAS PERTURBED IN 500 MONTE CARLO SIMULATIONS FOR THE DESERT TORTOISE 3 EXPERIMENT. (S.D.: STANDARD DEVIATION)	170
23.	MODEL UNCERTAINTIES FOR THE SLAB MODEL WHEN ONLY THE SURFACE ROUGHNESS WAS PERTURBED IN 500 MONTE CARLO SIMULATIONS FOR THE DESERT TORTOISE 3 EXPERIMENT. (S.D.: STANDARD DEVIATION)	170
24.	MODEL UNCERTAINTIES FOR THE SLAB MODEL WHEN ONLY THE SOURCE EMISSION RATE WAS PERTURBED IN 500 MONTE CARLO SIMULATIONS FOR THE DESERT TORTOISE 3 EXPERIMENT. (S.D.: STANDARD DEVIATION)	171
25.	MODEL UNCERTAINTIES FOR THE SLAB MODEL WHEN ONLY THE SOURCE DIAMETER WAS PERTURBED IN 500 MONTE CARLO SIMULATIONS FOR THE DESERT TORTOISE 3 EXPERIMENT. (S.D.: STANDARD DEVIATION)	171
26.	RATIOS OF RELATIVE MODEL UNCERTAINTIES, $\sigma_c/\bar{C}$, AND $\sigma_w/\bar{w}$, TO THE RELATIVE INPUT DATA UNCERTAINTIES, $\sigma_i/\bar{i}$, FOR THE SLAB MODEL WHEN THE SEVEN PRIMARY INPUT PARAMETERS WERE PERTURBED ONE AT A TIME IN 500 MONTE CARLO SIMULATIONS FOR THE DESERT TORTOISE 3 EXPERIMENT. (S.D.: STANDARD DEVIATION)	173



Documents

Document Sharing Hub

LIST OF TABLES
(CONTINUED)

<u>Table</u>	<u>Title</u>	<u>Page</u>
27.	RANKING OF MODELS ACCORDING TO FAC2 (FACTOR OF TWO) STATISTIC, WHICH EQUALS THE FRACTION OF TIME THAT THE PREDICTIONS ARE WITHIN A FACTOR OF TWO OF THE OBSERVATIONS	176
28.	SUMMARY OF PERFORMANCE EVALUATION BASED ON GEOMETRIC MEAN BIAS MG AND GEOMETRIC VARIANCE VG FOR CONCENTRATIONS, NEGLECTING INSTANTANEOUS PASSIVE DATASET. THE TERMS "OVER" AND "UNDER" REFER TO THE BIAS IN THE MEAN PREDICTIONS	178



EXECUTIVE SUMMARY

A. OBJECTIVE

The overall objective of this project is to develop and test computer software containing a quantitative method for estimating the uncertainty in PC-based hazard response models. This software is to be used by planners and engineers in order to evaluate the predictions of hazard response models with field observations and determine the confidence intervals on these predictions. This particular volume (II) provides an example of the application of the software to 14 typical hazard response models and 8 sets of field data:

B. BACKGROUND

The U.S. Air Force and the American Petroleum Institute, among others, have increased emphasis on calculating toxic corridors due to releases of hazardous chemicals into the air. There are dozens of PC-based computer models recently developed in order to calculate these toxic corridors. However, the uncertainties in these models have not been adequately determined, partly due to the lack of a standardized quantitative method that could be applied to these models. Individual model developers generally present a limited evaluation of their own model, and the USEPA has published some partial evaluations, but a comprehensive study has not been completed.

C. SCOPE

The scope of the overall project has included acquisition and testing of databases and models, development and application of model evaluation software, and assessment of the components of uncertainty. The current volume (II) emphasizes an example application of the model to a reasonably comprehensive set of 14 hazard response models and 8 independent field experiments. Both proprietary and publicly-available models are considered, and the field data cover a wide variety of source scenarios and thermodynamic behavior.

D. METHODOLOGY

The statistical performance measures are tabulated and discussed for six publicly-available computer models (AFTOX, DEGADIS, HEGADAS, INPUFF, OB/DG, and SLAB) and six proprietary computer models (AIRTOX, CHARM, FOCUS, GASTAR, PHAST, and TRACE). In addition, results are presented for two simple analytical models--the Gaussian plume model (GPM) and the Britter and McQuaid model (B&M). These models were applied to data from eight field tests, where the source scenarios include continuous dense gas releases (Burro, Coyote, Desert Tortoise, Goldfish, Maplin Sands, and Thorney Island-C), instantaneous dense gas releases (Thorney Island-I), continuous passive gas releases (Prairie Grass and Hanford-C), and instantaneous passive gas releases (Hanford-I).

The report contains discussions of the following major topics:

- **Creation of Modelers Data Archive (MDA)**--Each field experiment is described in detail and the data from all experiments are combined in a consistent Modelers Data Archive (MDA) that can be used to initialize and evaluate all of the models. The MDA is listed in an Appendix to Volume II, and a floppy disk containing the MDA is available to all interested persons.
- **Application of Models to MDA**--The 14 models are reviewed and methods of applying them to the MDA are discussed. In many cases, preprocessor and postprocessor software had to be written so that all 14 models could begin from the same set of input data and could produce consistent output data.
- **Statistical Model Evaluation**--The model performance measures (mean bias, mean square error, correlation coefficient, fraction within a factor of two) and their confidence limits are calculated for each model and each data group and are presented in tables and figures. The primary mode of graphical presentation is a figure with mean square error on the vertical axis and mean bias on the horizontal axis, on which points are plotted for each model. Summary tables are provided.

- Residual Plots-- Many figures are given, in which ratios of prediction to observation are plotted versus input parameter (for example, wind speed or stability) for each model. Conclusions are given in summary tables.
- Sensitivity Study--The Monte Carlo sensitivity software is used to determine the sensitivity of the SLAB model to variations in input parameters.

E. CONCLUSIONS

- A few models can successfully predict concentrations with a mean bias of 20 percent or less, a relative scatter of 50 percent or less, and little variability of the residual errors with input parameters.
- The four models (BM, GPM, SLAB, and HEGADAS) that produce the best "Factor of Two" agreement are on the list of six models (BM, GPM, SLAB, HEGADAS, CHARM, and PHAST) that produce the most consistent performance for the statistics describing the mean bias and the variance.
- The performance of any model is not related to its cost or complexity.
- In two of the three data groups, the "best" model is one which was not originally developed for that scenario (that is, GPM for continuous dense gas releases and SLAB for continuous passive gas releases).
- The BM, GPM, SLAB, and HEGADAS models demonstrate the most consistent performance for the "fraction within a factor of two" (FAC2) statistic.
- The results of the analyses in this section lead to the recommendation that the following simple, analytical formulas can be confidently used for screening purposes for sources over flat, open terrain:

BM (Britter and McQuaid) for continuous and instantaneous dense gas releases.

GPM (Gaussian Plume Model) for continuous passive gas releases.

There are insufficient field data to justify recommendations for instantaneous passive gas releases. However, the EPA's INPUFF model appears to perform reasonably well for the Hanford dataset in Figure 14b.

These screening models would not be appropriate for source scenarios and terrain types outside of those used in the model derivations. For example, because the screening models neglect variations in roughness length, they would be inappropriate for urban areas or heavily industrialized areas.

F. RECOMMENDATIONS

This evaluation exercise has been by no means independent, since all of the models have been previously tested by the developers with at least one of the datasets. Furthermore, some of the results may be fortuitous, since, in a few cases, certain models have been applied to source scenarios for which they were not originally intended.

In the future, our model evaluation software should be used to evaluate models with new independent datasets. An attempt should be made to set up standards for models so that they all conform to certain scenarios and to certain input and output data requirements.

SECTION I
INTRODUCTION

A. OBJECTIVES

This is Volume II of a three volume set describing the results of a project in which a quantitative method has been developed to determine the uncertainties in hazardous gas models. The first volume discusses the user's guide for this model evaluation method and the third volume discusses the three components of model uncertainty--data input errors, stochastic fluctuations, and model physics errors. The current volume provides an example of the application of the procedures.

The Phase II research has had the following eight technical objectives or tasks. The volume of the final report that deals with each of the following tasks is listed in parentheses at the end of the paragraphs.

Task 1: Archival of Data Sets and Preparation of Modelers Data Bases. A computerized archive of field data sets has been prepared. This archive includes a broad range of source conditions, meteorological conditions, and averaging times. The information in the data base is sufficient to run any of the models. (Volume II)

Task 2: Archival of Hazard Response Models, including Testing. A comprehensive archive of available microcomputer-based hazard response models has been prepared. This includes recently developed or modified publicly-available models such as SLAB and DEGADIS, as well as proprietary models that are in common use. (Volume II)

Task 3: Application of Models to Test Data. Predictions from the models obtained under Task 2 were produced for the field tests obtained under Task 1. In some cases it was necessary to make additional calculations so that the input data are in the form acceptable by the model, or so that the model output data are in the form required by the model evaluation software. (Volume II)

Task 4: Further Development of Model Evaluation Software. The statistical model evaluation software has been refined and further developed so that it is sufficiently general to take a wide variety of input data sets and calculate a complete set of possible performance measures. It is possible to calculate confidence intervals (that is, model uncertainties) from this procedure. (Volume I)

Task 5: Application of Model Evaluation Software. The model evaluation software was applied to the model predictions and data sets in our archive. Estimates of typical confidence limits for certain classes of models and sizes of data set were made. (Volume II)

Task 6: Assessment of Data Uncertainties. The contribution of data uncertainties to total model error were estimated. Part of this research involves investigation of Air Force meteorological instrumentation and quality control/quality assurance procedures, as well as field tests by NCAR scientists of instrument accuracy and representativeness. (Volume III)

Task 7: Assessment of Stochastic Uncertainties. The contribution of stochastic or random uncertainties to total model error was further studied, and a quantitative procedure was developed for estimating this component as a function of receptor position, source type, sampling and averaging time, and meteorological conditions. The effect of these fluctuations on relations for toxic response were studied. (Volume III)

Task 8: Assessment of Model Physics Errors. Dimensional analysis and various reduction procedures were applied to the complete archive of data sets and models in order to isolate the contribution of errors in model physics assumptions to the total model uncertainty. (Volume III)

B. BACKGROUND

The U.S. Air Force and the American Petroleum Institute, among others, have increased emphasis on calculating "toxic corridors" due to potential release of hazardous chemicals. The Ocean Breeze/Dry Gulch (OB/DG) model was originally used for calculating these corridors, and does contain an estimate of model uncertainty. However, the OB/DG model does not account for many important scientific phenomena, such as two-phase jets, evaporative emissions, and dense gas slumping. The new models mentioned above are more advanced scientifically, but do not include model uncertainty. The intent of this research is to fully develop quantitative model evaluation procedures, better estimate the components of the uncertainty (data input errors, stochastic uncertainties, and model physics errors), and test the procedures using a wide spectrum of field and laboratory experiments.

Several evaluations of dispersion models applicable to the release of toxic material to the atmosphere were reviewed in the Phase I report for this project. We repeat reviews of the more recent studies, and include an overview of a recent evaluation program sponsored by EPA.

1. EPA Model Evaluation Program

The EPA has been sponsoring a related dense gas model evaluation project being performed by TRC Environmental Consultants. We have exchanged ideas and information with the EPA scientists, and have reviewed a preliminary draft copy of their final report (Reference 1). The purpose of this section is to briefly compare the methods and results of the two studies.

The two studies are evaluating the models in the list below:

	EPA	USAF/API
Publicly Available	SLAB DEGADIS	SLAB DEGADIS GAUSSIAN PLUME MODEL INPUFF AFTOX HEGADAS OB/DG Britter & McQuaid

	EPA	USAF/API
Proprietary	AIRTOX	AIRTOX
	CHARM	CHARM
	TRACE	TRACE
	FOCUS	FOCUS
	SAFEMODE	PHAST
		GASTAR

It is seen that the USAF/API study includes six more publicly-available models and one more proprietary model.

The following field data sets are used:

	EPA	USAF/API
Dense Gas	Burro	Burro
	Desert Tortoise	Desert Tortoise
	Goldfish	Goldfish
		Coyote
		Maplin Sands
		Thorney Island

	EPA	USAF/API
Passive Gas		Prairie Grass
		Hanford Kr85

The EPA study was deliberately restricted to data sets in which dense gases were continuously released for periods of three to ten minutes. The total numbers of individual field tests in the EPA and USAF/API studies are 9 and 118, respectively.

The EPA contractor permitted the model developers to advise them on how to run the models (for example, definitions of input conditions and choices of model options), whereas the models were run in a more independent manner in the USAF/API study. The developers were asked to comment on the way their models were set up in the USAF/API study, but the final decision was made by us.

The model performance measures used in the two studies are similar. Both considered maximum concentrations and plume widths on monitoring arcs. In any given field test, there were about two to seven monitoring arcs.

The results of the EPA study were inconclusive. The TRACE, CHARM, DEGADIS, and SLAB model performances were not significantly different, and "none demonstrated good performance consistently for all three experimental programs". In contrast, as will be shown below, the USAF/API results were more conclusive, perhaps because of the much larger set of data.

2. Model Sensitivity Studies

During 1986 and 1987, Professor Carney of Florida State University prepared several papers for the AFESC on the sensitivity of the AFTOX, CHARM, and PUFF models to uncertainties in input data (Reference 2). His 1987 paper applied the uncertainty formula suggested by Freeman et al. (Reference 3), which has also been applied by Hanna (Reference 4) to a simplified air quality model. If concentration, C , is an analytical function of the variables x_i ($i = 1$ to n), then the uncertainty or variance $V_C = \sigma_C^2$ is given by the equation

$$V_C = \sum_{i=1}^n \left(\partial C / \partial x_i \right)^2 V_{xi} + \sum_{i=1}^n \sum_{j=1}^n \left(\partial^2 C / \partial x_i \partial x_j \right)^2 V_{xi} V_{xj} \quad (1)$$

$$+ 0.5 \sum_{i=1}^n \left(\partial^2 C / \partial x_i^2 \right) V_{xi}^2$$

where V_{xi} is the uncertainty or variance in input variable x_i . This equation is a Taylor expansion and implicitly assumes that the individual uncertainties are much less than one. Carney (Reference 2) finds that the wind speed, u , contributes the most uncertainty to the concentration, C , predicted by the AFTOX model.

3. Summary of Field Data

Ermak et al. (Reference 5) has put together a comprehensive summary of 26 "bench mark" field experiments, including data from Burro (LNG), Coyote (LNG), Eagle (N_2O_4), Desert Tortoise (NH_3), Maplin Sands (LNG and LPG) and Thorney Island (Freon). This study (funded by AFESC) presents input data

required by models and includes observed peak concentrations, average centerline concentrations, and average height and width of the cloud as a function of downwind distance. These data are sufficiently complete for anyone to run and evaluate his model.

4. A Methodology for Evaluating Heavy Gas Dispersion Models

In another recent draft report prepared for AFESC, Ermak and Merry (Reference 6) review methods for evaluating heavy gas dispersion models. They first list several specific criteria of interest to the Air Force:

- The methodology is to be based on comparison of model predictions with field-scale experimental observations.
- The methods of comparison must be quantitative and statistical in nature.
- The methods must help identify limitations of the models and levels of confidence.
- The methodology must be compatible with atmospheric dispersion models of interest to the Air Force.

These criteria are similar to those for our present study.

The Ermak and Merry (Reference 6) report is a review of general evaluation methods and heavy gas model data sets, and does not contain examples of applications of any new evaluation methods with field data sets. They first review the general philosophy of model evaluation, pointing out that sometimes evaluations of model physics are just as important as quantitative statistical evaluations. Much of their philosophical discussion follows the points made in a review paper by Venkatram (Reference 7). For example, a model whose predictions agree with field data but which contains an irrational physical assumption (for example, dense gas plumes accelerate upward) is not a good model. Also, they recognize that most model predictions represent ensemble averages, whereas field experiments represent only a single realization of the countless data that make up an

ensemble. They emphasize that observed concentrations are strong functions of averaging time, and that most heavy gas dispersion models do not include the effects of averaging time.

Heavy gas dispersion models are distinguished from other dispersion models by three effects: reduced turbulent mixing, gravity spreading, and lingering. The main parameters of interest in evaluations of these models are the maximum concentration, the average concentration over the cloud, and the cloud width and height (all as a function of downwind distance, x). Ermak and Merry emphasize the ratio of predicted to observed variables and define several statistics, such as the mean and the variance. Methods of estimating confidence limits on these statistics are suggested, and the report closes with an example of the application of some of their suggested procedures to a concocted data set drawn from a Gaussian distribution.

5. Comprehensive Model Evaluation Studies

Mercer's (Reference 8) review emphasizes estimation of variability or uncertainty in model predictions, which he finds is typically an order of magnitude when outliers are considered. He includes the following quote from Lamb (Reference 9), which is also appropriate for our discussion.

"The predictions even of a perfect model cannot be expected to agree with observations at all locations. Consequently, the common goal of model validation should be one of determining whether observed concentrations fall within the interval indicated by the model with the frequency indicated, and if not, whether the failure is attributable to sampling fluctuations or is due to the failure of the hypotheses on which the model is based. From the standpoint of regulatory needs the utility of a model is measured partly by the width of the interval in which a majority of observations can be expected to fall. If the width of the interval is very large, the model may provide no more information than one could gather simply by guessing the expected concentration. In particular, when the width of the interval of probable concentration values exceeds the allowable error bounds on the model's predictions, the model is of no value in that particular application."

Mercer (Reference 8) then produces concentration predictions of ten different models for a dense gas source equivalent to that used in the Thorney Island experiments. This comparison shows that the 10 model predictions range over an order of magnitude at any given downwind distance.

6. CMA Model Evaluation Program

The Chemical Manufacturers' Association (CMA) sponsored an evaluation of eight dense gas dispersion models and nine spill evaporation models (References 10 and 11). The authors ran some of the models themselves and requested the developers of proprietary models to run their own models using standard input data sets. Model uncertainty is typically a factor of two to five. The comparisons are clouded by the use of some data sets that had already been used to "tune" certain of the models tested.

C. SCOPE

This introductory section has provided an overview of the objectives of the entire project, which was initiated because there are no standard objective quantitative means of evaluating microcomputer-based hazard response models. There are dozens of such models including several sponsored wholly or in part by the U.S. Air Force and the American Petroleum Institute: ADAM, AFTOX, CHARM, DEGADIS, SLAB, and OB/DG. A few data sets exist for testing these models, but, up until now, the models have not been tested or intercompared with these data on the basis of standard statistical significance tests. The U.S. EPA recently sponsored a related model evaluation project (Reference 1), which had a more limited scope and considered fewer models and datasets.

In this volume, we focus on a demonstration of the system to evaluate the performance of micro-computer-based dispersion models that are applicable to releases of toxic chemicals into the atmosphere. The study includes a total of 14 models and 8 datasets. The datasets are described in Section II, and the models are described in Section III. Results of the statistical evaluations are presented in Section IV, and a scientific evaluation of the distribution of residuals is presented in Section V. One example of how Monte Carlo procedures described in Volume I can be used to investigate the

sensitivity of a model to uncertainties in the input data is discussed in Section VI. The overall results are presented in Section VII.

When reading about the evaluations presented in Section IV and V, it is important to remember that, in many cases, there can be more than one way to apply a given model to a given dataset. Our approach has been to retain a fair degree of "independence" from the developers of the models being tested. We assembled/developed the data required as input to the models, assembled/developed the data against which the models are compared, applied the models, and then requested comments on our approach from the developers of the models. We supplied each developer with a description of the datasets and the procedure used to apply the developer's model to each dataset. We also provided a list of the concentrations obtained from the model, and those concentrations against which the modeled concentrations are compared, but we did not provide any indication of model performance relative to other models used in the study. Comments solicited in this way resulted in changes to our evaluation only if errors in the application were discovered. In this way, we were able to maintain a uniform approach to all of the models, and we consider the results indicative of what would be obtained by modelers "in the field."

This approach did not, however, preclude earlier discussions with the model developers. Upon reading the user's manuals, clarifications were sometimes needed, and these were addressed by means of telephone conversations and/or letters. Some of the models in the study underwent revisions during the study, so that some interaction focused on implementing new versions of the models. Such new versions sometimes contained bugs that became obvious as we began to use them, and this information was immediately passed on to the developer, and generally resulted in a revision. We emphasize, however, that none of these interactions focused on model performance issues arising from work performed during this study. Section III B characterizes the nature of our interactions with each of the model developers.

SECTION II

DATASETS

A. CRITERIA FOR CHOOSING DATASETS

The hazard response models included in this study (see Section 3) possess widely varying capabilities, but the majority do have several traits that influence the choice of datasets for evaluating this group of models. Chief among these is a preference for treating near-surface releases. As a result, we have not included datasets in which an elevated (say, more than a meter or two above the surface) source is used. Beyond this restriction, our criteria for selecting the datasets include:

1. Concurrent meteorological data must be available, obtained from sensors located near the site of the trials.
2. Concentrations should be available at more than one distance downwind, with sufficient lateral resolution to document the spatial structure of the cloud.
3. Temporal resolution of the concentration measurements should be less than the smaller of the duration of the release or the time-of-travel from the point of release to the nearest monitor.
4. Datasets chosen should document dispersion over a wide range of meteorological dispersion regimes.
5. Datasets chosen should include passive or "tracer" gas releases as well as dense-gas releases.
6. Datasets chosen should include instantaneous releases and continuous releases.

Many field experiments have been conducted for the purpose of evaluating dispersion models. Draxler (Reference 12) reviews many carried out with positively or neutrally buoyant sources. Hanna and Drivas (Reference 13) review many carried out with negatively buoyant sources. A total of 16 datasets derived from these reviews were considered for inclusion in this

TABLE 1. LIST OF EXPERIMENTS THAT WERE CONSIDERED FOR THE MODEL EVALUATION DATA ARCHIVE.

Name	Material Released	Dense Gas	Type of Release	
			Quasi-Continuous	Instantaneous
Burro	LNG	✓	✓	
Coyote	LNG, LCH ₄	✓	✓	
Desert Tortoise	NH ₃	✓	✓	
*Eagle	N ₂ O ₄	✓	✓	
*Falcon	LNG	✓	✓	
Goldfish	HF	✓	✓	
*Porton Down	Freon-12	✓		✓
Thorney Island	Freon-12(N ₂)	✓	✓	✓
Maplin Sands	LNG, LP	✓	✓	✓
Prairie Grass	SO ₂		✓	
*Dry Gulch	FP		✓	
*Ocean Breeze	FP		✓	
*Green Glow	FP		✓	
Hanford Kr ⁸⁵	Kr ⁸⁵		✓	✓
*Sandstorm	Be		✓	
*Adobe	Be		✓	

* Not included in the modeling data archive

project, and are listed in Table 1. Nine involve releases of denser-than-air gases, while seven involve the release of gases or suspended particles in amounts small enough to act as passive tracers.

Based on a review of the data and the documentation for these 16 experiments, a decision was made not to consider seven of them. Neither the ADOBE nor Sandstorm experiments were included in the study, since they were concerned with the transport and diffusion of buoyant exhaust clouds from rocket motors. Few of the models tested in this project can accommodate a buoyant cloud, and furthermore, there are not sufficient data on the exhaust characteristics of the rocket motors in the data reports to adequately define the temperature and volume flux of the jet. Data from the Falcon Experiments were excluded from the study for two reasons: only one of the trials was successful from the point of view of evaluating diffusion models, and a data report is not available. The Eagle tests were also excluded, since some of the tests involved the use of a barrier to the flow, which sets them apart from the remainder of the datasets used in the study, and there were instrument problems with the remaining tests.

Of the remaining 11 experiments, 5 are tracer experiments (that is, the chemical that is released behaves as an inert or passive non-buoyant substance as it disperses downwind). The Prairie Grass experiment provides high quality dispersion data over a wide range of turbulence regimes at an ideal site. The Dry Gulch, Ocean Breeze, and Green Glow data are not included because they are similar to the Prairie Grass data, yet cover a more limited range of stabilities. The Kr⁸⁵ tracer experiment conducted in Hanford, WA is included because it provides good data for puff releases as well as quasi-continuous releases of neutral-density or passive gases.

One of the remaining dense-gas dispersion datasets was recently dropped from consideration as well. The Porton Down dataset includes 42 trials in which mixtures of Freon-12 and air were released in the form of an instantaneous cloud. Those trials include variations in initial cloud density, wind speed, and surface roughness, but they lack an extensive array of monitors capable of providing continuous concentration measurements. The primary monitors provided only dosage measurements. These dosages can be used to estimate a mean concentration during the time over which the cloud passed through the monitoring array, but we found that these estimates contribute little to the goal of quantifying model performance. We expected

that the models would tend to produce estimates of peak concentration which would exceed the average concentrations estimated from the dosages--and all of the models did. No additional information could be obtained from the dataset. As a result, we have excluded the Porton Down trials from any further discussion in this report.

Hence, the performance evaluations are based on a total of 8 datasets. In the remainder of this section, we provide: a description of each of the field studies (Section II B); a description of the MDA containing data from each dataset (Section II C); a summary of the methods used to calculate information required by the MDA (Section II D); and an overall summary of the datasets (Section II E).

B. DESCRIPTION OF INDIVIDUAL FIELD STUDIES

1. Burro and Coyote

Both the Burro (Reference 14) and Coyote (Reference 15) series of trials were conducted at the Naval Weapons Center (NWC) at China Lake, California. Sponsored by the U.S. Dept. of Energy and the Gas Research Institute, the trials consisted of releases of LNG onto the surface of a 1 m deep pool of water, 58 m in diameter. In addition, the Coyote series expanded on the earlier Burro trials by studying the occurrence of rapid-phase-transitions (RPT), and included releases of liquefied methane and liquid nitrogen. The Burro series focused on the transport and diffusion of vapor from spills of LNG on water. The Coyote series focused on the characteristics of fires resulting from ignition of clouds from LNG spills, and the series also focused on the RPT explosions. In all, eight trials from the Burro series and four trials from the Coyote series are suitable for testing transport and diffusion models.

For the Burro series, twenty cup-and-vane anemometers were located at a height of 2 m at various positions within the test array in order to map the wind field. There were six 10 m tall turbulence stations, one upwind and five downwind, which had bivane anemometers at three levels and thermocouples at four levels. Humidities were measured close to the array centerline at eight stations, including the upwind turbulence station. Ground heat-flux sensors were mounted at seven downwind stations along with the humidity sensors. Figure 1 shows the configuration of the test site.

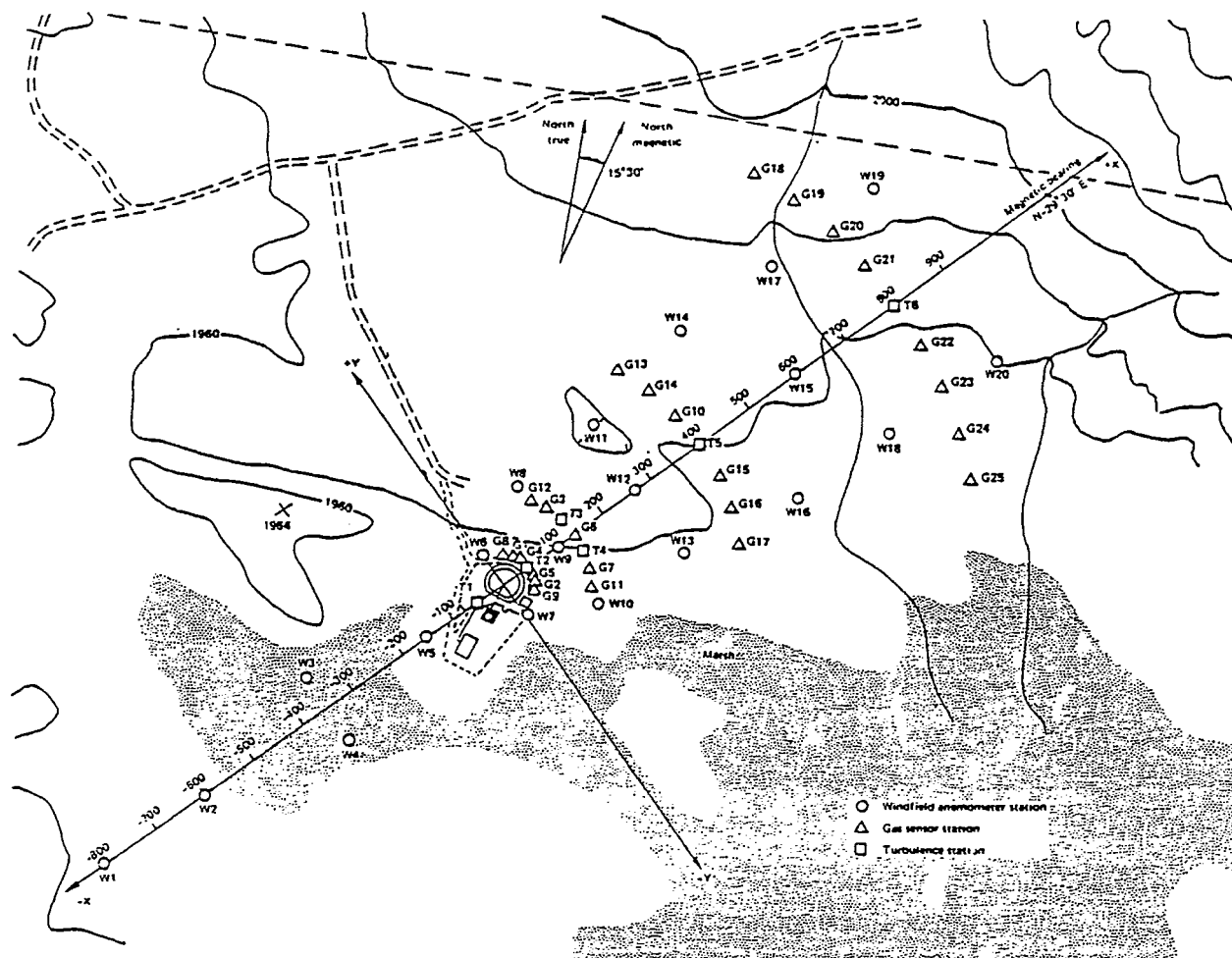


Figure 1. Instrumentation Array for the 1980 LNG Dispersion Tests at NWC, China Lake (Reference 14).

Concentrations were measured at heights of 1 m, 3 m, and 8 m at 25 gas-sampling stations and 5 turbulence stations arranged in arcs at distances of 57 m, 140 m, 400 m, and 800 m downwind from the spill point. The turbulence stations sample the data at a higher rate than the gas stations (3-5 Hz compared to 1 Hz). The lateral spacing between stations varied from 13 m at stations closer to the spill point, to 80 m at stations located 800 m downwind.

The Coyote series maintained a similar array of instrumentation. However, only two of the turbulence stations (one upwind, one at 300 m downwind of the spill site) were instrumented with bivane anemometers because of a concern that they might be damaged. Gas concentrations were measured at heights of 1 m, 3 m, and 8 m at 24 gas-sampling stations and 5 turbulence stations arranged in arcs at distances of 110 m, 140 m, 200 m, 300 m, 400 m and 500 m downwind from the spill point. Note that there were in fact only one and two gas sensors deployed at distances of 110 m and 500 m downwind, respectively. The lateral spacing between stations varied from 30 m at a distance of 140 m downwind to 60 m at a distance of 800 m downwind. Figure 2 shows the configuration of the test site.

Data from all eight Burro trials and three of the four Coyote trials are available on 9-track tape prepared by Lawrence Livermore National Laboratory (LLNL). Comparison data-reports (Burro, (Reference 14); Coyote, (Reference 15)) are also available, and proved very useful in preparing the data for use in the evaluations. The individual trials contained in these reports include

Burro: 2, 3, 4, 5, 6, 7, 8, 9

Coyote: 3, 5, 6

A brief summary of the characteristics of the source emissions and the meteorological conditions for the eight Burro trials and four Coyote trials is given in Table 2.

2. Desert Tortoise and Goldfish

These two series of field experiments were conducted at the Frenchman Flat area of the Nevada Test Site. The first in the series, Desert Tortoise (Reference 16) was designed to document the transport and

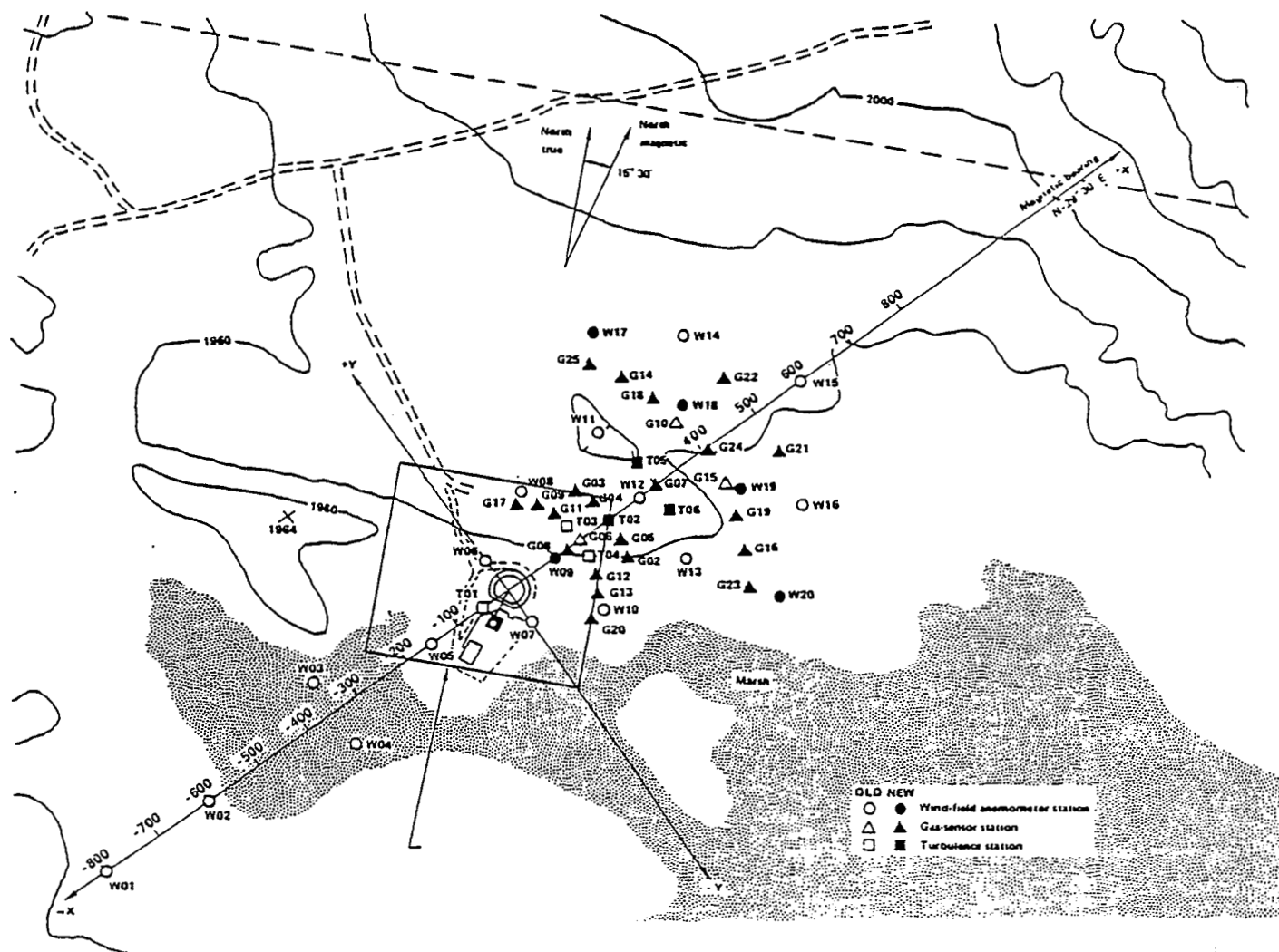


Figure 2. Instrumentation Array for the Coyote Series at NWC, China Lake.
 "Old" Locations Mark those used in the Burro Series (Reference 15).

TABLE 2. SUMMARY OF THE BURRO AND COYOTE TRIALS.

Test Name	Date	Material Spilled	Spill Volume (m ³)	Spill Rate (m ³ /min)	Averaged Wind Speed (m/s)	Averaged Wind Direction (degrees)	Atmospheric Stability Class
Burro 2	18 June	LNG	34.3	11.9	5.4	221	C
Burro 3	2 July	LNG	34.0	12.2	5.4	224	C
Burro 4	9 July	LNG	35.3	12.1	9.0	217	C
Burro 5	16 July	LNG	35.8	11.3	7.4	218	C
Burro 6	5 Aug.	LNG	27.5	12.8	9.1	220	C
Burro 7	27 Aug.	LNG	39.4	13.6	8.4	208	C/D
Burro 8	3 Sept.	LNG	28.4	16.0	1.8	235	E
Burro 9	17 Sept.	LNG	24.2	18.4	5.7	232	D
Coyote 3	2 Sept.	LNG	14.6	13.5	6.0	205	C
Coyote 5	7 Oct.	LNG	28.0	17.1	9.7	229	C
Coyote 6	27 Oct.	LNG	22.8	16.6	4.6	220	D

diffusion of ammonia vapor resulting from a cryogenic release of liquid ammonia. For each of the four trials, pressurized liquid NH_3 was released from a spill pipe pointing downwind at a height of about 1 m above the ground. The liquid jet flashed as it exited the pipe and its pressure decreased, resulting in about 18 percent of the liquid changing phase to become a gas. The remaining 82 percent of the NH_3 -jet remained as a liquid, which was broken up into an aerosol by the turbulence inside the jet. Very little, if any of the unflashed liquid was observed to form a pool on the ground. Dispersion of the vapor-aerosol cloud was dominated by the dynamics of the turbulent jet near the point of release, but the slumping and horizontal spreading of the cloud downwind of the jet zone indicated the dominance of dense-gas dynamics at later stages.

Figure 3 shows the configuration of instrumentation used during Desert Tortoise. Eleven cup-and-vane anemometers were located at a height of 2 m at various positions within the test array in order to define the wind field for the planning of field experiments and the subsequent calculation of plume trajectories. In addition, a 20 m tall meteorological tower was located just upwind of the spill area, with temperature measured at four levels and wind speed and turbulence at three levels. Ground heat fluxes were measured at that tower and at three locations just downwind of the spill.

NH_3 concentrations and temperatures were obtained at elevations of 1, 2.5, and 6 m on seven towers located along an arc at a distance of 100 m downwind of the source. In most cases, nearly all of the plume was below the 6 m level of the towers and within the lateral domain of the towers. Additional NH_3 concentration observations at elevations of 1, 3.5, and 8.5 m were taken on five monitoring towers at a distance of 800 m from the source, where the lateral spacing of the towers was 100 m. Finally there were two arcs with up to eight portable ground-level stations at distances of 1400 m or 2800 m, and on occasion at 5500 m downwind. No information on vertical distribution of NH_3 concentration was available from these more distance arcs.

The Goldfish trials are very similar to the Desert Tortoise NH_3 trials described above. Hydrogen fluoride (HF) was released using a similar release mechanism and some of the same sets of instruments. Note that

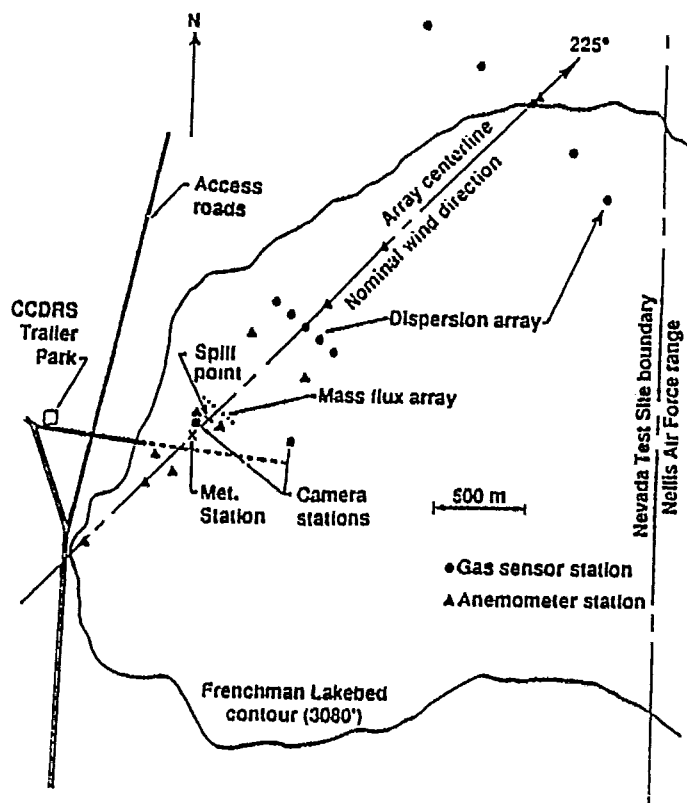


Figure 3. Sensor Array for the Desert Tortoise Series Experiments.

although six trials were conducted, the last three involved a study of the effectiveness of water sprays, and are not included in this evaluation demonstration. A portion of the liquid HF flashed upon release, creating a turbulent jet in which the unflashed liquid was broken up into an aerosol that remained in the jet-cloud. No pooling of the liquid was observed.

HF samplers were located on cross-wind lines at distance of 300, 1000, and 3000 m from the source. The closest line has 11 sampling locations, with instruments at heights of 1, 3, and 8 m at the inner 5 positions and instruments at a height of 1 m on the outer 6 positions. The 1000 m line has 13 sampling locations, with three levels of measurement on the inner 9 and only one level on the outer 4. The 3000 m line has 11 sampling locations, with a similar variation in sampler heights. In general the observed height of the HF cloud was less than the highest sampler level at the 300 m sampling line, but appeared to extend above the highest sample levels at the larger distances. The maximum ground level concentration and the cloud width could be accurately estimated in each test.

Data from the Desert Tortoise experiment are available on a 9-track tape from LLNL, and a companion report similar to the ones prepared for Burro and Coyote is also available (Reference 16). No such report is scheduled to be produced for the Goldfish experiment. Data for the 3 dispersion trials (not the three mitigation effectiveness trials) were obtained from Mr. D. Blewitt of AMOCO (one of the sponsors of Goldfish), and much of the documentation for these trials may be found in a paper that appeared in the International Conference on Vapor Cloud Modeling (Reference 17). The individual trials contained in these reports include:

Desert Tortoise: 1, 2, 3, 4

Goldfish: 1, 2, 3

Table 3 provides an overview of these two field experiments. Most of the trials were performed during "neutral" stability conditions, with moderate wind speeds of 3 to 7 m/s. Although generally similar, note that Desert Tortoise trials differ from Goldfish trials in that the spill rates are about an order of magnitude greater.

3. Hanford Kr⁸⁵

The results from 13 dispersion trials conducted at the Atomic Energy Commission's Hanford reservation are reported by Nickola (Reference 18). Five of these trials involved the instantaneous release of small quantities of the inert radioactive gas krypton-85 (Kr⁸⁵), and the other eight involved short-period releases of Kr⁸⁵ over periods of ten to twenty minutes.

Up to as many as 64 detectors were operated along arcs located 200 m and 800 m downwind of the point of release. This section of the Hanford field diffusion grid is nearly flat, and is covered with sagebrush and steppe grasses. Most of the detector locations consisted of one detector set at 1.5 m above the surface. However, each row also included three towers on which five detectors provided a vertical profile of the Kr⁸⁵ clouds. The configuration is shown in Figure 4. Note that the uppermost detectors did not extend above the top of the diffusing clouds.

Meteorological data are reported for averaging periods of 1 minute, 5 minutes, and the period over which data were collected during a trial. These data are taken from the faster-response instruments mounted on the 25 m tower located near the source, when available. Otherwise, the data are reported from strip-charts recorded by instruments on the 122 m tower. Tabulations of time-series of meteorological and concentration data for both the instantaneous and continuous releases of Kr⁸⁵ are printed in the data report for the study (Reference 18). Wind speed, the standard deviation of wind speed and wind direction, and temperature are reported for consecutive 1-minute periods during each trial. Concentration data from the near-surface samplers (1.5 m above the ground) and the elevated sampling masts are reported at intervals of 38.4 seconds for the continuous release trials, and are reported at intervals of either 1.2, 2.4, or 4.8 seconds for the instantaneous release trials.

TABLE 3. SUMMARY OF DESERT TORTOISE AND GOLDFISH EXPERIMENTS.

Trial Name	Date	Duration (sec)	Spill Rate (m ³ min)	Averaged Wind Speed (m/s)	Averaged Wind Direction (degrees)	Atmospheric Stability Class
DT 1	24 Aug.	126	7.0	7.4	224	D
DT 2	29 Aug.	255	10.3	5.7	226	D
DT 3	1 Sept.	166	11.7	7.4	219	D
DT 4	6 Sept.	381	9.5	4.5	229	E
GF 1	1 Aug.	125	1.78	5.6	-	D
GF 2	14 Aug.	360	0.66	4.2	-	D
GF 3	20 Aug.	360	0.65	5.4	-	D

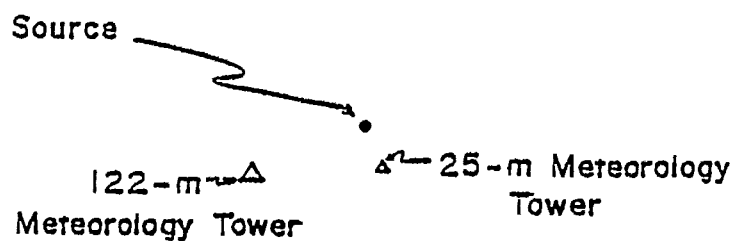
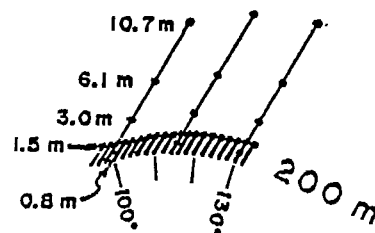
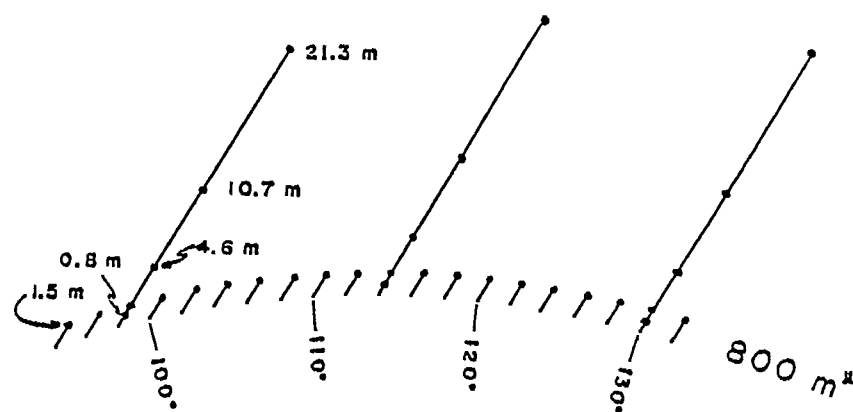


Figure 4. Configuration of Meteorological Towers and Kr^{85} Detectors for the Hanford Kr^{85} Trials (Reference 18).

As part of a project funded by the EPA, TRC Environmental Consultants, Inc. had entered the concentration data and meteorological data for six of the eight instantaneous release trials into computer files. The two trials dropped from use for this project were less desirable than the others because portions of the clouds drifted to the side of the array of detectors. We obtained these data and entered data for the five continuous-release trials into LOTUS 1-2-3 worksheets, preserving all of the information and structure of the original tables. The following trials comprise the data recorded on magnetic media:

Continuous-Release Trials: C1, C2, C3, C4, C5

Instantaneous Release Trials: P2, P3, P5, P6, P7, P8

A summary of the meteorological data for six of the eight instantaneous-release trials and all five continuous-release trials is presented in Table 4.

4. Maplin Sands

The dispersion and combustion trials conducted at Maplin Sands in 1980 (Reference 19) involved the release of liquefied natural gas (LNG) and refrigerated liquid propane (LPG) onto the surface of the sea. Each liquid was released in both a continuous and an instantaneous mode. The size of a spill during each trial was approximately 20 m^3 .

Because the objective of the trials was to study the behavior of LNG and LPG vapor clouds over the sea, the site was located on the tidal flats of the Thames estuary. A shallow dike 300 m in diameter was constructed around the spill area to meet the requirement that the spill occur on the sea surface. Pontoons with either 4 m masts or 10 m masts were used to position meteorological instruments and sampling instruments along arcs downwind of the spill area. Figure 5 shows the pontoon configuration at the start of the series, and Figure 6 shows the revised configuration used after Trial 35. A

TABLE 4. SUMMARY OF HANFORD KRYPTON-85 TRACER RELEASES.

Trial No. *	Date (1967)	Start (PST)	End (PST)	Duration (min & sec)	Total Emitted (Ci)	Release Rate (Ci/Sec)	Wind Speed at 1.5 m (mps)	Qualitative Thermal Stability
P2	Sep 14	2300:00	-	-	10.0	-	1.3	Very Stable
C1	Sep 15	0000:00	0015:28	15:28	10.9	0.0117	1.3	Very Stable
P3	Oct 17	0738:00	-	-	10.0	-	4.2	Neutral
C2	Oct 17	0801:50	0801:50	15:05	10.9	0.0120	3.9	Unstable
P5	Oct 23	1052:40	-	-	10.0	-	8.0	Unstable
C3	Oct 23	1101:25	1115:40	14:15	23.8	0.0278	7.1	Unstable
P6	Oct 23	1130:00	-	-	10.0	-	7.3	Unstable
P7	Oct 24	1052:30	-	-	10.0	-	4.6	Unstable
C4	Oct 24	1104:30	1114:28	9:58	22.8	0.0388	3.9	Unstable
C5	Nov 8	0512:22	0532:13	19:51	20.4	0.0171	2.6	Stable
P8	Nov 8	0602:00	-	-	10.0	-	1.5	Stable

* P: Denotes a puff (instantaneous) release
 C: Denotes a continuous (short-period) release

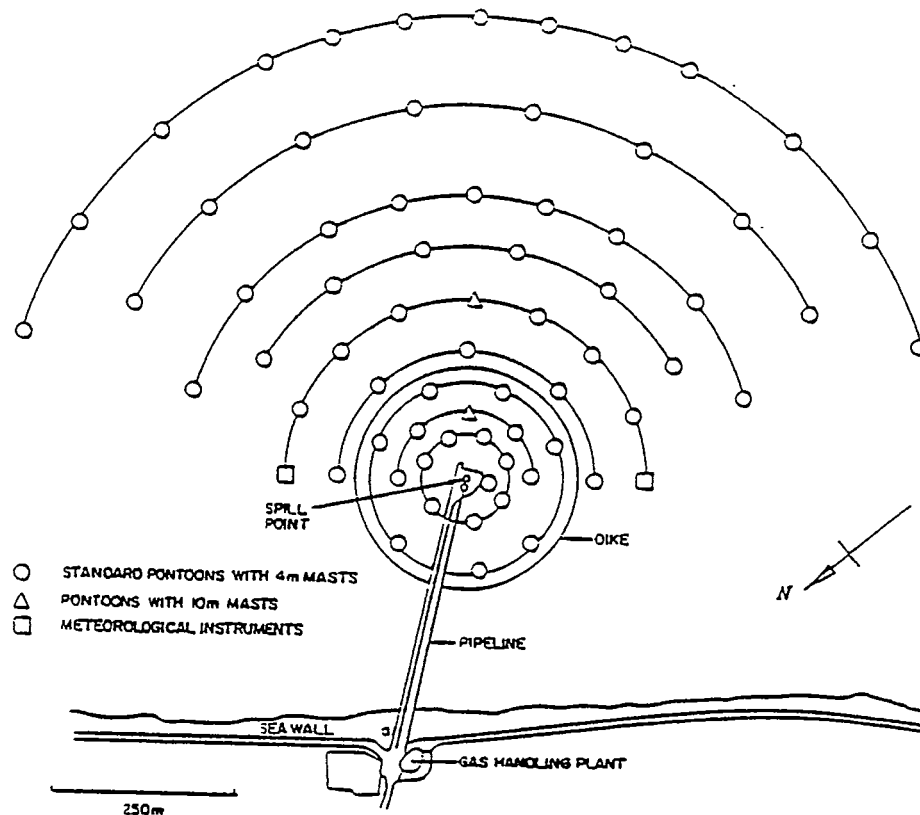


Figure 5. Initial Configuration of the Maplin Sands Site.

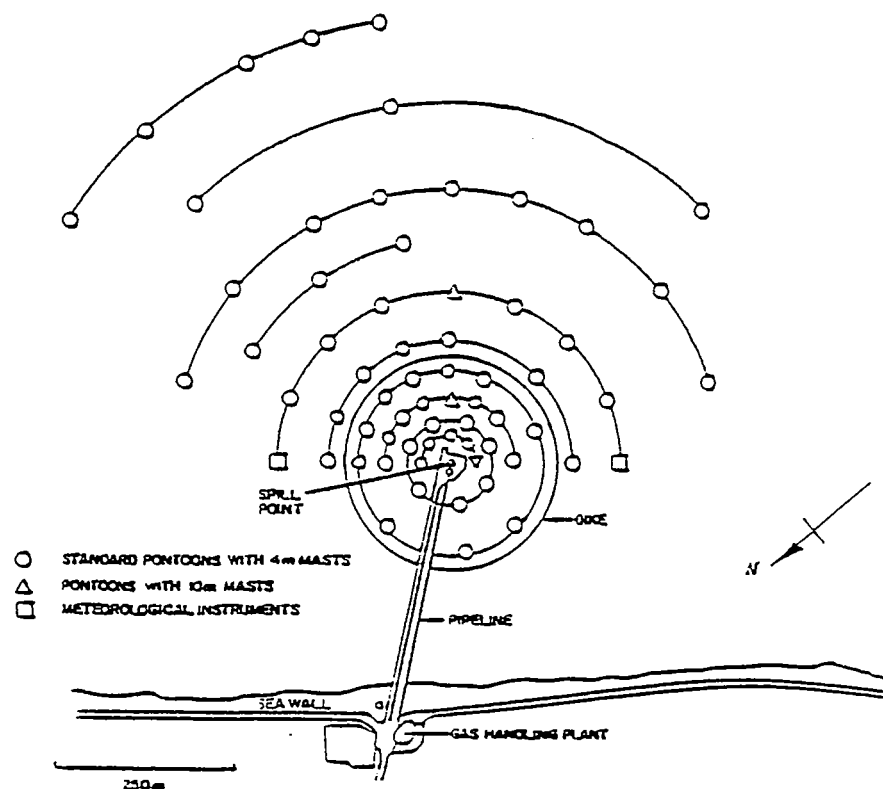


Figure 6. Revised Configuration of the Maplin Sands Site (After Trial 35).

total of 360 sensors were deployed in these trials, 200 of which were gas concentration sensors. Other types of sensors included:

<u>Parameter</u>	<u>Instrument Type</u>	<u>Number Of Sensors</u>
Wind speed	Cup anemometer	6
Wind direction	Vane	5
Turbulence	Ultrasonic anemometer	6
Air temperature	Platinum resistance	8
Relative humidity	Humicap	2
Insolation	Solarimeter	2
Sea surface roughness	Conductivity probe	1
Sea current	Turbine	2
Sea temperature	Platinum resistance	2
Cloud temperature	Thermocouple	66

Table 5 summarizes features of each of the Maplin Sands trials. The combustion aspects of some of these trials removes the vapor cloud, so the dispersion data are available only up to the moment of ignition. Not all of these trials are used in the performance evaluations. None of the instantaneous trials are retained. Within the continuous propane trials, trial 45 is dropped due to unsteady winds, and trials 51 and 55 are dropped because the vapor-clouds largely "missed" the sampler array. Within the continuous LNG trials, trial 37 is dropped due to the buoyant nature of the cloud, trial 39 is dropped because the cloud was ignited within 1 minute of the release, and trial 56 is dropped because much of the cloud did not pass through the sensor array. Trials 9, 12, and 15 are also dropped from the study because much of the LNG evaporated in the air prior to reaching the water (Reference 20), thereby complicating the nature of the release (a simple evaporating pool description is not appropriate). Therefore, the trials actually used in the performance evaluation are:

LNG: 27, 29, 34, 35

LPG: 42, 43, 46, 47, 49, 50, 52, 54

5. Prairie Grass

Project Prairie Grass, designed by Air Force Cambridge Research Center personnel, was held in north central Nebraska near O'Neill in the summer of 1956 (Reference 21). Small amounts of SO_2 were released continuously over 10-minute periods from ground level in the 70 trials that

TABLE 5. SUMMARY OF THE MAPLIN SANDS EXPERIMENT.

Trial Number	Volume (m ³)	Rate (m ³ /min)	Duration of steady flow (s)	Wind Speed (m/s)	Comments
Continuous Propane					
42		2.5	180	3.7	Underwater release
43		2.3	330	5.5	
*45		4.6	330	~2	Wind very unsteady
46		2.8	360	8.1	
47		3.9	210	5.6	
49		2.0	90	6.2	Ignited
50		4.3	160	7.9	Ignited
*51		5.6	140	6.9	Ignited
					Plume center missed sensors
52		5.3	140	7.9	Underwater release
54		2.3	180	3.8	
*55		5.2	150	5.5	At edge of sensor array
Continuous LNG					
* 9		1.6	300	8.9	
*12		0.7-1.1	340	1.5	
*15		2.9	285	3.6	
27		3.2	160	5.5	Ignited
29		4.1	225	7.4	
34		3.0	95	8.6	
35		3.8	135	9.8	
*37		4.1	230	4.7	Pipe end below water surface
					Buoyant plume
*39		4.7	60	4.1	Ignited
*56		2.5	80	5.1	Plume only briefly over sensors
Instantaneous Propane					
*63	17.	-	-	3.4	
Instantaneous LNG					
*22	12.	-	-	5.5	Ignited
*23	8.5	-	-	6.6	Ignited briefly

- These trials are not included in the performance evaluation

comprised the project. Dosage measurements were made on arcs located at distances of 50, 100, 200, 400, and 800 meters downwind. About half of the trials were conducted during unstable daytime conditions and the rest were held at night with temperature inversions present. Meteorological measurements included wind speed, direction, and fluctuations in direction from cup anemometers and airfoil type wind vanes. Micrometeorological data, rawinsonde data, and aircraft soundings were also taken.

The site was located on virtually flat land covered with natural prairie grasses. The roughness length determined for the site by some of the researchers was 0.6 centimeters. Dosages were measured at a height of 1.5 meters along the arcs using midget impingers. The meteorological data were given as 10-minute averages.

Earlier, the Porton Down dataset was dropped because most of the data obtained in the monitoring array are in the form of dosages. Why, then, are the dosages obtained during Prairie Grass acceptable? The reason is that the duration of the Prairie Grass releases (10 minutes) is long enough to create a quasi-steady plume over the monitoring array. In the absence of meandering, the time series of concentrations that might have been measured would have a plateau-like appearance. The average concentration estimated from the dosage (assuming a time-scale equal to the duration of the release = 10 minutes) would then be a fair estimate of the peak concentration. The Porton Down data, on the other hand, involve instantaneous releases, which would result in a time series of concentrations that might have been measured which would have a peak-like appearance. The mean concentration estimated from the dosage and the time it takes such a cloud to pass a monitor is a poor estimate of the peak concentration. Hence, the dosages from the Prairie Grass dataset are more useful for evaluating model performance. Note, however, that the average concentrations are still expected to be less than the peak concentrations.

Table 6 provides a summary of the meteorology for a subset of 44 trials that will be used on this project. These 44 represent the best of the program, and have been used extensively by other researchers (for example, Reference 22; Reference 23; Reference 24).

TABLE 6. SUMMARY OF SELECTED METEOROLOGICAL DATA FROM PRAIRIE GRASS TRIALS.

TRIAL	U (m/s)	STABILITY CLASS	TRIAL	U (m/s)	STABILITY CLASS
7	4.2	B	37	4.6	D
8	4.9	C	38	4.1	D
9	6.9	C	41	4.0	E
10	4.6	B	42	5.3	D
13	1.3	F	43	5.0	C
15	3.4	A	44	5.7	C
16	3.2	A	45	6.1	D
17	3.3	D	46	5.2	D
18	3.5	E	48	8.0	D
19	5.8	C	49	6.3	C
20	8.6	D	50	6.6	C
21	6.1	D	51	6.1	D
22	6.4	D	53	2.5	F
23	5.9	D	54	4.0	D
24	6.2	D	55	5.4	D
25	2.8	A	56	4.3	D
28	2.6	E	57	6.7	D
29	3.5	D	58	1.9	F
32	2.2	F	59	2.6	F
33	8.5	D	60	4.9	D
34	9.0	D	61	8.0	D
36	1.9	F	62	5.2	C

6. Thorney Island

The Heavy Gas Dispersion Trials project at Thorney Island (Reference 25) organized by the British Health and Safety Executive consists of the following five types of trials:

- (1) Phase I - the instantaneous release of a preformed cloud of approximated 2000 m^3 of dense gas over flat terrain. Sixteen trials were carried out.
- (2) Phase II - Ten trials were carried out to study the effects of obstacles on Phase I-type releases.
- (3) Continuous release trials - Three trials in which approximately 2000 m^3 of heavy gas was released at a rate of $5 \text{ m}^3/\text{sec}$ over flat terrain.
- (4) GRI trial - A single Phase I-type of release.
- (5) Phase III- Six continuous release trials in which a fenced enclosure surrounded the gas container.

For this project, we are focusing on the Phase I trials (item #1) and the continuous-release trials (item #3). The instantaneous-release trials of Phase I are similar in design to those conducted earlier at Porton Down, but the size of the source is approximately fifty times larger, and continuous monitors were used to obtain concentration measurements.

A gas container with a volume of 2000 m^3 was filled with a mixture of freon and nitrogen. For instantaneous release trials, the sides of the gas container collapsed to the ground upon release. For continuous release trials, the gas container simply served as a storage tank. The gas would then be ducted below the ground to the chosen release position. The release mechanism was designed to give a ground-level release with zero vertical momentum.

A 30 m tall meteorological tower was located 150 m upwind from the release point. The instrumentation consisted of five cup anemometers, five temperature sensors, two sonic anemometers, and one sensor each for relative humidity, solar radiation and barometric pressure. Four trailer-mounted towers, with a total of eight sonic anemometers, were also deployed. Note that the 30 m tall tower was replaced by a 20 m tall tower for continuous release trials.

Thirty-eight towers were used to measure gas concentrations. Measurements were taken at four levels. The lowest gas sensor was positioned at a height of 0.4 m; and the highest at 4 m on towers close to the spill point, 10 m at most other towers, and 14.5 m at towers in the far field. The towers were placed on a rectangular grid with distances up to about 800 m from the release point. The four trailer-mounted towers mentioned previously also had gas sensors mounted at four different heights. The configuration of instrumentation at the site is shown in Figure 7.

Copies of the 9-track tapes containing data for the instantaneous release trials were obtained through the API, who had contributed to the experiments, and who are co-funding this model performance evaluation. We have averaged the 20-Hz data in blocks of 0.6 seconds each. Data for the continuous release trials were obtained from TRC Environmental Consultants, Inc., who had digitized plots of the data to produce data corresponding to 30-s averages. A total of 9 of the 16 continuous release trials and 2 of the 3 instantaneous release trials were retained for the evaluation. Trials 10, 11, and 46 were dropped because of wind-shifts during the trial; trial 4 was dropped because the cloud became elevated; trial 5 was dropped because the release mechanism malfunctioned, producing 2 clouds rather than 1; and trials 14, 15, and 16 were dropped because the density of the initial cloud appeared to be stratified. The trials included in this study are:

Instantaneous releases: 6, 7, 8, 9, 12, 13, 17, 18, 19

Continuous releases: 45, 47

A brief summary of the characteristics of the source emissions and the meteorological conditions for the Phase I and continuous release trials is given in Table 7.