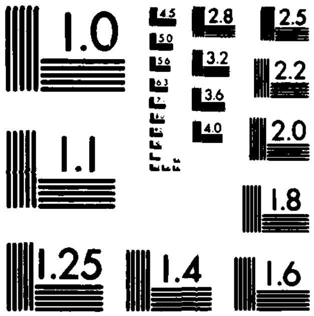
AD-A146 789 AIDS TO NAVIGATION PRINCIPAL FINDINGS REPORT: 1/2
VALIDATION FOR A SIMULATOR-..(U) ECLECTECH ASSOCIATES
INC NORTH STONINGTON CT M W SMITH ET AL. JUL 84
UNCLASSIFIED EA-84-U-203 USCG-D-06-84 DOT-CG-835285-A F/G 13/10 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

CG-D-06-84

(12)

# AIDS TO NAVIGATION PRINCIPAL FINDINGS REPORT: VALIDATION FOR A SIMULATOR-BASED DESIGN PROJECT

Eclectech Associates Division of
Ship Analytics, Incorporated
North Stonington Professional Center
North Stonington, Connecticut 06359

July 1984

Interim Report

Prepared for

U.S. Coast Guard
U.S. Department of Transportation
Office of Research and Development
Washington, D.C. 20593

# NOTICE

Technical Report Documentation Page

| | | |
|---|---|---|
| **1. Report No.** | **2. Government Accession No.**<br>AD A [illegible] | **3. Recipient's Catalog No.** |

**4. Title and Subtitle**
Aids to Navigation Principal Findings Report: Validation for a Simulator-Based Design Project

**5. Report Date**
July 1984

**6. Performing Organization Code**

**7. Author(s)** M.W. Smith, K.L. Marino, J. Multer, and J.D. Moynehan

**8. Performing Organization Report No.**
84-U-203

**9. Performing Organization Name and Address**
Eclectech Associates Division of
Ship Analytics, Inc.
North Stonington Professional Center
North Stonington, Connecticut 06359

**10. Work Unit No. (TRAIS)**

**11. Contract or Grant No.**
DOT-CG-835285-A

**12. Sponsoring Agency Name and Address**
Department of Transportation
United States Coast Guard
North Stonington Professional Center
North Stonington, Connecticut 20593

**13. Type of Report and Period Covered**
Interim Report

**14. Sponsoring Agency Code**
G-DST-1

**15. Supplementary Notes**

**16. Abstract**

The research reported here is part of the United States Coast Guard's Performance of Aids to Navigation (AN) Systems Project. The objective of the project is to provide guidelines for evaluation and design of AN systems in restricted waterways. The major effort has been the evaluation of aid systems under a variety of conditions on a marine simulator developed for the project, at Ship Analytics, Inc., North Stonington, Connecticut.

This report describes the validation of the simulator by comparing performance data collected at sea with performance data collected on the simulator. Ships were tracked electronically at sea and other data was collected manually in Chesapeake Bay in the approach to Baltimore, Maryland and Narragansett Bay in the approach to Providence, Rhode Island. The conditions observed in these restricted waterways were modeled on the USCG/SA simulator and pilot performance was compared. This evaluation identifies the strengths and limitations of the simulator and the need for possible adjustments.

**17. Key Words** at-sea data collection, ship bridge simulator, validation short range aids to navigation radio aids, Loran C, Raydist, ship tracking, piloting, buoy arrangement

**18. Distribution Statement**
Document is available to the public through the National Technical Information Service, Springfield, VA 22161

| **19. Security Classif (of this report)** | **20. Security Classif (of this page)** | **21. No. of Pages** | **22. Price** |
|---|---|---|---|
| UNCLASSIFIED | UNCLASSIFIED | 155 | |

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

# PREFACE

The objective of the United States Coast Guard's Performance of Aids to Navigation Systems project is to prepare guidelines for the design and evaluation of aid systems in restricted waterways. The Coast Guard's interest includes fixed and floating visual aids, radar, and radio aids. To provide quantitative data on which to base these guidelines, a series of experiments was done on two simulators, the Maritime Administration's Computer Aided Operation Research Facility (CAORF) at Kings Points, New York and a simulator developed for the project at Ship Analytics, Incorporated in North Stonington, Connecticut.

In 1982 at an interim point in the project, a draft manual was published summarizing completed components on the performance of visual and radio aids in a form useful as guidelines. "Draft SRA/RA Systems Design Manual for Restricted Waterways" is available from NTIS as AD-A113236.

The project is ongoing. The present phase of the work has included new simulator experiments done at Ship Analytics, Incorporated on the effectiveness of turnmarkings for nighttime plotting, on the effectiveness of short-range aids (buoys) for radar piloting, and on the effect of shiphandling factors on pilot performance. As of this writing, an experiment on the special needs of the meeting traffic situation is in progress at CAORF.

The continuing project includes two additional components meant to maximize the transfer of the findings to sea. To validate the USCG/SA simulator on which most of the experiments were run, performance data was collected in Chesapeake and Narragansett Bays. The present report describes the comparison of this at-sea data to performance data collected on the simulator. A model implementation of the draft manual is in progress in Narragansett Bay. Recommended aid changes have been made. New at-sea data will be collected and compared with the first at-sea sample to evaluate those recommendations.

A final SRA/RA design manual incorporating new data and experience with the draft manual is planned for 1985.

# ACKNOWLEDGEMENTS

The validation process has been spread over several states and five years in time. It was only possible through the involvement and cooperation of a great many people. As authors of the last report on validation, we hope we have recalled them all.

First, we want to thank those involved in the 1980 data collection in Chesapeake Bay. We can't do better than to repeat the acknowledgements that appeared in a preliminary report on those data:

> "The authors, on behalf of the U.S. Coast Guard, wish to express their sincere appreciation to the Association of Maryland Pilots and in particular Captain George A. Quick, President, without whose close support and involvement in this project it could not have been accomplished. From the office staff, dispatchers, auto drivers, and launch captains to the pilots, themselves, all went out of their way to accommodate our unusual requests and make us feel very welcomed. We thank you! We are also indebted to the many masters who permitted us aboard their ships for the brief voyage through Baltimore Harbor, and the special hospitality shown by the their crew members.
>
> "Finally, those of us on the 'tracking team' would like to thank Captain J.T. Montonye, USCG, for his close support in the project; and QM1 Bruce McIntosh who performed as the RAYDIST equipment operator. Quartermaster McIntosh accepted enormous responsibility during this experiment, both in coordinating activities with the pilot office, and in ensuring installation, checkout, and operation of the portable tracking equipment carried aboard each ship. As evident in this report, the collection of experimental data at sea requires extensive logistical support and cooperation among all personnel involved. The success of this particular study must, therefore, be attributed to the keen interest and special dedication of all its participants. Again, our thanks!"

We are also grateful to the Association of Maryland Pilots, Inc. for the matching simulator data. Members of that association made the trip to Connecticut out of interest in the aids to navigation project and the simulator. We are especially indebted to Captain Frederic S. Delano, Jr. who arranged their participation for us and who served as a consultant in the preparation of the simulation. We are grateful to each of those who came. They were Captains Duke Adams III, James A. Hannon, John Herbert, Brian H. Hope, Patrick G. Lynch, Thomas M. Quirk, Richard H. Schafer, and Davis R. Van Metre. They not only brought their expertise to the simulator, but also each one contributed to our understanding of piloting, of the Baltimore Approach and of simulation.

We want to express our gratitude to all those who helped in the Narragansett Bay data collection. Again, a repetition of the acknowledgements that appear in a preliminary report seems the most appropriate way:

"An at-sea data collection effort requires cooperation and logistical support from many directions. In conjunction with the Coast Guard, Ship Analytics would like to express appreciation to some of those groups and individuals who made this data collection project on Narragansett Bay possible.

"To the pilots, dispatchers, and pilot boat operators of the Northeast Marine Pilots, Inc., whose assistance in meeting the vessels and cooperation in accommodating data collection requests at sea were central to the completion of the project.

To the Marine Transportation Department of Mobil Oil Corporation, who allowed Ship Analytics' data collection teams to board their tankers; to the masters and crew members of the tankers themselves for their cooperation and hospitality; and to Mobil office personnel at the pier site who provided advance scheduling information on ship arrivals.

"Finally, to Coast Guard Research and Development personnel, who provided LORAN C and HP9915 tracking equipment, performed the waypoint survey of Narragansett Bay and provided programs for data collection and reduction of this data to a format compatible with Ship Analytics systems so it could be presented as the track plots contained in this report. In particular, the data collection team would like to thank LT Douglas Taggart at the Coast Guard's Avery Point, Connecticut, R&D Center who offered continual, unstinting assistance with both equipment problems and programming needs."

We want to thank the Northeast Marine Pilots, Incorporated of Newport, Rhode Island, for their participation in the Narragansett Bay validation, both at sea and on the simulator, and for much more. We have been heavily dependent on them as consultants and as pilots for most of the simulator experiments as well. At the time of this writing, they are helping us in a second at-sea data collection for a model implementation of the aids to navigation systems design manual and are participating in an additional experiment being run at CAORF, the Maritime Administration's Computer Aided Operation Research Facility in Kings Point, New York. All their members have been cooperative and supportive. We are especially indebted to Captain Kenneth Warner, President, who has expended considerable time and effort to meet our needs. The group has been central to the aids to navigation project.

It is not possible to mention and thank all the United States Coast Guard personnel who were connected with some aspect of the simulator validation. We can only include as representatives those who were central or recent. Commander Donald Murphy and Lieutenant H.H. Sharpe from the Office of Navigation, and Commander Robert Bates and Lieutenants John Anthony (now Lieutenant Commander) and Wallace Ridley from the Office of Development provided guidance and support for much of the validation effort. Lieutenants Anthony and Ridley also provided practical assistance in preparing for the collection of electronic data on Narragansett Bay and Lieutenant Ridley guided the preparation of this report. And we are grateful to Mr. Karl Schroeder of the Office of Development for providing continuity for us back to the beginning of the validation effort and the beginning of the project.

Even more than the Aids to Navigation project as a whole, the validation task was the effort of a very extended group.

# EXECUTIVE SUMMARY

## INTRODUCTION

The research reported here is a part of the United States Coast Guard's Performance of Aids to Navigation (AN) Systems Project. The objective of this project is to provide guidelines for the evaluation and design of AN systems in restricted waterways. The major effort in this project has been the experimental collection of performance data on aid systems under a variety of conditions on a marine simulator developed for the project at Ship Analytics, Inc., North Stonington, Connecticut. Performance data available at an interim point in 1982 was used to develop the "Draft SRA/RA Systems Design Manual for Restricted Waterways." Components of the project since then include additional simulator experiments, a validation of the simulator, and a test implementation at sea. A revision of the manual, benefiting from these later components, is planned for the spring of 1985.

The present report describes the validation of the USCG/SA simulator, a simulator consisting of complex ship hydrodynamic models and a simple computer-generated visual image designed to represent a waterway by its aids alone. The objective of validation was the evaluation of the simulator and its capability in supporting the AN project. At-sea and simulator conditions that matched as closely as possible in channel dimensions, environment, ship characteristics, shiphandler population, and aid arrangement were selected. Performance data collected in the two domains were compared. The strengths and limitations of the simulator and possible needs for adjustment were identified. The result of the validation is confidence in the simulator, the performance data collected on it, and the design manual developed from these data.

## CHESAPEAKE BAY VALIDATION

The first sample of at-sea performance data was collected in Chesapeake Bay in the approach to Baltimore, Maryland. This approach is an 800-foot wide channel marked by buoys and ranges, transited by a large variety of ship types. Observers rode ships, recording pilot behavior and, electronically, recording the ship track. The recorded transits were organized into two groups, similar in environmental conditions, but differing in ship size: smaller than 45,000 dwt and larger than 45,000 dwt.

A simulation of the channel represented by buoys and ranges was prepared for members of the Association of Maryland Pilots, Inc. They transited the "channel" with both the 30,000 and the 80,000 dwt tanker models used in the AN experiments. Comparison of performance data taken at sea and on the simulator produced the following findings.

- Performance in the straightaways provided a good match. Statistical tests demonstrated that the means of the two sets of tracks were the same within errors of 30 to 60 feet for much of the transit. These data came from a situation in which a ship's beam was approximately 100 feet and the channel width was 800 feet.

- There were some differences in the approach to the turn and in the turn, differences attributable to differences in pilot strategy.

- The 30,000 dwt tanker model provided a very good match to the performance of the group of small ships tracked in Chesapeake Bay. The 80,000 dwt tanker model showed more accurate trackkeeping than the group of large ships tracked.

The simulator provided useful performance data for a situation in which the dominant aids are visual ranges.

## NARRAGANSETT BAY VALIDATION

A second at-sea data collection was made in Narragansett Bay in the approach to Providence, Rhode Island. (This same data was used as a baseline for the test implementation of the manual.) This channel is 600 feet wide with a number of turns widened by dredging and is marked with lighted and unlighted buoys and a number of lighted structures. The ships tracked were tankers very much like the 30,000 dwt tanker model used in the AN experiments. The available data was divided into day and night transits. There were meaningful differences at sea from day to night in pilot strategy, mostly in the turns and in the vicinity of unlighted aids.

A simulation of the channel represented by its aids was prepared for members of Northeast Marine Pilots, Inc. They made the simulator transits under day and night conditions. Comparison of performance data taken at sea and on the simulator provided the following findings.

- Performance in the straightaways again provided a good match. Statistical tests demonstrated that the means of the two sets of tracks were the same within errors of approximately 70 feet for much of the transit. These data came from a situation in which a ship's beam was approximately 85 feet and the channel width was 600 feet.

- Performance in the turns provided a good match when local pilot strategies transferred from sea to simulator. Statistical tests demonstrated that the means were the same within errors of 50 to 80 feet through the turn. These data came from a situation in which the ship's beam was approximately 85 feet and the channel turn area was widened beyond 600 feet.

- Local strategies for making uniquely configured turns did not always transfer from sea to the simulator.

- For sections of the transit where at-sea performance was not directly dependent on the aids, there was not a good match.

The simulator provided useful performance data for situations in which performance is dependent on visual aids on the channel edge.

VALIDATION OF GENERIC DATA

The primary objective of simulator use for the AN project was the collection of generic -- not specific to any single waterway -- performance data for the experiments and for the design manual. For the objectives of the experiments and of the manual, most of these data were collected under conditions that would not be expected at sea with any frequency. However, some sample of the experimental conditions can be matched by at-sea conditions. The comparisons provided the following findings.

- In buoyed channels, in channels marked by ranges, for daytime transits, and for nighttime transits, simulator performance provided a good match for at-sea performance. Statistical tests demonstrated that the means of the two sets of tracks were the same within errors of 20 to 60 feet for critical points in the transits. These data came from situations in which the ship's beam was 85 and the channel width was 500 feet.
- Trackkeeping performance with the 30,000 dwt tanker model again matched the performance of small ships at sea better than trackkeeping performance with the 80,000 dwt tanker model matched the performance of large ships.
- An increase in the cautiousness of turn strategy from day to night was observed at sea in Narragansett Bay and in the generic turn data. (In both cases the pilots avoid the uncertain outside edge by turning harder at night.) Differences in turn geometry made a quantitative comparison impossible, but the observation of this unexpected behavior both at sea and on the simulator supported the general validity of turn performance on the simulator.

The comparisons made between a sample of at-sea performance and a sample of generic performance data supports the validity of the whole.

CONCLUSIONS

The findings of the validation effort support conclusions on the future use of the USCG/SA simulator for several purposes and on the use of the available performance data already collected.

- The use of the USCG/SA simulator for the experimental or generic evaluation of aid arrangements is the least demanding task for the simulator. In a generic waterway there is little expectation that very subtle cues be present or that the pilots transfer very specific behaviors. A visual scene that includes only simple representations of the aids is a conservative and appropriate test of generic conditions. The validation findings support confidence in the simulator for this use, in the data already collected, and in the use of these data for the development of the design manual.
- The use of the USCG/SA simulator for the evaluation of aid systems in specific waterways is also a relatively undemanding task for the simulator. A visual simulation that consists only of the aids in the

waterway simply presented is, again, a conservative test of the aids. Poor performance on the simulator identifies those regions of the waterways where the aids are potentially weak, where they function at sea because they are supplemented by other factors like cultural objectives or well-rehearsed pilot strategies relatively independent of visual cues. The present simulator capability and experimental methodology is appropriate for this use.

- The use of the USCG/SA simulator to reproduce the unique performance to be expected at sea in a specific waterway is the most demanding task for the simulator. Such reproduction will occur only if the pilots transfer accustomed behavior to the simulator and if conditions on the simulator are sufficiently similar to allow them to implement that behavior. Any limitations to reproduction that were observed have consequences only for the validation effort itself. They have no consequences for the project's other uses of simulator performance data.

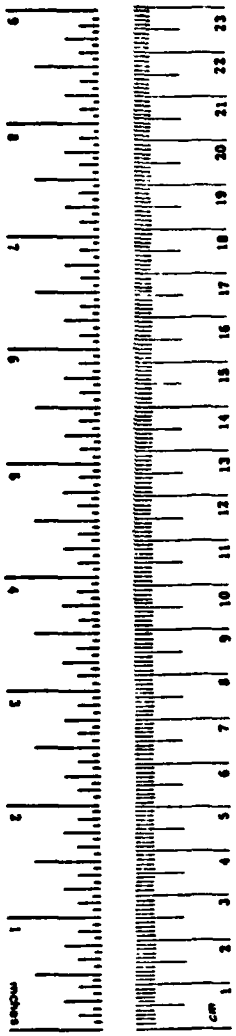
# METRIC CONVERSION FACTORS

## Approximate Conversions to Metric Measures

| Symbol | When You Know | Multiply by | To Find | Symbol |
|---|---|---|---|---|
|  |  | **LENGTH** |  |  |
| in | inches | *2.5 | centimeters | cm |
| ft | feet | 30 | centimeters | cm |
| yd | yards | 0.9 | meters | m |
| mi | miles | 1.6 | kilometers | km |
|  |  | **AREA** |  |  |
| in$^2$ | square inches | 6.5 | square centimeters | cm$^2$ |
| ft$^2$ | square feet | 0.09 | square meters | m$^2$ |
| yd$^2$ | square yards | 0.8 | square meters | m$^2$ |
| mi$^2$ | square miles | 2.6 | square kilometers | km$^2$ |
|  | acres | 0.4 | hectares | ha |
|  |  | **MASS (weight)** |  |  |
| oz | ounces | 28 | grams | g |
| lb | pounds | 0.45 | kilograms | kg |
|  | short tons<br>(2000 lb) | 0.9 | tonnes | t |
|  |  | **VOLUME** |  |  |
| tsp | teaspoons | 5 | milliliters | ml |
| Tbsp | tablespoons | 15 | milliliters | ml |
| fl oz | fluid ounces | 30 | milliliters | ml |
| c | cups | 0.24 | liters | l |
| pt | pints | 0.47 | liters | l |
| qt | quarts | 0.95 | liters | l |
| gal | gallons | 3.8 | liters | l |
| ft$^3$ | cubic feet | 0.03 | cubic meters | m$^3$ |
| yd$^3$ | cubic yards | 0.76 | cubic meters | m$^3$ |
|  |  | **TEMPERATURE (exact)** |  |  |
| °F | Fahrenheit temperature | 5/9 (after subtracting 32) | Celsius temperature | °C |

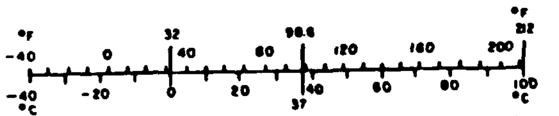
---

*1 in = 2.54 (exactly). For other exact conversions and more detailed tables, see NBS Misc. Publ. 286, Units of Weights and Measures, Price $2.25, SD Catalog No. C13.10:286.



## Approximate Conversions from Metric Measures

| Symbol | When You Know | Multiply by | To Find | Symbol |
|---|---|---|---|---|
|  |  | **LENGTH** |  |  |
| mm | millimeters | 0.04 | inches | in |
| cm | centimeters | 0.4 | inches | in |
| m | meters | 3.3 | feet | ft |
| m | meters | 1.1 | yards | yd |
| km | kilometers | 0.6 | miles | mi |
|  |  | **AREA** |  |  |
| cm$^2$ | square centimeters | 0.16 | square inches | in$^2$ |
| m$^2$ | square meters | 1.2 | square yards | yd$^2$ |
| km$^2$ | square kilometers | 0.4 | square miles | mi$^2$ |
| ha | hectares (10,000 m$^2$) | 2.5 | acres |  |
|  |  | **MASS (weight)** |  |  |
| g | grams | 0.035 | ounces | oz |
| kg | kilograms | 2.2 | pounds | lb |
| t | tonnes (1000 kg) | 1.1 | short tons |  |
|  |  | **VOLUME** |  |  |
| ml | milliliters | 0.03 | fluid ounces | fl oz |
| l | liters | 2.1 | pints | pt |
| l | liters | 1.06 | quarts | qt |
| l | liters | 0.26 | gallons | gal |
| m$^3$ | cubic meters | 35 | cubic feet | ft$^3$ |
| m$^3$ | cubic meters | 1.3 | cubic yards | yd$^3$ |
|  |  | **TEMPERATURE (exact)** |  |  |
| °C | Celsius temperature | 9/5 (then add 32) | Fahrenheit temperature | °F |



## TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

## LIST OF ILLUSTRATIONS

## LIST OF TABLES

# Section 1

# INTRODUCTION

## 1.1 THE AIDS TO NAVIGATION PROJECT

The research reported here is a part of the United States Coast Guard's performance of Aids to Navigation Systems Project (AN). The objective of this project is to prepare a decisionmaking tool for the design and evaluation of aids to navigation systems in restricted waterways. To provide quantitative data for the design process, the major effort in the project has been simulator evaluation of aids to navigation systems under a variety of channel, ship, and environmental conditions. The evaluations have included fixed and floating visual aids, radio aids (Loran C), and radar. In 1982 at an interim point in the ongoing project, the available data was used in the "Draft SRA/RA Systems Design Manual for Restricted Waterways."[1] The SRA/RA manual considers both Short-Range Aid and Radio Aid Systems. Recent simulator evaluations include the performance of buoy light flash characteristics for nighttime piloting,[2] the performance of aids with passive reflectors for radar piloting[3], and the performance of floating aids under varied shiphandling conditions.[4] An experiment evaluating the performance of aids for meeting traffic experiment is also planned.[5] Present components of the project include a model implementation of the draft manual and the validation of the simulator on which most of the performance evaluations were made. A revision of the manual benefiting from the new data, the implementation and validation, and the experience with the draft manual is planned for the spring of 1985.

---

[1] W.R. Bertsche, M.W. Smith, K.L. Marino and R.B. Cooper. "Draft SRA/RA Systems Manual for Restricted Waterways." CG-D-77-81, U.S. Coast Guard, Washington, D.C., February 1982. NTIS AD-A113236.

[2] J. Multer and M.W. Smith. "Aids to Navigation Turn Lights Principal Findings: Effect of Turn Lighting Characteristics, Buoy Arrangements, and Ship Size on Nighttime Piloting." CG-D-49-82, U.S. Coast Guard, Washington, D.C., February 1983. NTIS AD-A-126080.

[3] J. Multer and M.W. Smith. "Aids to Navigation Radar I Principal Findings: Performance in Limited Visibility of Short-Range Aids with Passive Reflectors." CG-D-79-83, U.S. Coast Guard, Washington, D.C., December 1983.

[4] K.L. Marino, M.W. Smith and J.D. Moynehan. "Aids to Navigation SRA Supplemental Experiment Principal Findings: Performance of Short-Range Aids Under Varied Shiphandling Conditions." U.S. Coast Guard, Washington, D.C., December 1983.

[5] J.D. Moynehan, M.W. Smith and J.W. Gynther. "Aids to Navigation Presimulation Report: Aids in the Meeting Traffic Situation." U.S. Coast Guard, Washington, D.C. February 1984.

Most of the performance data was collected on a marine simulator designed for the project at Ship Analytics, Incorporated, in North Stonington, Connecticut. The simulator is described in Appendix A. The validation is an evaluation of the simulator in its ability to represent at-sea conditions. This report describes two sets of at-sea data, collected in Chesapeake and Narragansett Bays, and their use in the evaluation of the USCG/SA simulator. A summary of the reports leading up to this evaluation is presented in Appendix B.

## 1.2 THE VALIDATION PROCESS

Research or training is conducted on a marine simulator rather than on a ship, when the ship is not a practical alternative. A simulator is used because of difficulties associated with at-sea effort: it is too expensive or too time-consuming, critical conditions occur too infrequently or are too dangerous to induce, even frequent conditions cannot be selected or replicated, and accurate and complete data are frequently impossible to collect. Despite these pragmatic and compelling reasons for simulator use, there are also reasons for caution. A simulator effort is worthwhile only to the extent that the results transfer to sea and they transfer to sea only to the extent that the simulator is able to represent real-world conditions. "Validation" is an evaluation of that representativeness and of the potential for transfer. The objective of validation is not an all-or-none acceptance or rejection of simulator use. Rather, it is the identification of its strengths and limitations and its possible need for adjustment.

Some general discussion about the nature of the simulator validation for AN is a useful introduction to the present report. The general methodology for simulator validation is illustrated in Figure 1. At-sea and simulator conditions are selected to be as similar as possible. Similar conditions include similar pilot populations and behavior. Data on the performance of interest must be collected in similar form. The evaluation is done by comparing the two sets of performance data. The intention is to evaluate the contribution of simulator capability to the similarities and differences observed in the performance data, but other steps in the process contribute and need to be considered.

### 1.2.1 Validation Is Dependent On At-sea Data

There is a question about the validity of a simulator until it is demonstrated. There is no question but that the user has a high degree of control over the selection of simulator conditions (within the capability of the simulator), their replication, and the collection of a variety of data. The situation is reversed for the at-sea sample to be matched. There can be no question about the "validity" of the real-world, but for the same reasons that a project is not conducted at-sea in its entirety, the data sample presents problems. It is inevitable that at-sea data be characterized by small sample size, uncertainty in specifying the conditions under which it was collected, and, when electronic ship tracking is involved, positioning error that may be meaningful in a narrow channel. The degree of similarity possible between performance at sea and on the simulator is determined by

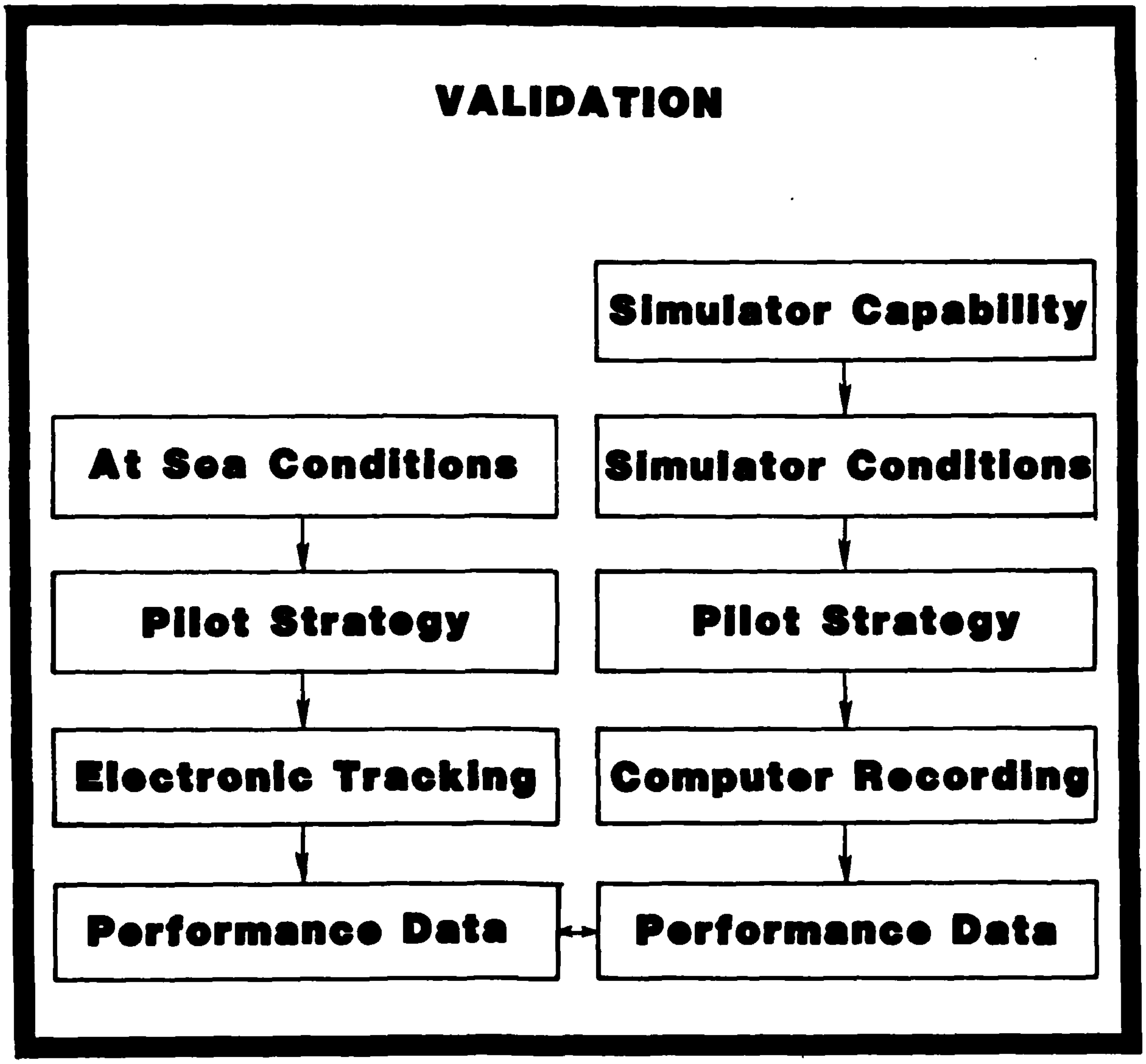


Figure 1. The General Methodololgy for Simulator Validation

the characteristics of the at-sea data as well as by the effectiveness of the simulator.

While AN's at-sea data samples present a unique opportunity for validation, there have been difficulties with both sets of data. In Chesapeake Bay[6,7] tracking the ship with Raydist equipment involved some error and some loss of runs. The ships available varied widely in configurations, dimensions, and maneuverability. Information on maneuverability was not always available in any useful form. Environmental conditions (time of day, current, wind, visibility, etc.) varied widely. Wind and current especially were difficult to specify with any accuracy. Traffic in the channel resulted in excursions in ship tracks that obliterated the effects of any other conditions. In Narragansett Bay[8] Loran C equipment also involved some error and loss of runs. There, the ships were quite similar, even repetition of the same ship; but they were less frequent than anticipated. While ships were less frequent in Narragansett Bay, there, too, some data was discarded because of traffic. Limitations to the data must affect the comparisons that can be made and the conclusions that can be drawn.

### 1.2.2 Validation Is Designed For A Simulator Configuration

At the beginning of a validation project, it is necessary to specify what is to comprise the "simulator". Early in the history of the AN project, some assumptions were made about how elaborate a simulation was needed to evaluate the performance of aid arrangements under varying combinations of channel dimensions, environmental conditions, and ship characteristics. The capabilities of the USCG/SA simulator included only a simple computer-generated visual scene, both to keep costs low and on the assumption that this simplicity would meet research objectives. It was assumed that the performance of the aids could be evaluated using a visual scene that included only ownship's bow and simple representation of the buoys for day and single-intensity flashing lights at night. The absence of landmasses and other fixed objects that might provide additional cues or interfere with the identification of the aids allows unambiguous interpretation of the findings as an evaluation of the aids arrangements. (The lack of capability for landmass also means the lack of a capability for traffic. The meeting traffic situation was investigated in an early AN experiment[9] done at CAORF, the Maritime Administration's Computer Aided

---

[6] Eclectech Associates. "Aids to Navigation Preliminary Data Analysis for the At-sea Data Collection Program." Technical Memorandum. U.S. Coast Guard, Washington, D.C., August 1981.

[7] R.B. Cooper, R.C. Cook and K.L. Marino. "At-sea Data Collection for the Validation of Piloting Simulation." CG-D-60-81. U.S. Coast Guard, Washington, D.C., December 1981. NTIS AD-A111978.

[8] J.D. Moynehan. "Aids to Navigation Implementation At-Sea Preimplementation Draft Principal Findings." U.S. Coast Guard, September 1983.

[9] M.W. Smith and W.R. Bertsche. "Aids to Navigation Principal Findings on the CAORF Experiments: The Performance of Visual Aids to Navigation as Evaluated by Simulation." CG-D-51-81, U.S. Coast Guard, Washington, D.C., February 1981. NTIS AD-A107045

Operations Research Facility at Kings Point, New York, and will be further investigated in a second experiment[10] planned there for the near future.) The ship hydrodynamic models are relatively sophisticated and include shallow water effects for transits in restricted waterways. Because of the possibility that bank forces might provide cues that supplement the aids in keeping a ship in a channel, these were never developed for inclusion in the AN simulations. (The contribution of bank forces will be evaluated in the planned traffic experiment at CAORF.) Wind and current effects were available and were used in most of the AN experiments.

Any evaluation of the simulator validity is specific to the components that were in use at the time performance was observed. Several earlier AN reports document the magnitude of performance changes that can be expected when the visual scene is held constant but the components that affect the associated shiphandling task are changed.[11,12,13,14] Several of those performance changes are summarized in Appendix C. Because these effects can be sizable each comparison of performance on the simulator to a sample of at-sea performance must specify the simulator components involved.

The AN validation includes the evaluation of the contribution to the full simulation of several of its components. The Chesapeake Bay performance sample is split into transits with smaller and larger ships and these are compared to simulator performance using the 30,000 dwt tanker model and the 80,000 dwt tanker model, respectively. The Narragansett Bay sample is split into day and night runs and each of these is compared to an appropriate simulation. Differences in the performance match from small to large ship, from day to night, are relatively specific evaluations of components of the simulator.

### 1.2.3 Validation Is Designed For A Simulator Application

Validation for AN depends on more than the simulator narrowly defined as the visual scene, ship models, environmental forces, etc. The objective of simulator use is to generate performance data that will substitute for at-sea performance data in the analysis of a specific waterway (Implementation) or in the preparation of the SRA/RA design manual. More than the simulator contributes to the observed performance. Therefore, more than the simulator needs to be considered in validation. The validity of the performance data may depend on such factors as the conditions selected

---

[10] J.D. Moynehan, M.W. Smith and J.W. Gynther, February 1984, op. cit.

[11] W.R. Bertsche, D.A. Atkins and M.W. Smith. "Aids to Navigation Principal Findings Report on the Ship Variables Experiment: The Effect of Ship Characteristics and Related Variables on Piloting Performance." CG-D-55-81, U.S. Coast Guard, Washington, D.C., October 1981. NTIS AD-A108771.

[12] J. Multer and M.W. Smith, February 1983, op. cit.

[13] J. Multer and M.W. Smith, December 1983, op. cit.

[14] K.L. Marino, M.W. Smith and J.D. Moynehan, December 1983, op. cit.

for simulation, the criterion shiphandling tasks assigned the pilot, the measures used to quantify performance, etc. These less-concrete factors contribute to the validity of the performance observed on the simulator and the degree to which it will represent at-sea performance.

For the AN project the primary form of performance data has been the "trackplot", a plot of the crosstrack position of the ship's center of gravity as a function of its alongtrack progress in the channel. Human factors data, data on how the pilots achieve that track, has been secondary. This priority might not be appropriate if the simulation was intended for another application such as training.

## 1.3 VALIDATION FOR THE AIDS TO NAVIGATION PROJECT

The AN project used the simulator for conceptually different objectives. The objective for the Validation and Implementation tasks was to generate performance data to match that collected or expected in a specific waterway. The other objective was to generate performance data to represent generic conditions for the experiments or for the manual. The two uses present slightly different validation problems.

### 1.3.1 Simulation for a Specific Waterway

The method chosen for the AN simulator validation was a comparison of performance at sea in a specific waterway with performance in a simulation of that waterway. This objective is the most demanding for a simulator and its validity. A simulation is never meant to be a complete reproduction of the real world conditions. It is meant to represent the real world conditions by including sufficient critical elements to support the local pilots' habitual strategies. This means, first, the simulator must have the capability to provide the critical elements. Second, within the simulator's capability, the elements selected for simulation must be sufficient. For example, in the simulations discussed in the present report, landmasses that might have helped or hindered performance at sea were omitted because they were beyond the simulator's present capability. By selection, off-the-channel lights visible at night were omitted. Minimal wind and current effects were omitted. Available ship models were used rather than models of the specific ships that were tracked. While the local pilots association was represented in the simulation, only rarely were the same individuals observed in both domains. The simulator, narrowly defined, is only a part of what is validated by a comparison of the two performance samples. The selection of what is appropriately representative of the at-sea conditions is a very important part as well.

### 1.3.2 Simulation for Generic Performance

A second use of the simulator in the AN project was to generate performance data under generic conditions for the experiments or for the manual. Here, "generic" performance is meant to be applicable in a wide range of waterway situations, although the conditions might not be a highly representative for any one. For example, trackkeeping in a narrow channel with gated buoys happens frequently, but for a specific waterway the ship characteristics, buoy spacing, etc., might not be exactly right. This objective is less demanding for the simulator and its validity. There is no expectation that the pilots will transfer very specific strategies.

The collection of generic performance data was designed to meet the objectives of specific experiments or of the manual. Often, to meet these objectives, the conditions under which it was collected were deliberately nonrepresentative of at-sea conditions. A conspicuous example of nonrepresentative conditions is the difficult shiphandling problem imposed on the experimental scenarios to ensure differentiation among conditions for the experiments and to ensure conservatism to estimates of risk for the manual. Another example is the concentration in the experiments on the higher-risk, higher-angle noncutoff turns which do not occur with a very high frequency in the real world.

The principal difference between the two uses of simulation for the AN project is the degree of expectation that the conditions be comparable to real-world conditions or that the performance be comparable to real-world performance and thus, amenable to validation. The expectation is not as great for the generic performance data. However, the available at-sea data presents a unique opportunity to evaluate some sample of the generic data. Some scenarios or parts of scenarios are more representative of at-sea conditions that others. Validation of any sample of performance increases confidence in the whole.

## 1.4 ORGANIZATION OF THIS REPORT

Section 2 describes data collected on the Upper Chesapeake Bay, and Section 3 describes data collected in the Upper Narragansett Bay. The at-sea data is compared to the simulation of the seaway to quantitatively determine similarities and differences in performance and apply these to a simulator evaluation. Section 4 compares the at-sea data to selected generic data to further the evaluation of the simulator. Section 5 summarizes the findings and conclusions of the simulator evaluation. Because of the size and complexity of the validation effort, not all relevant material is included in the body of this report. There are six supplemental appendices: Appendix A describes the characteristics of the USCG/SA simulator, Appendix B identifies the components of the validation tasks and includes an annotated bibliography, Appendix C reviews validation-related issues that were analyzed in other reports, Appendix D presents the Chesapeake Bay simulator and at-sea track plots, Appendix E presents the Narragansett Bay simulator and at-sea track plots, and Appendix F provides a brief description of the statistical procedures used in the data analysis.

# Section 2

# CHESAPEAKE BAY: AT SEA AND ON THE SIMULATOR

## 2.1 INTRODUCTION

Chesapeake Bay was used as the setting for the first validation effort of the Aids to Navigation project. The general methodology by which the conditions tested on the USCG/SA simulator and similar conditions in the real world were compared is illustrated in Figure 2. Due to the fact that some of the variables studied in Aids to Navigation experiments such as wind and current, visibility as well as some extraneous variables such as traffic could not be controlled by the investigators, the at-sea data was collected before the simulated Chesapeake scenarios were designed. The simulated scenarios of Chesapeake Bay were designed after collecting at-sea data to increase the likelihood that the two situations compared were as similar as possible. Following the data collection in Chesapeake Bay, a selection procedure was used to choose the at-sea data to be compared with the simulator data and to group it according to the variable to be evaluated. The remaining data was analyzed to prepare for the development of a simulator experiment that would produce results that could be compared directly to the at-sea data. The data collection and analyses for the simulation experiment followed a parallel course to the at-sea data with one exception. Greater control over the environmental conditions and data recording procedures reduce the data selection process to an absolute minimum. The validation process was completed by comparing the results of the at-sea analysis with the analyses of the simulated Chesapeake scenarios.

This section reviews the methodology used to collect the at-sea data and perform the simulator validation experiment. Steps that have been described in earlier reports are identified in the discussion that follows. An annotated bibliography in Appendix B briefly summarizes each of these reports. This section adds a quantitative analysis of the comparability of performance in the two domains to what has been reported before.

### 2.1.1 At-Sea Data Collection Methodology

The Craighill Channel and Craighill Channel Upper Range in the upper Chesapeake Bay shown in Figure 3 were selected as the sites for the collection of at-sea data. This choice was based on its similarity, in terms of channel width, angle of bend and buoyage, to scenarios used in previous AN simulator experiments. In addition, the choice of this restricted waterway was based upon an expected high frequency of transits. It was possible to collect data under a variety of conditions needed to evaluate the USCG/SA simulator within reasonable time.

Pilots in the Association of Maryland Pilots participated in the at-sea data collection according to their normal rotation cycle assigned to them by the Association of Maryland Pilots. A total of 20 participated, each serving only once.

Teams of investigators measured pilot performance by using Raydist electronic tracking equipment to record the vessel's crosstrack position in

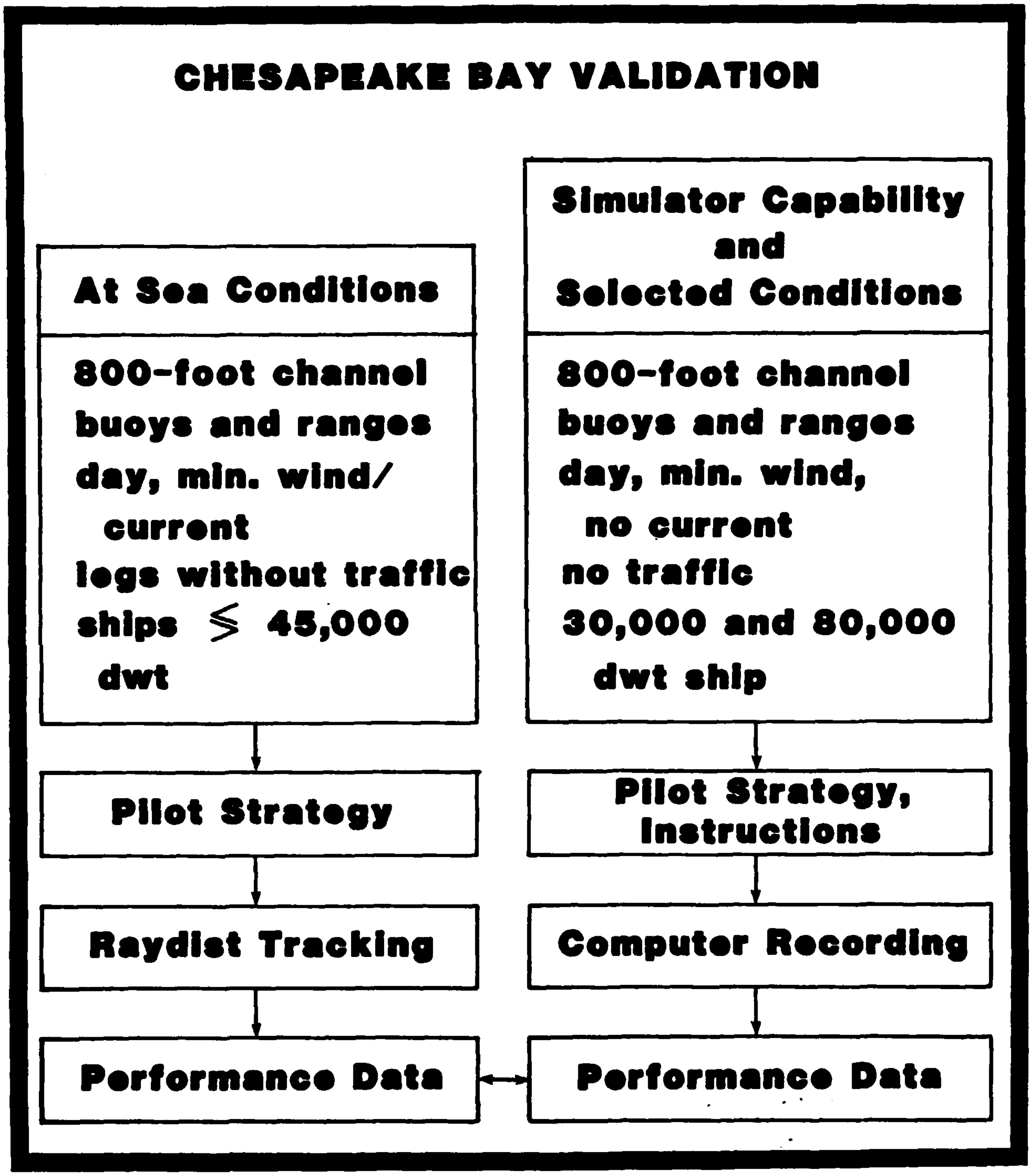


Figure 2. The Validation Methodology for Chesapeake Bay

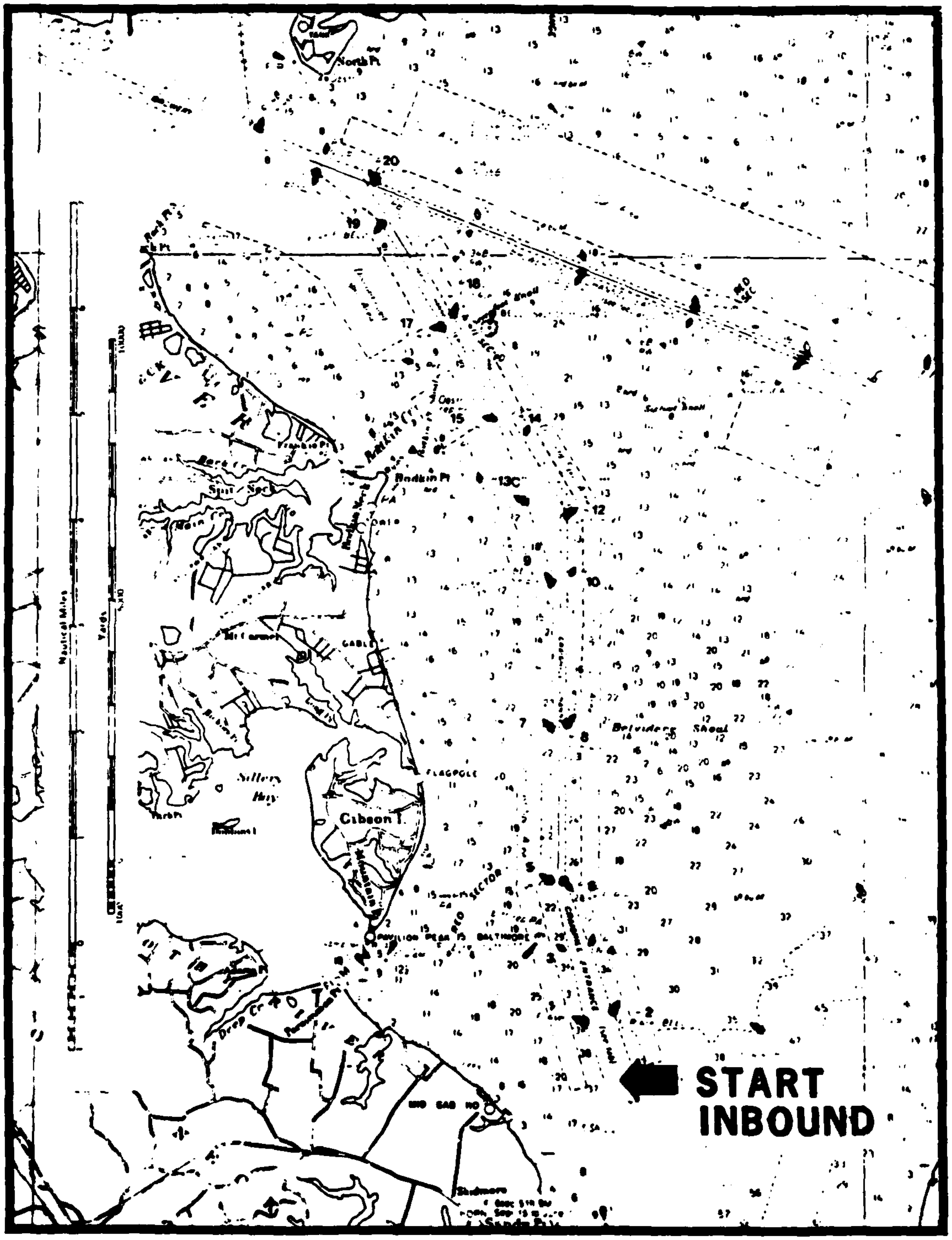


Figure 3. Upper Chesapeake Bay Where At-Sea Data Was Collected
Chart No. 12273 of April 17, 1982

the channel. In addition, the investigators interviewed the pilots and recorded their helm and engine order commands. The detailed procedure is described in the report entitled "At-Sea Data Collection Program for the Validation of Piloting Simulation and Evaluation of Radar Aided Piloting Techniques."[15]

### 2.1.2 At-Sea Data Selection

The inherent lack of control over all environment conditions made a screening process necessary to select data that would represent the pilot's at-sea performance. Of the 42 leg transits that were collected, 27 were eliminated. There were two reasons for eliminating data from the subsequent analyses. Several runs were eliminated because of difficulties with the electronic recording equipment. Runs were also eliminated because of frequent traffic encounters in which avoidance maneuvers prevented the pilot from accomplishing the experimental task of staying on the centerline of the channel. This selection process is described in the report entitled "At-Sea Data Collection for the Validation of Piloting Simulation.[16]

The remaining leg transits were analyzed by grouping according to a number of variables: these included ship size, ship type, wind, current, day/night, etc. Only one variable, ship size, had an effect on performance and was represented by a reasonable sample size. The available leg transits were split into ships larger and smaller than 45,000 dwt as summarized in Table 1. This analysis is also described in the report entitled "At-Sea Data Collection for the Validation of Piloting Simulation."[17] These data were used for comparison with simulator performance.

### 2.1.3 Simulator Data Collection

In order to accurately compare at-sea and simulator data, the conditions produced on the simulator were designed to be as close as possible to the conditions representative of those found in Chesapeake Bay. Ship size was the only variable manipulated in the simulated scenarios. All other variables remained constant. Table 2 shows the conditions for the Chesapeake Bay scenarios. Pilots from the Association of Maryland Pilots served as the subjects in this experiment. Their task was the same as in the at-sea collection phase; to pilot the simulated vessel up the channel shown in Figure 2, maintaining a track as close as possible to the center of the channel. Finally, the same performance measures were collected as well. This included crosstrack position throughout the channel, helm and

---

[15] R.C. Cook. "At-Sea Data Collection Program for the Validation of Piloting Simulation and Evaluation of Radar Aided Piloting Techniques." U.S. Coast Guard, Washington, D.C., March 1981.

[16] R.B. Cooper, R.C. Cook, and K.L. Marino. "At-Sea Data Collection for the Validation of Piloting Simulation." U.S. Coast Guard, Washington, D.C., December 1981.

[17] Ibid.

TABLE 1. SHIPS COMPRISING AT-SEA GROUPS

| Classification | Ship Size (1000 dwt) | Length (feet) | Beam (feet) | Draft (feet) | Type | Data Used for Leg 1 | Data Used for Leg 2 |
|---|---|---|---|---|---|---|---|
| Small | 16 | 610 | 78 | 26 | Cont | | X |
| | 18 | 481 | 67 | 24 | Bulk | X | X |
| | 31 | 565 | 85 | 32 | Tank | X | |
| | 32 | 530 | 85 | 33 | Tank | X | X |
| | 44 | 680 | 90 | 21 | Bulk | X | |
| Large | 60 | 744 | 104 | 23 | Bulk | X | |
| | 62 | 705 | 106 | 22 | Bulk | X | X |
| | 63 | 698 | 105 | 20 | Bulk | X | X |
| | 67 | 793 | 106 | 23 | Bulk | | X |
| | 70 | 800 | 105 | 24 | Bulk | X | X |

TABLE 2. CHESAPEAKE BAY CONDITIONS SIMULATED

| Scenario | Current | | Wind | | Speed (kts) | Size (1,000 dwt) | Day/Night |
|---|---|---|---|---|---|---|---|
| | Vel (kts) | Dir (deg) | Vel (kts) | Dir (deg) | | | |
| 5 | Slack | NA | 10 | 320 | 14 | 30 | Day |
| 6 | Slack | NA | 10 | 320 | 14 | 80 | Day |

engine order commands as well as discussions with the pilots. The experimental design and procedures used are discussed in the Chesapeake Bay Draft Principal Findings Report.[18]

### 2.1.4 Combined Data Analysis

As the last step in the validation process the at-sea data and simulator data were compared. Since ship size was the only variable to affect performance, it received the greatest emphasis in the combined data analyses. The ship size comparison was made in two ways. First an evaluation was made to compare the simulator data with the at-sea data for each ship size. Secondly, the at-sea group differences in performance between the large ship group and small ship group was compared to the simulator group differences. These results are discussed in detail in the Chesapeake Bay Draft Principal Findings Report.[19]

## 2.2 VALIDATION OF SHIP SIZE EFFECTS

The following discussion presents an overview of the results of the ship size validation. The goal in this section is to present a more quantitative evaluation of the results described in an earlier Chesapeake Bay Validation Report.[20]

### 2.2.1 Evaluation of the 30,000 dwt Tanker

Pilot performance with the smaller ships at sea and 30,000 dwt ship on the simulator is shown by Figure 4. The ordinates on the plots are the crosstrack position in the 800-foot channel, the abscissas are 475 feet of alongtrack distance. The data plotted are the crosstrack mean and standard deviation of a set of transits, four for the at-sea data, eight for the simulator data. At-sea performance is indicated by a solid line, simulator performance by a dotted line.

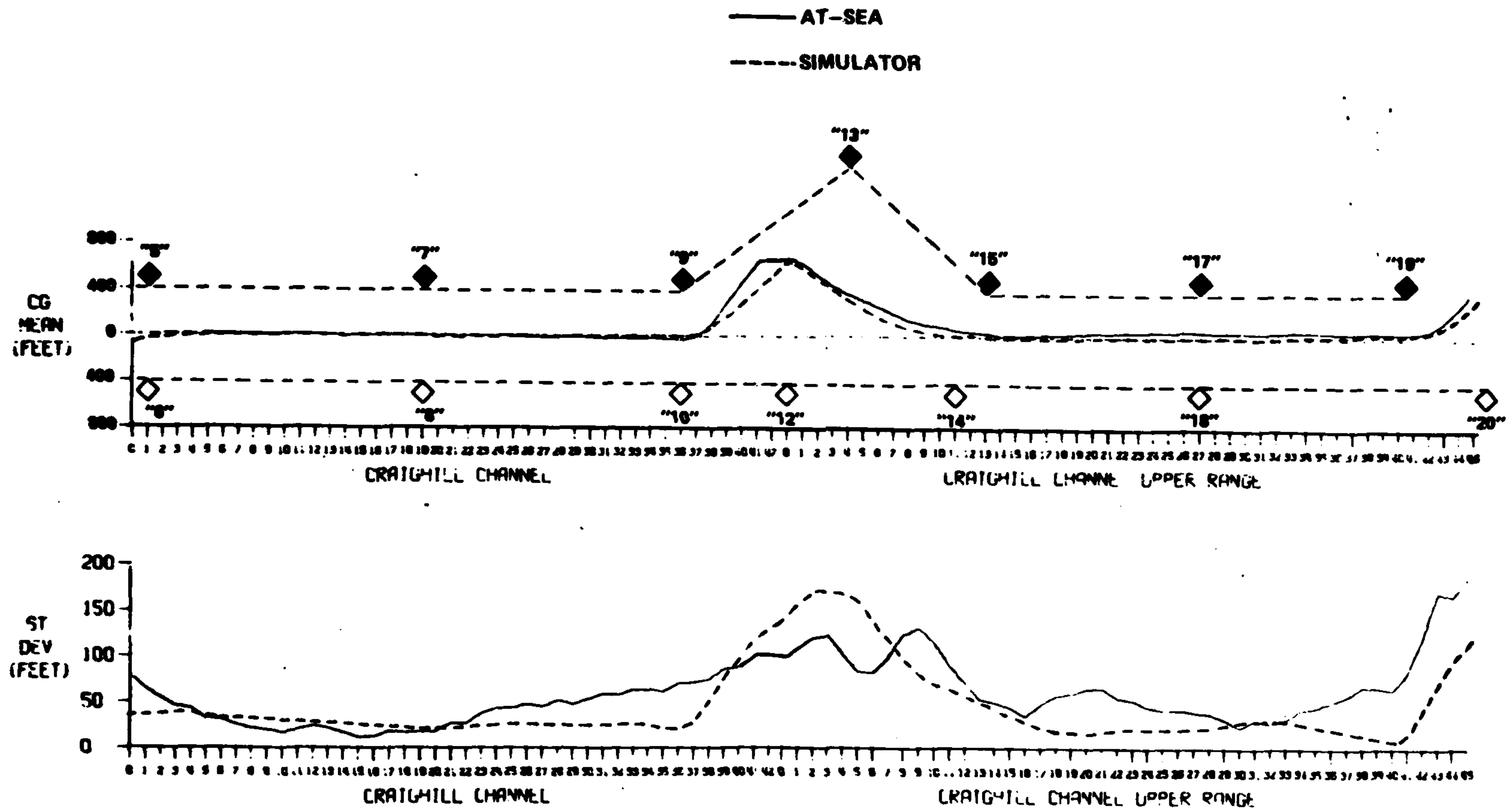
In the Craighill Channel, there is a fairly high degree of similarity between the performance measured on the simulator and performance measured in Chesapeake Bay. The similarity between the real world and the simulator was greatest near the center trackkeeping region and deteriorated, the closer the ship was near the turn. Table 3 shows the means and standard deviations for performance of both the simulator 30,000 dwt tanker group and the small ship group. The differences between the means were tested for significance using a t-test; the differences between the standard deviations were tested using an F test. These procedures are discussed further in Appendix F. The degree of similarity (or difference) between conditions is indicated by the closeness of the respective means and standard deviations in the simulator group to the at-sea group, and the lack of statistical

---

[18] G.E. Grant, J.D. Moynehan and M.W. Smith. "Aids to Navigation Validation Draft Principal Findings Report: Simulation I: Chesapeake Bay." U.S. Coast Guard, Washington, D.C., February 1983.

[19] Ibid.

[20] Ibid.



**BUOYS ARE POSITIONED FOR ILLUSTRATION AND MAY NOT APPEAR IN EXACT CHARTED LOCATION**

Figure 4. Chesapeake Bay Small Ship Comparison: Craighill Channel and Craighill Channel Upper Range

## TABLE 3. PERFORMANCE FOR SMALL SHIPS AT SEA AND THE 30,000 DWT TANKER ON THE SIMULATOR

| | | Craighill Channel Data Lines[1] | | | Upper Range Data Lines[1] | | |
|---|---|---|---|---|---|---|---|
| | Condition | 6-10 | 15-21 | 32-36 | 16-19 | 25-29 | 38-41 |
| Mean[2] | At sea | 5.37L | 3.52L | 27.50R | 48.63L[4] | 94.67L[4] | 110.75L[4] |
| | Simulator | 3.84R | 7.10R | 19.47R | 38.60R | 27.50R | 4.40R |
| Standard[3] Deviation | At sea | 23.96 | 23.88 | 63.30 | 49.38 | 40.92 | 69.35 |
| | Simulator | 32.71 | 17.19 | 26.10 | 23.52 | 26.87 | 5.86 |

[1] Performance was averaged over several data lines that seemed to represent a constant performance. (The degrees of freedom were not changed.)

[2] Means are expressed as feet to right or left of centerline. Differences were tested by t-test. Arrows indicate differences significant at $p \leq 0.10$.

[3] Standard deviations are expressed in feet. Differences were tested by F-test. Arrows indicate differences significant at $p \leq 0.10$.

[4] Means were not tested for difference because at-sea displacement was attributed to electronic tracking error.

significance of difference between the groups for much of the run. In the Craighill Channel, the only trackkeeping region in which there was any statistical significance was in the approach to the turn.

In the Craighill Channel Upper Range, the crosstrack mean was not evaluated because of problems with the electronic tracking equipment that skewed the results of the at-sea data. The crosstrack deviation, however, was not affected by this problem. Performance in the Upper Range shows trends similar to those found in the Craighill Channel. The only statistically significant differences occur, again, in the approach to the turn.

The differences in performance, in the approach to the turn may reflect the degree to which the pilot followed the experimental instructions. The pilots were instructed to maneuver the ship on the channel centerline, except in the turn where they could maneuver as they wished. The point at which the pilot left the channel center and began the turn was determined by the pilot. The pilots in the simulator condition stayed near the channel center longer than the pilots in the at-sea condition, before initiating the turn. At sea, the pilots are more conservative where the possibility of an error has more tangible consequences (i.e., vessel grounding) so they initiate the turn earlier than in the simulator condition. They attend less to the experimental instructions and more to the environmental constraints imposed by the real world while the pilots in the simulator condition are able to follow the experimental instruction more closely without worrying about the real world consequences.

The lack of significant difference between the means and standard deviations of performance data taken at sea and on the simulator suggests that the two sets of data are similar, but the lack of difference is not a conservative measure of sameness. The conventional use of the t-test is to provide a conservative measure of difference. A consideration of both alpha and beta error allows a conservative measure of sameness. It requires the specification of how large a difference between the two means will be accepted as "same". Then, it is possible to calculate the probability that the means are the same to that criterion. A summary of such an analysis for those parts of the Chesapeake Bay transit that showed no significance of difference appears in Table 4. As an example, the first entry in the table indicates that for data lines 6-10, if 30 feet is accepted as same, the probability that that criterion is reached is 0.68. If 50 feet is accepted as same, the probability that criterion is reached is 0.88. The procedure sets an upper limit of 0.88 to the probability than can be achieved for AN's data. Therefore, the degree of sameness is measured by how few or how many feet must be accepted as same to reach that maximum. In Table 4 all portions of the transit tested reach that maximum by 60 feet, or 7.5 percent of the 800-foot channel width. The statistical procedure used is discussed further in Appendix F.

### 2.2.2 Evaluation of the 80,000 dwt Tanker

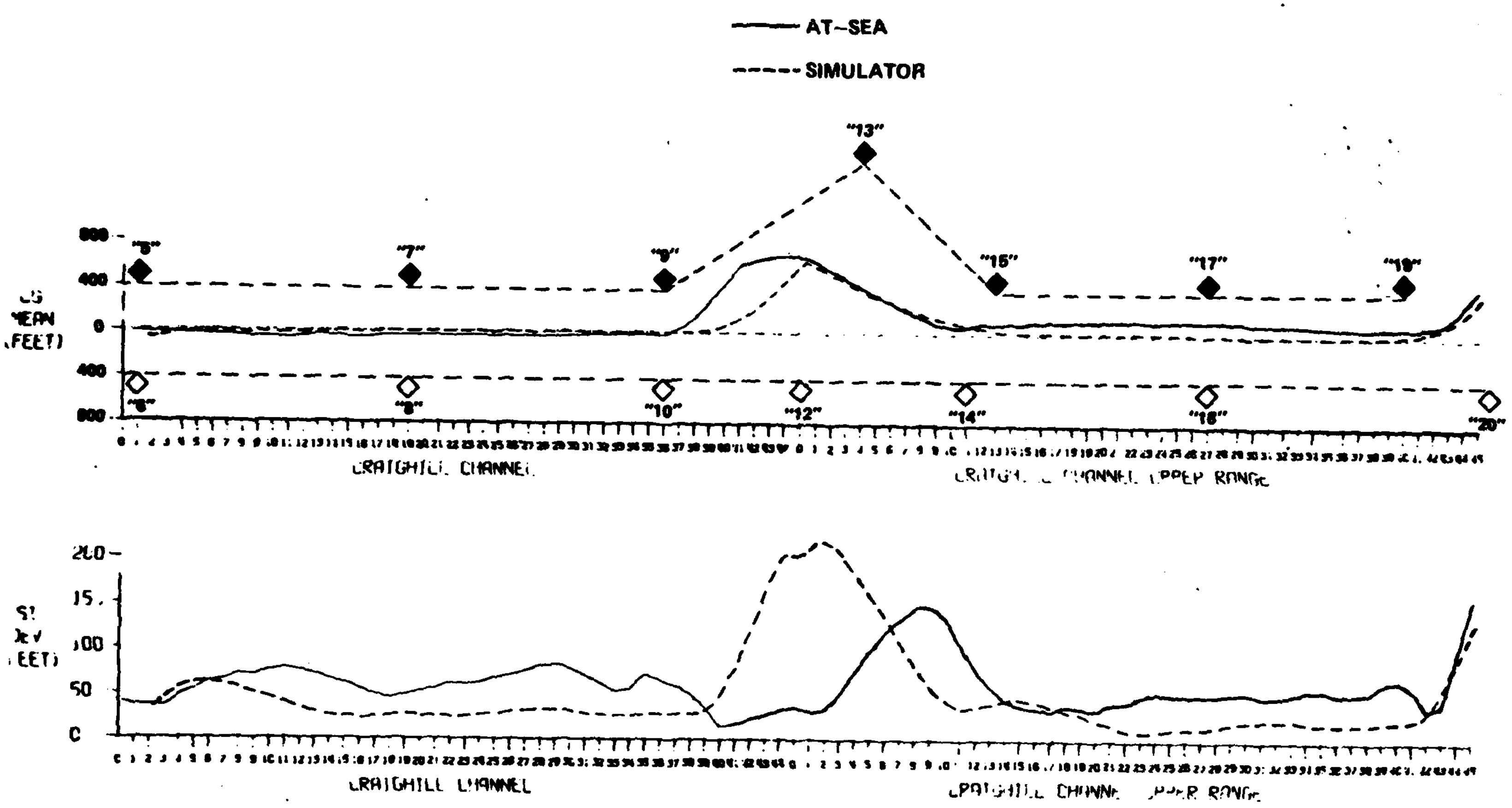
Pilot performance with large ships at sea and the 80,000 dwt ship on the simulator is shown by Figure 5. The mean and standard deviation is shown by a solid line for the at-sea data and by a dotted line for the simulator data.

## TABLE 4. PROBABILITY THE CROSSTRACK MEANS IN THE SIMULATOR AND AT-SEA CONDITIONS ARE THE SAME FOR BOTH SHIP SIZES

| | Small Ships and 30,000 dwt Tanker | | | | | | Large Ships and 80,000 dwt Tanker | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Craighill Channel | | | Upper Range[2] | | | Craighill Channel | | | Upper Range[2] | | |
| | Data Lines | | | | | | Data Lines | | | | | |
| $M_1$-$M_2$[1] | 6-10 | 15-21 | 32-36 | 16-19 | 5-29 | 38-41 | 6-10 | 15-21 | 32-36 | 16-19 | 5-29 | 38-41 |
| 20 | -- | 0.56 | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| 30 | 0.68 | 0.88 | 0.30 | 0.54 | 0.58 | 0.52 | 0.18 | 0.58 | 0.24 | 0.52 | 0.52 | 0.53 |
| 40 | 0.876 | 0.88 | 0.70 | 0.84 | 0.84 | 0.82 | 0.64 | 0.84 | 0.68 | 0.82 | 0.84 | 0.52 |
| 50 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.84 | 0.88 | 0.86 | 0.88 | 0.88 | 0.84 |
| 60 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 |
| 70 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 |
| 80 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 |

[1] $M_1$-$M_2$ represents the difference in feet between crosstrack means.

[2] The displacement due to electronic tracking error in the Craighill Channel Upper Range was corrected for in this analysis.



**BUOYS ARE POSITIONED FOR ILLUSTRATION AND MAY NOT APPEAR IN EXACT CHARTED LOCATION**

Figure 5. Chesapeake Bay Large Ship Comparison: Craighill Channel and Craighill Channel Upper Range

For the large ship comparison the performance patterns are similar to those found in the small ship comparison. The similarity between the large ship group and the 80,000 dwt tanker is again greatest in the center trackkeeping region. The similarity is least in the region approaching the turn. In comparison to the small ship evaluation, the magnitude of the differences between the at-sea conditions and simulator conditions is greater for the large ship. The difference in the means and standard deviation shown in Table 5 between the at-sea conditions and simulator conditions is larger for all regions of the channel. The analysis of degree of sameness also reflects the greater dissimilarity between conditions for the large ship as shown in Table 4. When the difference between the crosstrack means is 50 feet or less, the probability that they are the same is always lower for the large ship comparison. The lower probability for the large ship compared to the small ship suggests that the 30,000 dwt tanker model approximates the group of small ships tracked more closely than the 80,000 dwt tanker model approximates the group of large ships tracked.

## 2.3 SUMMARY

The quantitative analysis supports the qualitative trends reported in an earlier report.[21] Both ships exhibit performance in the trackkeeping regions that is characteristic of performance in the corresponding regions at sea. Performance is most similar in the center of the trackkeeping regions and less similar near the turns.

The performance of a group of small ships was well matched by that with the simulated 30,000 dwt tanker. However, the performance of a group of large ships was not as well matched by that with the simulated 80,000 dwt tanker. While trackkeeping on the range, the 80,000 dwt tanker showed a precision as good as that of the 30,000 dwt tanker, while at sea the larger ships showed a larger crosstrack error and larger standard deviation than did the smaller ships. If a limitation of the simulation was identified by the Chesapeake Bay validation, it would be that the 80,000 dwt tanker was too accurate when trackkeeping. This is presumably the result of a lack of random movement in the ship model during trackkeeping. However, there are only very limited consequences to the 80,000 dwt tankers' precision in trackkeeping for a simulation of performance in a specific waterway.

---

[21] Ibid.

## TABLE 5. PERFORMANCE FOR LARGE SHIPS AT SEA AND THE 80,000 DWT TANKER ON THE SIMULATOR

| | Condition | Craighill Channel Data Lines[1] | | | Upper Range Data Lines[1] | | |
|---|---|---|---|---|---|---|---|
| | | 6-10 | 15-21 | 32-36 | 16-19 | 25-29 | 38-41 |
| Mean[2] | At sea | 51.00R | 46.63R | 19.43R | 14.14L[4] | 162.65L[4] | 143.83L[4] |
| | Simulator | 3.66R | 7.81R | 5.77R | 18.89R | 22.05R | 1.66R |
| Standard[3] Deviation | At sea | 54.30 | 52.62 | 61.97 | 36.16 | 51.30 | 61.06 |
| | Simulator | 36.44 | 21.60 | 28.80 | 34.56 | 24.49 | 14.90 |

[1] Performance was averaged over several data lines that seemed to represent a constant performance. (The degrees of freedom were not changed.)

[2] Means are expressed as feet to right or left of centerline. Differences were tested by t-test. Arrows indicate differences significant at $p \leq 0.10$.

[3] Standard deviations are expressed in feet. Differences were tested by F-test. Arrows indicate differences significant at $p \leq 0.10$.

[4] Means were not tested for difference because at-sea displacement was attributed to electronic tracking error.

# Section 3

# NARRAGANSETT BAY: AT SEA AND ON THE SIMULATOR

## 3.1 INTRODUCTION

Two types of data, at sea and on the simulator, were collected in the Narragansett Bay. As outlined in the "At-Sea Data Collection Plan,"[22] the Narragansett Bay was selected because: (1) it was a buoyed channel with some markings that did not adhere to the SRA/RA manual recommendations; (2) 30,000 dwt midship bridge tankers, the same type of ship modeled in simulator experiments, made up a substantial portion of bay traffic; (3) it was close to Ship Analytics (SA) so ship transits were accessible; and (4) there was a working relationship with the pilots from Northeast Marine Pilots, Inc. since they participated in previous simulator experiments.

The methodology for the Narragansett Bay validation is illustrated in Figure 6. The Narragansett comparison differs in one respect from the Chesapeake. The comparison variable is day versus night performance whereas in the Chesapeake experiment it was large ship versus small ship.

### 3.1.1 At-Sea Data Collection Methodology

For the at-sea transits an Internav 404 Loran C receiver and an HP9915 computer for recording were supplied by the Coast Guard for data collection. The Coast Guard charted a waypoint survey to set reference points for the data collection and reduction. They established a station in Bristol, Rhode Island to continuously monitor the Loran C time differential (TD) signals to estimate the local distortion between Loran C grid position and actual geographic position. A data analysis program and facilities to use the program were provided at the Coast Guard's Research and Development Center in Groton, Connecticut.

About 50 minutes after boarding the tanker, the ship was abeam buoy "1" in Upper Narragansett Bay (see Figure 7) and the data collection program was initialized. Data were taken every 15 seconds. During the transit the pilot's helm and engine orders were recorded manually by the SA data-takers. The data collection ended at buoy "33" (see Figure 8) as the tanker met the tugs. For more details see the report listed entitled "At-Sea Preimplementation Draft Principal Findings."[23]

### 3.1.2 At-Sea Data Selection

In a real world data collection, problems are not unexpected and several occurred when collecting data in the Narragansett Bay. Thirteen ships were

---

[22] J.W. Gynther and R.B. Cooper. "At-Sea Data Collection Plan for Prototype Implementation of Aid System Design Guidelines." U.S. Coast Guard, Washington, D.C., April 1982.

[23] J.D. Moynehan. "Aids to Navigation Implementation At-Sea Preimplementation Draft Principal Findings." U.S. Coast Guard, Washington, D.C., September 1983.

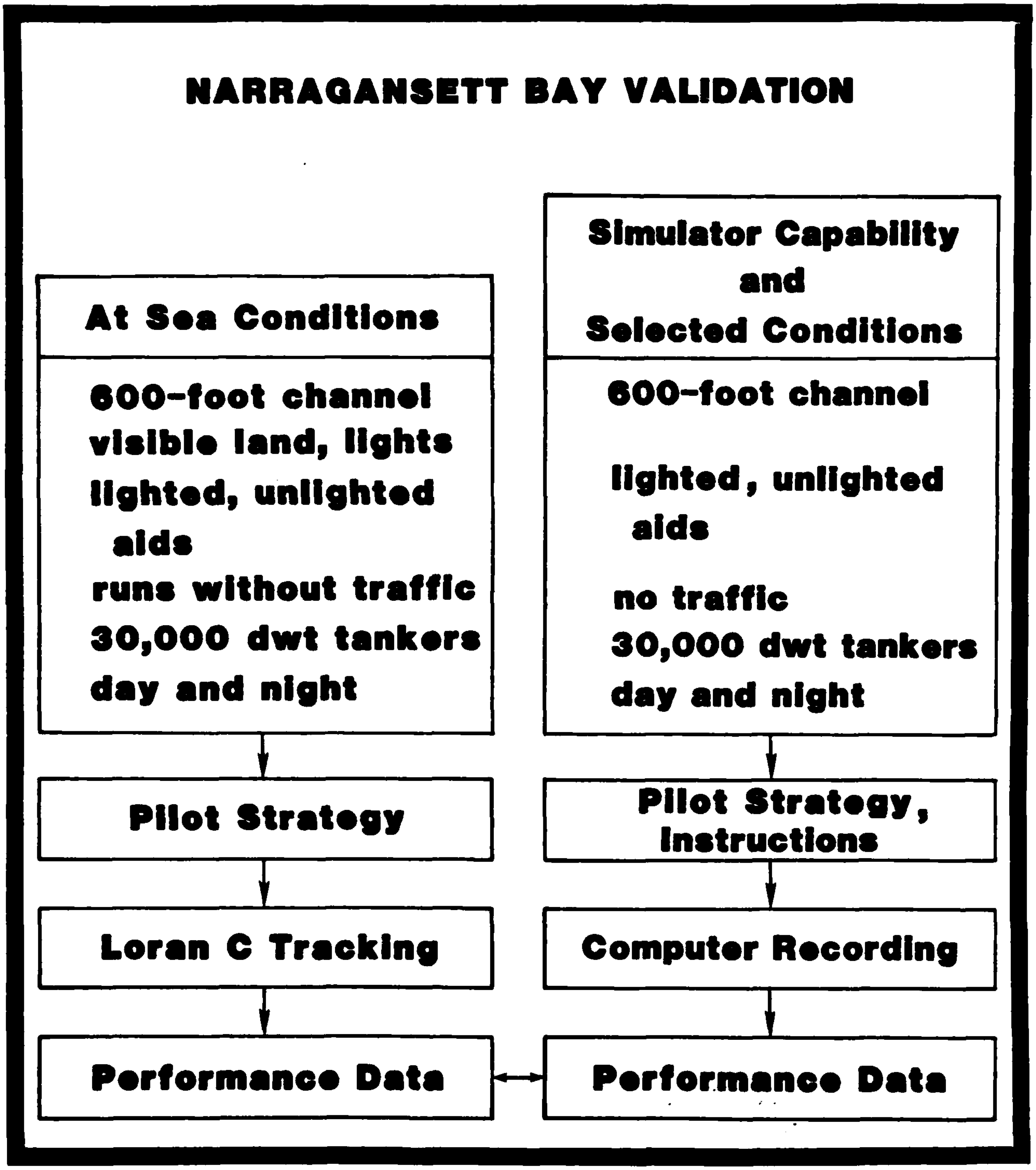


Figure 6. The Validation Methodology for Narragansett Bay

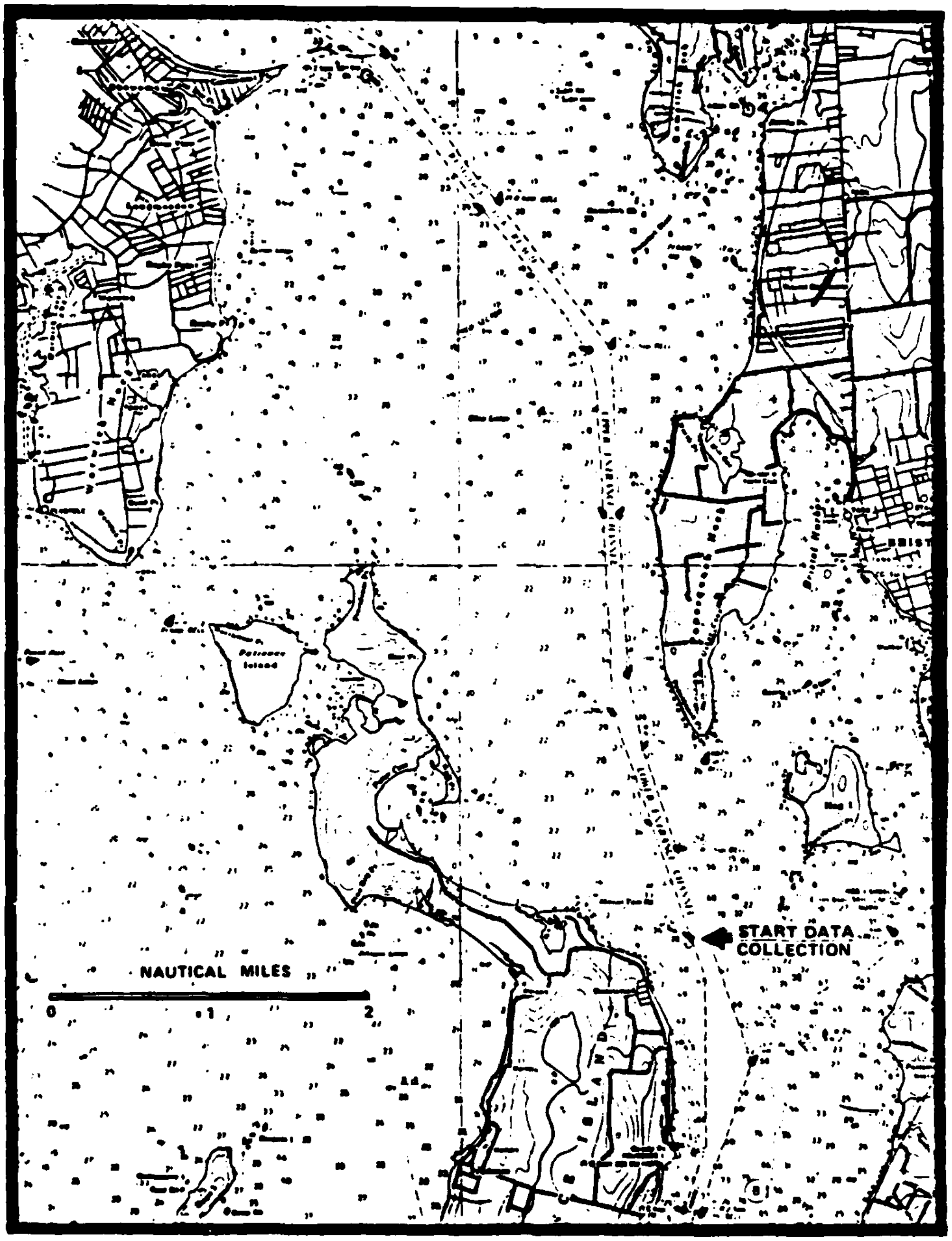


Figure 7. Narragansett Bay Where At-Sea Data Was Collected
Chart No. 13221 of March 28, 1981

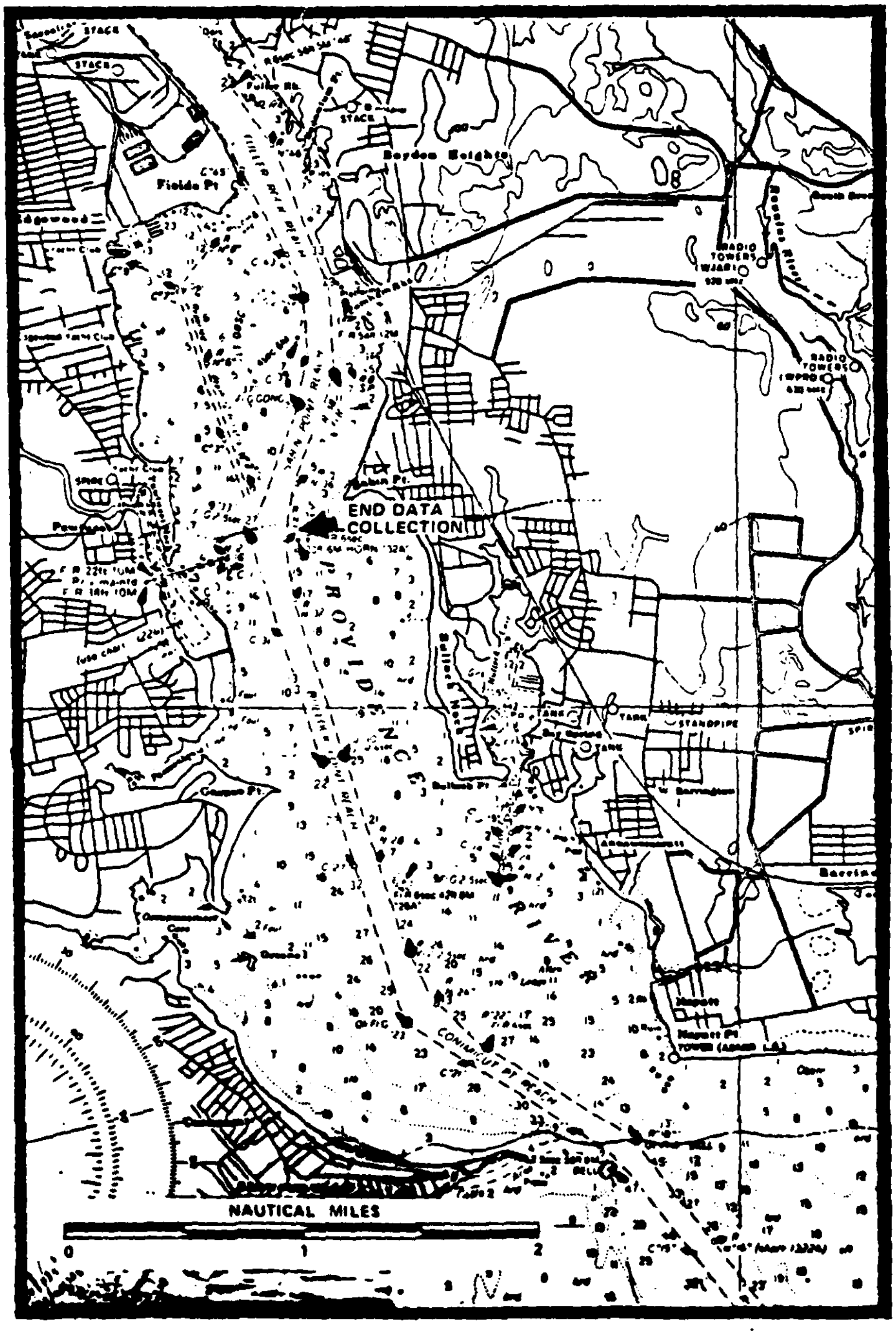


Figure 8. Narragansett Bay Where at Sea Data Was Collected
Chart No. 13221 of March 28, 1981 (Continued)

ridden and of these eight runs were acceptable with four runs comprising the daytime data and four runs comprising the nighttime data. A portion of one nighttime run, however, was lost due to a traffic ship interfering with the Loran signal. The five remaining transits were lost due to system problems or failures. In addition, the frequency of 30,000 dwt tankers which SA rode was lower than anticipated averaging only two per month. This prolonged the task through the winter when severe weather precluded riding four available tankers.

Beyond these logistical and equipment problems, there is the issue of data accuracy and the possibility of a tracking error. Inherent accuracy shortcomings occur with the Loran C signal and recording equipment which are generally not as accurate as the data obtained on the simulator. With optimal atmospheric conditions, which most likely will not occur, accuracy within 10 feet is the best which can be expected from the Loran receiver. See Appendix A of the Preimplementation Draft Principal Findings Report for further details.[24]

### 3.1.3 At-Sea Data Analysis

The raw TD data of each transit was broken down into its component legs by waypoint and trackpoint using a Coast Guard supplied program on the HP9845. The TD corrections were applied to each leg in 15-minute increments. The TD data is referenced against the waypoint survey by the program to produce alongtrack and crosstrack positions which are then converted to an X-Y grid. These X-Y values are then reduced to a format compatible with SA's computer system, broken into legs of each at-sea run.

These legs are then combined, according to night and day runs, to produce the composite plots contained in Appendix E. A mathematical smoothing function is incorporated in these composite plots to separate the "noise" deviation from the total standard deviation. The standard deviation remaining is assumed to be the piloted variation, which is the performance measure of interest. (See the draft At-Sea Preimplementation Principal Findings Report for more details.)

### 3.1.4 Simulator Data Collection

For this second validation experiment, simulator data was collected under day and night conditions on the USCG/SA simulator, similar to the at-sea data collection conducted on the Upper Narragansett Bay. Experiment planning and preliminary data analysis are discussed in separate reports.[25,26] Eight pilots from the Northeast Marine Pilots, Inc. (the pilot group observed at sea in Narragansett Bay) ran through the simulations.

---

[24] Ibid.

[25] G.E. Grant and M.W. Smith. "Aids to Navigation Presimulation Report for Validation: Validation for a Simulator-Based Design Project." U.S. Coast Guard, Washington, D.C., September 1982.

[26] G. Grant and J. Moynehan. "Aids to Navigation II/Implementation: Preliminary Observations and Data Analysis, U.S. Coast Guard, Washington, D.C., May 1983.

The criterion task (transit close to the centerline at a realistic speed using preferred strategy in the turns) and the primary performance measure (ship tracks) were the same at sea and on the simulator.

### 3.1.5 Combined Data Analysis

The following at-sea and simulator data comparisons are available: day for at-sea versus simulator, night for at-sea versus simulator, and the effect of day/night at sea versus simulator.

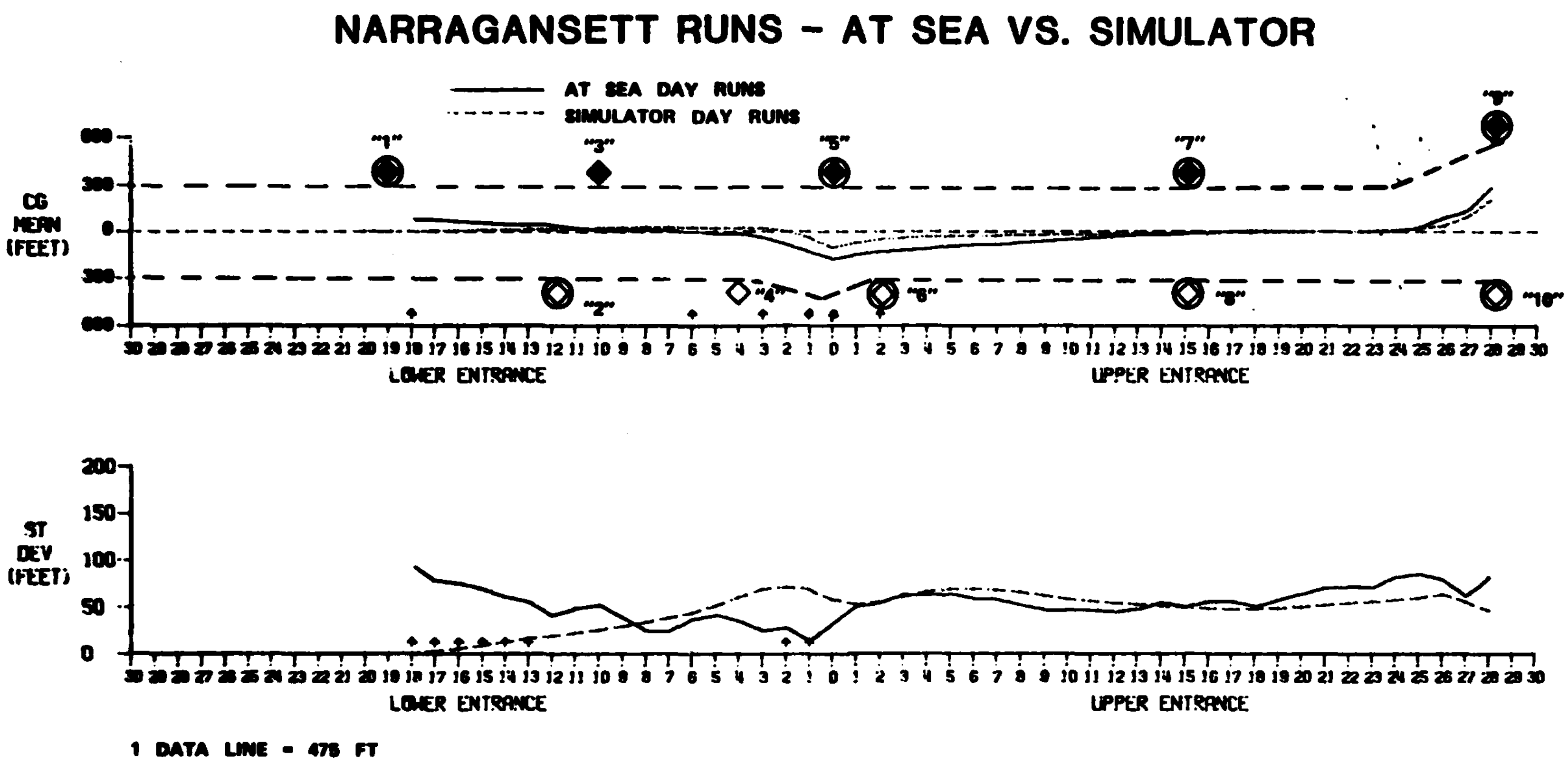
## 3.2 DAYTIME AT SEA AND ON THE SIMULATOR

The daytime Narragansett Bay at-sea and simulator data have some striking similarities, but also contain some strategy differences. At sea, pilots generally emphasized the turns by starting early, moving to the inside of the cutoffs (the area widened by dredging), and returning to the centerline late. One turn at sea was not marked to indicate the cutoff, discouraging the pilots from using this strategy. On the simulator, pilots emphasized the centerline by staying on it longer, staying closer to the centerline in the turn, and coming back on the centerline earlier in the new leg making little use of the space available in the cutoff. In the one turn not marked to indicate the cutoff, the at-sea and simulator strategies converged and performance was similar. In the straight legs of the channel, the at sea and simulator tracks appear identical.

The Narragansett Bay data was collected from buoy "1" in the Lower Entrance to buoy "33" in the Bullock Point Reach (see charts in Figures 7 and 8). Daytime pilot performance in the Narragansett Bay is shown by Figures 9, 10, and 11. The mean and standard deviation is shown by a solid line for the at-sea data and by a dotted line for the simulation data. The arrows on the plots identify areas of statistical difference resulting from F and t-tests at a 0.10 level of significance. (See Appendix E for more detailed track plots and Appendix F for a discussion of statistical tests used to analyze the data.)

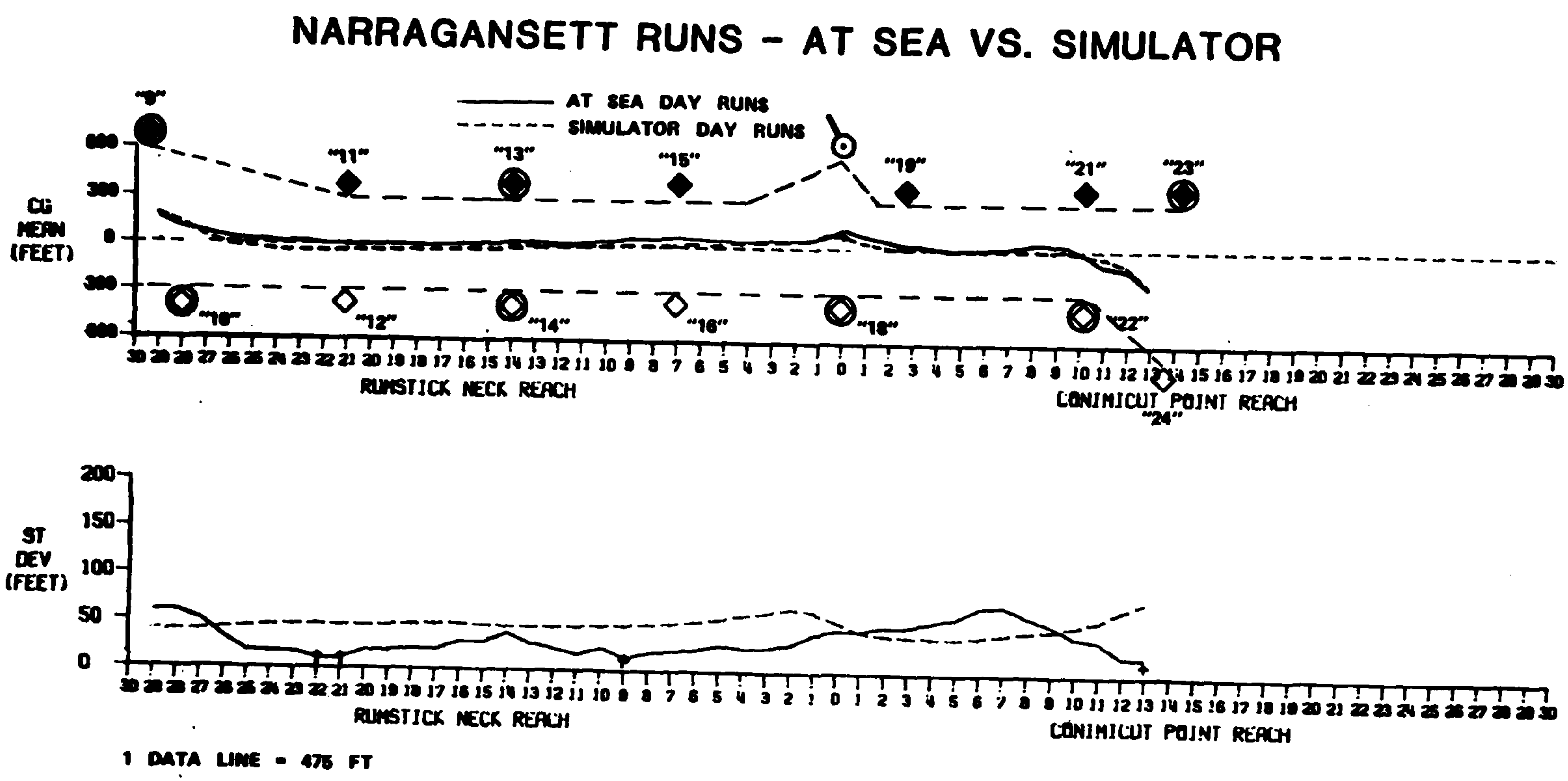
In the Lower Entrance, shown by Figure 9, the statistical differences between tracks in the first 0.4 nm of the run are not meaningful due to different start points of the transits. At sea the ships have just made a left turn in an extremely wide cutoff, and on the simulator the ships start at the same point. When preparing for the turn off Popasquash Point, there are interesting differences between tracks. The mean tracks show that at sea pilots ease through the leg gradually, sliding through the centerline to take a well-marked, slight cutoff early and inside; on simulator they are more rigorous about maintaining centerline, turn later, and stay close to the center of turn. The tendency is the same to turn inside, however, the degree is less. The standard deviation indicates that at sea pilots make the Popasquash turn similarly, while on the simulator the standard deviation increases as pilots approach and negotiate the turn differently (some take the turn inside, and some stay left in an effort to maintain the centerline as directed).

In the Upper Entrance also shown by Figure 9, once the pilots recover from the Popasquash Point turn there are no statistical differences in



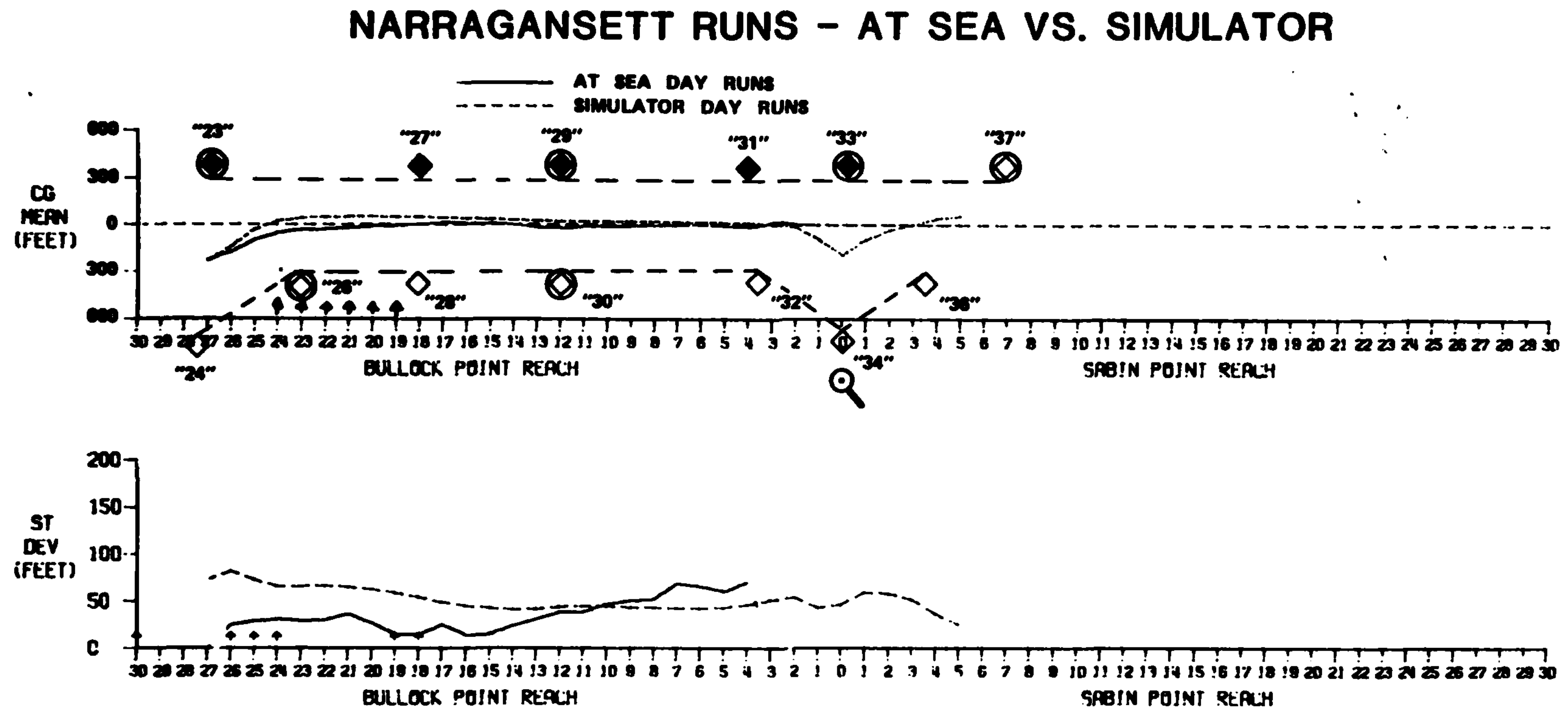
BUOYS ARE POSITIONED FOR ILLUSTRATION AND MAY NOT APPEAR IN EXACT CHARTED LOCATION

Figure 9. Narragansett Bay Daytime Comparison: Lower Entrance and Upper Entrance



BUOYS ARE POSITIONED FOR ILLUSTRATION AND MAY NOT APPEAR IN EXACT CHARTED LOCATION

Figure 10. Narragansett Bay Daytime Comparison: Rumstick Neck and Conimicut Point



1 DATA LINE = 475 FT

**BUOYS ARE POSITIONED FOR ILLUSTRATION AND MAY NOT APPEAR IN EXACT CHARTED LOCATION**

Figure 11. Narragansett Bay Daytime Comparison: Bullock Point Reach

piloted performance at the 0.10 level of significance. At sea pilots are more lenient in returning to the centerline of this relatively easy leg. At sea and on the simulator, the pilots use the first set of gates located 1.2 nm past the turn to return to the centerline. Both at sea and simulator means are within 30 feet of the centerline at data line 11 or 0.85 nm past the turn buoy "5". From this point through the rest of the straight leg (data lines 11 through 24), the mean tracks appear identical. Both alpha and beta error were analyzed to determine the probability that the tracks are the same. (See Appendix F for a discussion of the analysis.) Table 6 shows the outcome of this analysis for selected data lines. For example, if a 40-foot difference is acceptable as the same, then at data line 11 the probability is 0.20 that they are the same. It reaches maximum probability of 0.88 that the means are the same at 80 feet. The standard deviation in the Upper Entrance is approximately 50 feet. This fairly high standard deviation reflects the light marking of the leg, there is only one gate (1.2 nm past the turn) marking over 2.2 nm, both standard deviations dip on the approach to this gate at data line 15. Although the standard deviations appear similar, there is no corresponding procedure for finding the probability that they are the same. The standard deviations rise as the ship approaches the 35-degree turn onto Rumstick Neck Reach.

The turn from Upper Entrance to Rumstick Neck Reach is dredged as a cutout turn but not marked as one. The arrangement is illustrated in Figure 7. Rather than the three buoys that would be needed to outline the cutoff area, the turn is marked with two buoys, arranged as a gate across the channel. The uncertainty of the outline of the turn discourages the pilots from taking advantage of it at sea. The mean tracks for at-sea and simulator performance are very similar both in the entrance to the turn in Figure 9 and the exit from it in Figure 10. The mean tracks in the turn pullout at data line 25 are the same within 30 feet with a probability of 0.54, within 40 feet with a probability of 0.80, and within 50 feet with a probability of 0.88. The standard deviation is slightly higher at sea, but not significantly so, suggesting that some pilots did make some use of the cutoff. The probability analysis is presented in Table 6.

In Rumstick Neck Reach, shown by Figure 10, the mean tracks are similar and are generally within 30 feet of the centerline. This flat mean is a result of the high buoy density. The mean shows a uniform strategy of coming slowly left through the leg to take the Conimicut Point turn inside. The results of the beta analysis is presented in Table 6. Data line 25 represents the turn pullout, data line 13 represents trackkeeping, and data lines 3 and 1 representing turn initiation and turn maneuver. The probability that the means are the same is at its maximum at 0.88. The standard deviations of the tracks are different. The at-sea standard deviation is high exiting the turn and entering Rumstick Neck Reach (60 feet at data line 28), it dips as the ship passes through the unlighted gated buoys (at data line 21). Here the at-sea standard deviation is significantly lower than the simulator (10 feet compared to 47 feet). The simulator standard deviation is high but flat throughout the leg, rising into the Conimicut Point turn. The flat standard deviation represents the balance between pilots who took the leg slightly to the right and left of centerline. The lower standard deviation at sea throughout Rumstick Neck Reach shows rigorous attention to the task in the real world; the rise and

## TABLE 6. PROBABILITIES THE CROSSTRACK MEANS IN THE SIMULATOR AND AT-SEA CONDITIONS REFLECT THE SAME POPULATION FOR DAYTIME CONDITIONS

| $M_1$-$M_2$ in feet | Upper Entrance Data Lines | | | | Rumstick Neck Data Lines | | | |
|---|---|---|---|---|---|---|---|---|
| | 11 | 17 | 22 | 27 | 25 | 13 | 3 | 1 |
| 20 | -- | -- | -- | -- | -- | -- | -- | -- |
| 30 | -- | -- | -- | -- | 0.54 | 0.16 | -- | -- |
| 40 | 0.20 | 0.30 | -- | 0.10 | 0.80 | 0.68 | 0.16 | 0.10 |
| 50 | 0.46 | 0.66 | 0.42 | 0.48 | 0.88 | 0.85 | 0.60 | 0.50 |
| 60 | 0.78 | 0.83 | 0.66 | 0.78 | 0.88 | 0.88 | 0.79 | 0.74 |
| 70 | 0.87 | 0.88 | 0.82 | 0.86 | 0.88 | 0.88 | 0.88 | 0.83 |
| 80 | 0.88 | 0.88 | 0.87 | 0.88 | 0.88 | 0.88 | 0.88 | 0.88 |

| $M_1$-$M_2$ in feet | Conimicut Point Data Lines | | | Bullock Point Data Lines |
|---|---|---|---|---|
| | 5 | 4 | 11 | 12 |
| 20 | -- | -- | -- | -- |
| 30 | 0.10 | -- | -- | 0.16 |
| 40 | 0.60 | 0.54 | 0.30 | 0.72 |
| 50 | 0.82 | 0.68 | 0.73 | 0.83 |
| 60 | 0.88 | 0.83 | 0.82 | 0.88 |
| 70 | 0.88 | 0.88 | 0.88 | 0.88 |
| 80 | 0.88 | 0.88 | 0.88 | 0.88 |

fall of the standard deviation shows the efficiency of the gates. The turn into Conimicut Point is made like that off Popasquash; at sea the pilots turn to the inside, while on the simulator the pilots maintain a centerline track longer before turn initiation. The lower standard deviation at sea at turn initiation, while not statistically significant, is due to a "canned" strategy, according to the pilots, of making an "S-shaped" turn.

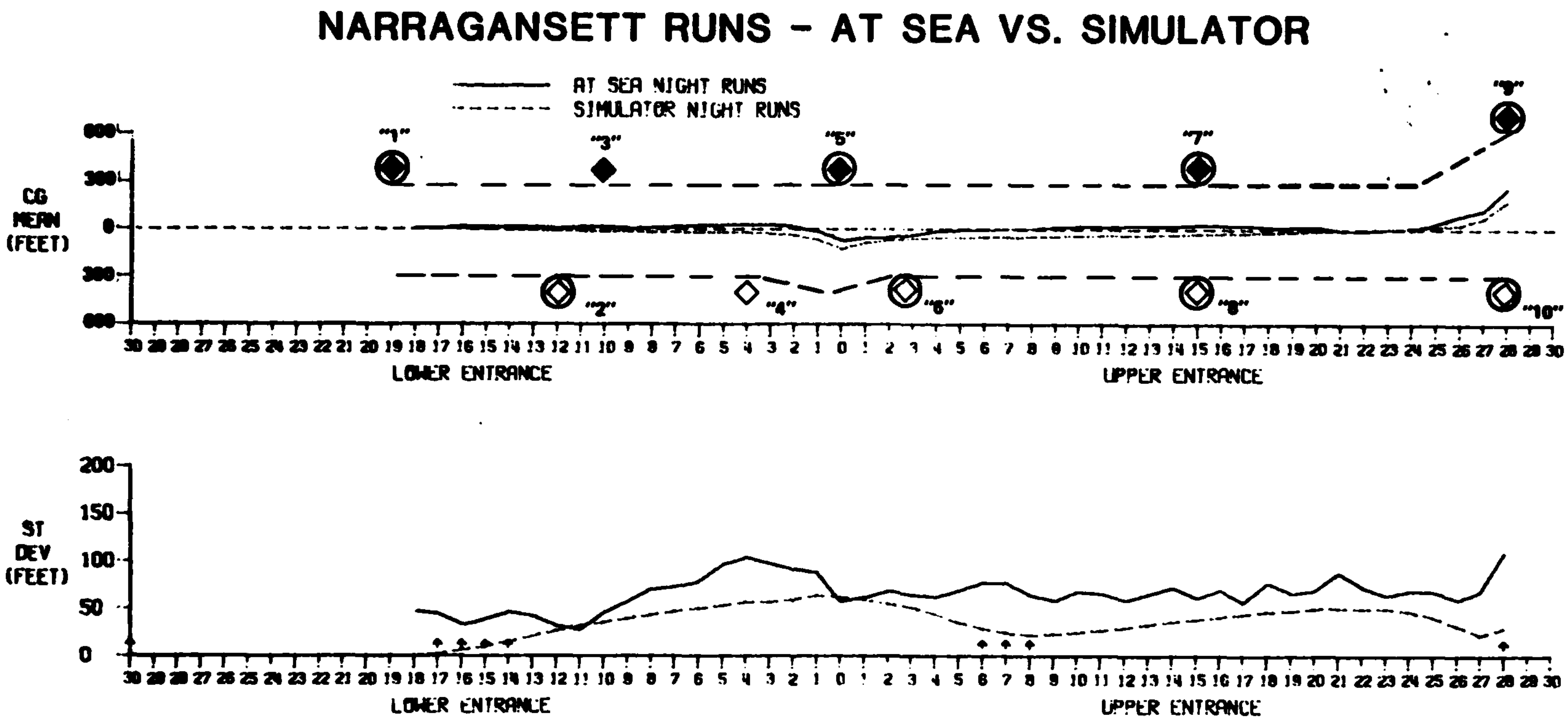
In Conimicut Point Reach, shown by Figure 10, the mean tracks appear identical. Table 6 with probability analysis results shows that even with a small difference in means of 40 feet, the probability is over 0.50 at data lines 4 and 8 the means are the same. There are no statistical differences in the standard deviation until turn initiation onto Bullock Point Reach. The standard deviation at sea is 19 feet as compared to 50 feet on the simulator. At sea, the turn strategy is more standardized, thus the lower standard deviation. At sea pilots treat the entire maneuver as an S-shaped turn because Conimicut Point Reach is under 1 nm (0.8 nm). The turn onto this leg is 18 degrees and the turn out of this leg is 40 degrees. On the simulator, the pilots' strategy is not as consistent, thus the high standard deviation.

In Bullock Point Reach, shown by Figure 11, at sea the pilots complete the turn later and the mean track reaches the centerline about 0.6 nm past the turn buoy (data line 20). On the simulator, the turn is made sharply and the mean track crosses the centerline about 0.3 nm past the turn buoy (data line 24). The significant differences between the means in the turn pullout region, before the first set of gated buoys, is due to an overshoot of the centerline of the simulator track. The high standard deviation of the simulator data in the turn pullout, 80 feet maximum as compared to 30 feet maximum, is due to differing strategies (some pilots take the turn as they normally would at sea and some try harder to maintain a centerline track).

## 3.3 NIGHTTIME AT SEA AND ON THE SIMULATOR

The nighttime Narragansett Bay at-sea and simulator data have some similarities and some differences as did the daytime data. Nighttime piloted performance in the Narragansett Bay is shown by Figures 12, 13, and 14. The arrows on the plots identify areas of statistical difference resulting from F and t-tests at a 0.10 level of significance. (See Appendix E for more detailed track plots and Appendix F for a discussion of statistical tests used to analyze the data.)

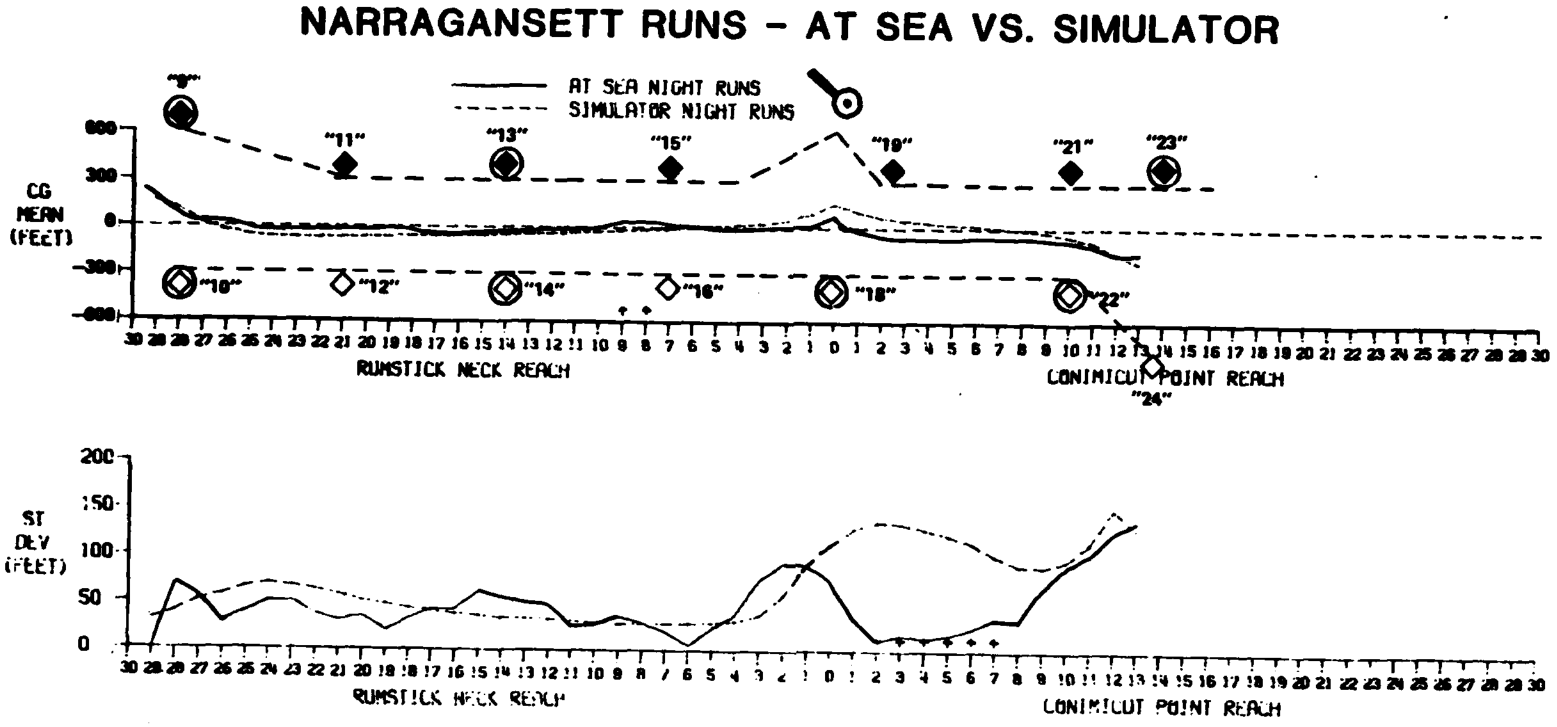
In the Lower Entrance, shown by Figure 12, there are no statistical differences between the means but there are differences between the standard deviation at the beginning of the leg. The at-sea tracks have more variation because pilots are entering a new leg and although the mean is on the centerline, all tracks are not necessarily at the same point as they are when the simulation began. In preparing for the turn off Popasquash Point, there are some differences in strategies, however, these are not statistically significant. At sea, the pilots react to unlighted buoy "4" by steering away from it, while on the simulator they maintain their centerline track. At sea the standard deviation rises to 100 feet compared to 60 feet on the simulator. This reflects pilot uncertainty of the



**1 DATA LINE = 475 FT**

**BUOYS ARE POSITIONED FOR ILLUSTRATION AND MAY NOT APPEAR IN EXACT CHARTED LOCATION**

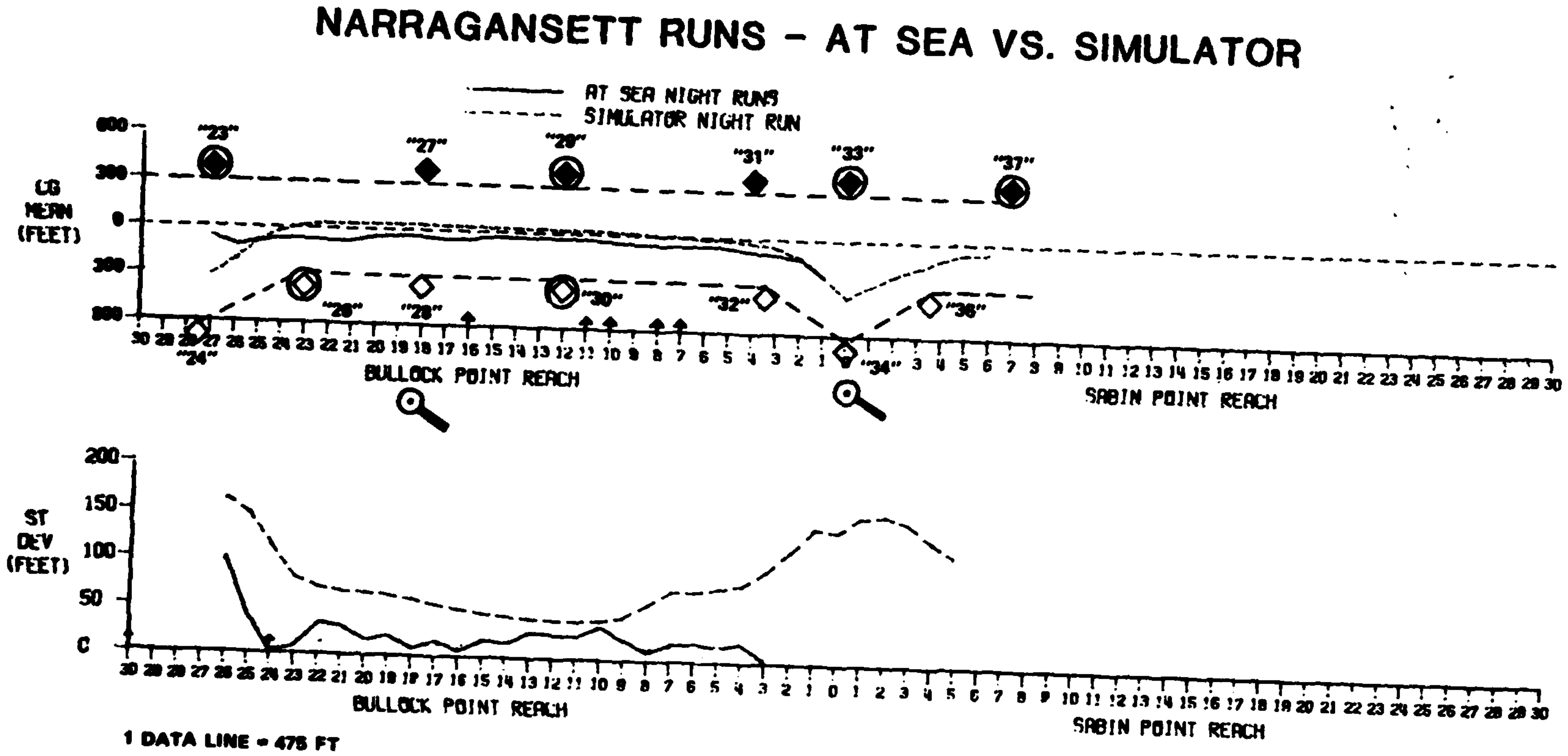
Figure 12. Narragansett Bay Nighttime Comparison: Lower Entrance and Upper Entrance



1 DATA LINE = 475 FT

**BUOYS ARE POSITIONED FOR ILLUSTRATION AND MAY NOT APPEAR IN EXACT CHARTED LOCATION**

Figure 13. Narragansett Bay Nighttime Comparison: Rumstick Neck and Conimicut Point Reaches



**BUOYS ARE POSITIONED FOR ILLUSTRATION AND MAY NOT APPEAR IN EXACT CHARTED LOCATION**

Figure 14. Narragansett Bay Nighttime Comparison: Bullock Point Reach

unlighted buoy position. Therefore, at sea the pilots turn closer to the centerline and on the simulator they turn further to the inside. The simulator data is more similar to daytime performance while the at-sea turn performance is more conservative. (See Section 3.4 for a discussion of day/night differences.)

In the Upper Entrance, also shown by Figure 12, the mean tracks are similar. The at-sea track reflects the pilots turn strategy so the centerline is reached and maintained sooner. The simulator track is slightly to the right of the centerline since the ship exited the turn farther right. An analysis was conducted to determine the probability that the tracks are the same. Resulting probabilities are listed in Table 7 for data lines 16 and 26. If a 40-foot difference is acceptable as the same at data line 16 the probability is 0.34 and at data line 26 the probability is 0.64 that the means are the same. The probability that the means are the same is at its maximum at 0.88. This occurs in 70 feet for data line 10 and 60 feet for data line 20. Overall, the standard deviation is larger at sea than it is on the simulator. Areas of statistical difference occur only as the simulator standard deviation decreases on exiting the Popasquash turn as they line up on lighted gated buoys "7" and "8" at data line 15. After passing through the gates, the standard deviation rises closer to the at-sea tracks. The standard deviations again become significantly different when the simulator dips due to a common turn technique onto Rumstick Neck Reach.

The comparison between at-sea and simulator performance in the turn from Upper Entrance to Rumstick Neck Reach shows a pattern at night similar to that in the daytime. The mean tracks are similar going into the turn in Figure 12 and coming out in Figure 13. The probability analysis is presented in Table 7. At data line 26 going into the turn, the probability that the means are the same within 50 feet is 0.84, at 60 feet is 0.88. At the turn pullout, represented by data line 25, the probability that the means are the same within 80 feet is 0.85. Performance is slightly less similar coming out of the turn than going into it. The nighttime mean suggests that the later initiation of the turn results in a later completion. As in the daytime, the standard deviation is higher at sea, but not significantly so, suggesting pilots did make some use of the cutoff.

In Rumstick Neck Reach, shown by Figure 13, the mean tracks appear similar through most of the leg. Probabilities listed in Table 7 show that if 50 feet is an acceptable difference in tracks, there is a probability of 78 percent at data line 13 and 0.52 at data line 3 that the tracks are the same. The maximum probability that the means are the same is at 0.88. This occurs in 70 feet for data lines 13 and 3. The only area of statistical difference occurs immediately before gated buoys 15 and 16 (at data lines 9 and 8). At-sea tracks are 35 feet to the left of the centerline as compared to 15 feet to the right of the centerline on the simulator. There is a strategy difference at sea versus the simulator. One explanation is that at sea the pilots center on Conimicut Point light and buoy "18" while on the simulator the light was not as heavily relied upon. The at-sea and simulator standard deviation in Rumstick Neck Reach are similar. As the ships approach the set of lighted buoys "13" and "14", the standard deviation tends to drop. It begins to rise when the ship nears the end of the leg and the pilots prepare for the turn onto Conimicut Point.

TABLE 7. PROBABILITIES THE CROSSTRACK MEANS IN THE SIMULATOR AND AT-SEA CONDITIONS REFLECT THE SAME POPULATION FOR NIGHTTIME CONDITIONS

| $M_1$-$M_2$ in feet | Upper Entrance Data Lines |  | Rumstick Neck Data Lines |  |  |
|---|---|---|---|---|---|
|  | 16 | 26 | 25 | 13 | 3 |
| 20 | -- | -- | -- | -- | -- |
| 30 | -- | 0.20 | -- | -- | -- |
| 40 | 0.34 | 0.64 | -- | 0.46 | 0.18 |
| 50 | 0.68 | 0.84 | 0.04 | 0.76 | 0.52 |
| 60 | 0.83 | 0.88 | 0.36 | 0.87 | 0.87 |
| 70 | 0.88 | 0.88 | 0.60 | 0.88 | 0.88 |
| 80 | 0.88 | 0.88 | 0.85 | 0.88 | 0.88 |

| $M_1$-$M_2$ in feet | Conimicut Point Data Lines |  |  | Bullock Point Data Lines |
|---|---|---|---|---|
|  | 4 | 8 | 11 | 12 |
| 20 | -- | -- | -- | -- |
| 30 | -- | -- | -- | -- |
| 40 | -- | -- | -- | -- |
| 50 | -- | -- | -- | 0.25 |
| 60 | -- | -- | -- | 0.54 |
| 70 | -- | 0.12 | -- | 0.74 |
| 80 | -- | 0.34 | -- | 0.84 |

In Conimicut Point Reach, also shown by Figure 13, there are interesting differences in pilot performance at sea and the simulator. Although the means are not statistically different except at the turn buoy at data line 0, different strategies are used at sea versus on the simulator. This is confirmed by beta analysis at data lines 4, 8, and 11. Even if an 80-foot difference in means is the same, there is 0.0 probability that the means are the same at data lines 4 and 11, and only a 0.34 probability at data line 8. At sea the pilots make a gentler turn into Conimicut Point by steering toward the Conimicut Point light which brings the ship to the left of the channel when initiating the turn and then to the right of centerline in the new leg. At sea the pilots do not attempt to bring the ship back to the centerline but keep about 65 feet to the right to make the sharp 40-degree right turn onto Bullock Point. On the simulator, pilots attempt to turn closer to the centerline in Rumstick Neck Reach and after turn completion, they attempt to maintain a centerline track. Between data lines 2 and 7, there is a very high standard deviation (137 feet at its highest) that is significantly larger than the at-sea standard deviation (36 feet at its highest). This high standard deviation represents a simulation discrepancy which appears for the first time. At night, there are only two lighted buoys marking this reach. But at sea, in addition to the two lighted buoys, the pilots have other cues. Unlighted buoy "19" is not as useful at night as it is in the daytime, but it is visible when the ship gets close because of the bright light from Conimicut Point light. In addition, other lights on the shoreline which are visible provide additional position fixing information. At sea the pilots track closer to the lighted aid to avoid unlighted buoys. Since the mean track is further to the right and closer to lighted buoys "18" and "22", the standard deviation is smaller. At sea, the turn technique is obviously standardized, the low standard deviation, however, rises sharply to be more similar to the simulator in the 40-degree turn onto Bullock Point Reach since unlighted buoy "24" causes uncertainty. On the simulator the pilots have difficulty judging the ship position with the limited aids. Since the leg is short, they never steady up but recover from the turn and prepare for a new one. Since the simulation was designed to evaluate the aids, there was no background lighting or lights on the land so in this leg the pilots were at a disadvantage on the simulator.

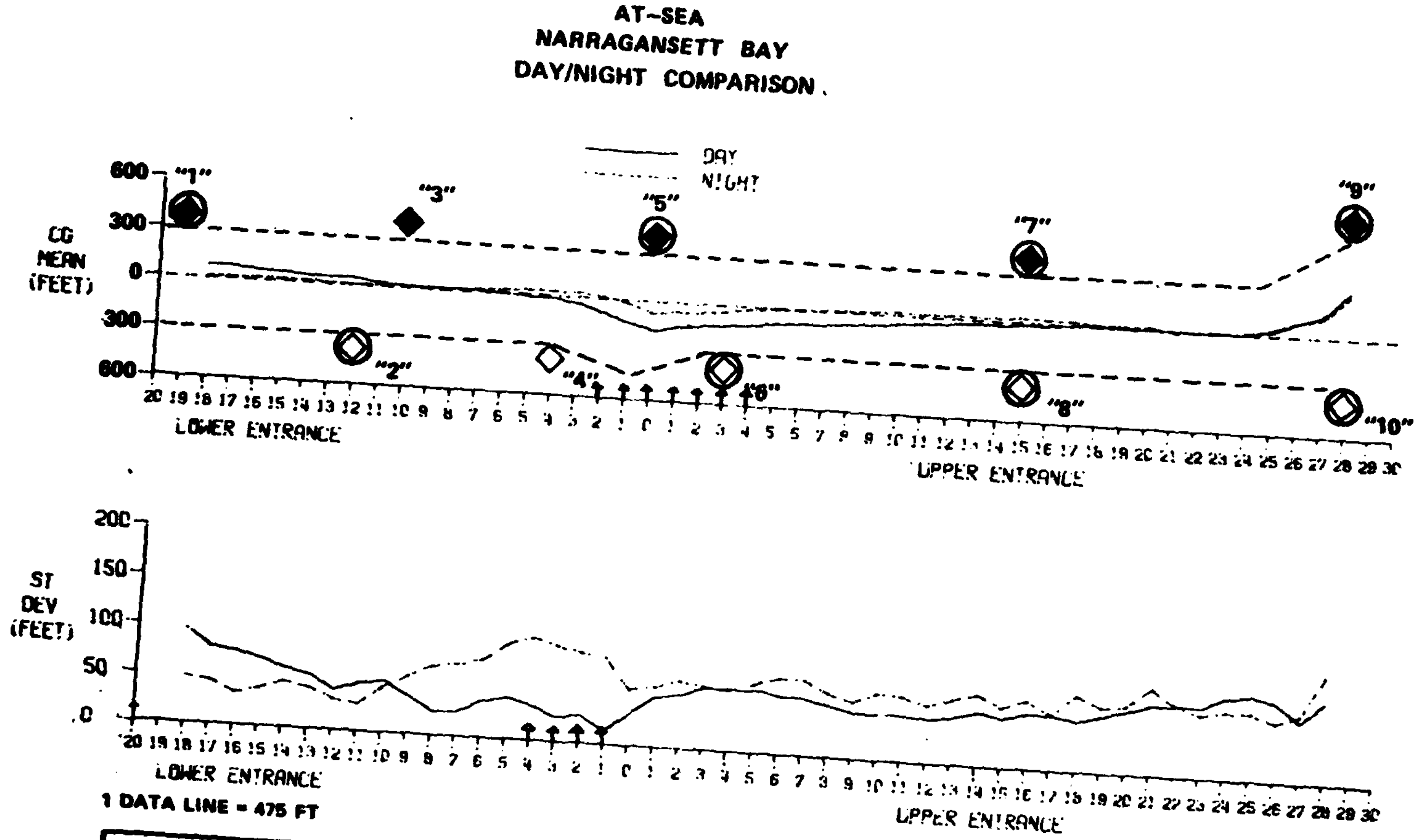
In Bullock Point Reach, shown by Figure 14, the at-sea mean track exits the turn right of centerline and maintains this track. On the simulator the pilots find the centerline 0.2 nm past the turn, swing left, and return the ship to the centerline. There are statistical differences between the means, and these reflect the different strategies. On the simulator, the pilots tend to be stricter maintaining a centerline track. At sea, pilots are more concerned about practical aspects of the transit and less concerned about maintaining a centerline track. Also at sea in Bullock Point Reach, the pilots slow the ship to prepare for the tugs while on the simulator there are no tugs. Although the standard deviation is higher on the simulator, statistical differences only occur in the turn regions. The high standard deviation of the simulator tracks is "safe" because the mean is on the centerline so the track envelop is in the channel boundaries.

## 3.4 DAY/NIGHT DIFFERENCES AT SEA AND ON THE SIMULATOR

This subsection evaluates the four sets of data (at sea versus the simulator and day versus night) in a different way than that presented in previous subsections. The differences in piloting strategies at sea and on the simulator are described in subsections 3.2 and 3.3. At sea, pilots tend to maneuver deep into the cutoff turn in the day, but not as deep at night because the pilots are less sure of the channel edges. As part of that general trend, at night the pilots tended to move away from unlighted buoys to avoid them. On the simulator, the pilots attempted to maneuver for a turn centerline-to-centerline, and they did not use the cutoff turn as much as when at sea. The pilots tried to maintain this strategy regardless of day or night. Therefore, on the simulator performance was very similar although there was a slight tendency to overshoot the turn at night when compared to the day.

Day/night performance differences at sea and on the simulator are compared in the Lower Entrance, Popasquash turn, and Upper Entrance of the channel by Figures 15 and 16. At sea there are differences in turn strategies in day versus night with nighttime techniques being more conservative. During the day the pilots make the Popasquash turn deep inside the cutoff while at night they steer away from unlighted buoy "4" and make the turn close to the centerline. Therefore, in the turn region of Figure 13, between data lines 2 and 4, there are statistical differences between mean tracks. At night, between unlighted buoy "4" and lighted turn buoy 5, the standard deviation is high and statistically different from that during the day (100 versus 30 feet). This indicates pilot uncertainty of the unlighted buoy position which affects nighttime strategy while during the day the buoy aids the pilot by cueing turn strategy. At day, the low standard deviation indicates pilots are certain of their position so strategies are more similar and firm. On the simulator, there are no statistical differences between the means and standard deviation in the turn region. It appears that on the simulator the pilots concentrate harder to maintain a centerline track regardless of day or night. There are day/night differences in the standard deviation between data lines 7 through 11 and 27 and 28 in the Upper Entrance. At day, the standard deviation is higher but constant while at night the standard deviation dips to approximately 25 feet between data lines 7 through 11. At night, as pilots recover from the Popasquash turn, they center on the lighted gated buoys "7" and "8". The dip at night was not seen at sea, but appears here as the pilots try harder to trackkeep on the centerline. The nighttime standard deviation is also lower in turn initiation onto Rumstick Neck Reach.

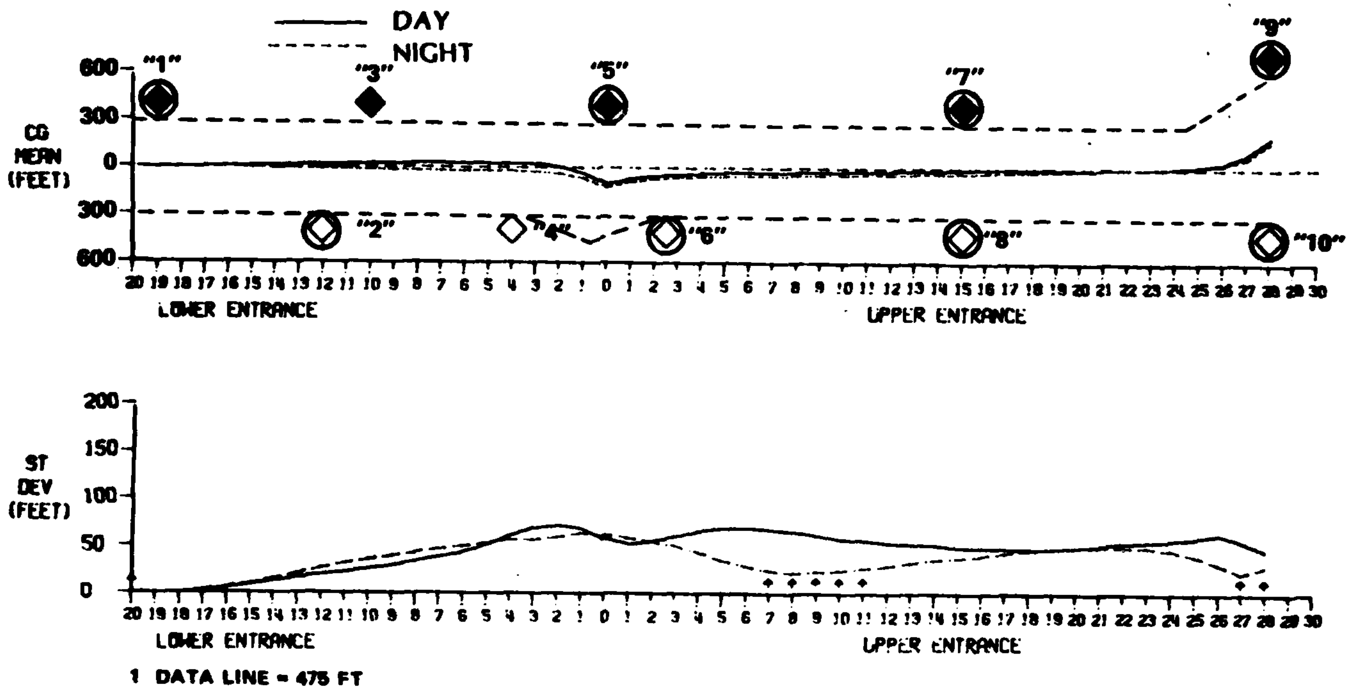
Performance in the turn from Upper Entrance to Rumstick Neck Reach is similar in all four conditions. This similarity is consistent with the pilots' apparent reluctance to make use of the unmarked cutoff. The remaining strategy is turning centerline to centerline. At-sea performance going into the turn is compared for day/night in Figure 15. Performance is virtually identical. The pullout is compared in Figure 17. Performance is very similar. Simulator performance going into the turn is compared for day/night in Figure 16. The means are identical. The standard deviation drops below what it is in the other three conditions, indicating a consistency in strategy. This consistency probably results from the



**BUOYS ARE POSITIONED FOR ILLUSTRATION AND MAY NOT APPEAR IN EXACT CHARTED LOCATION**

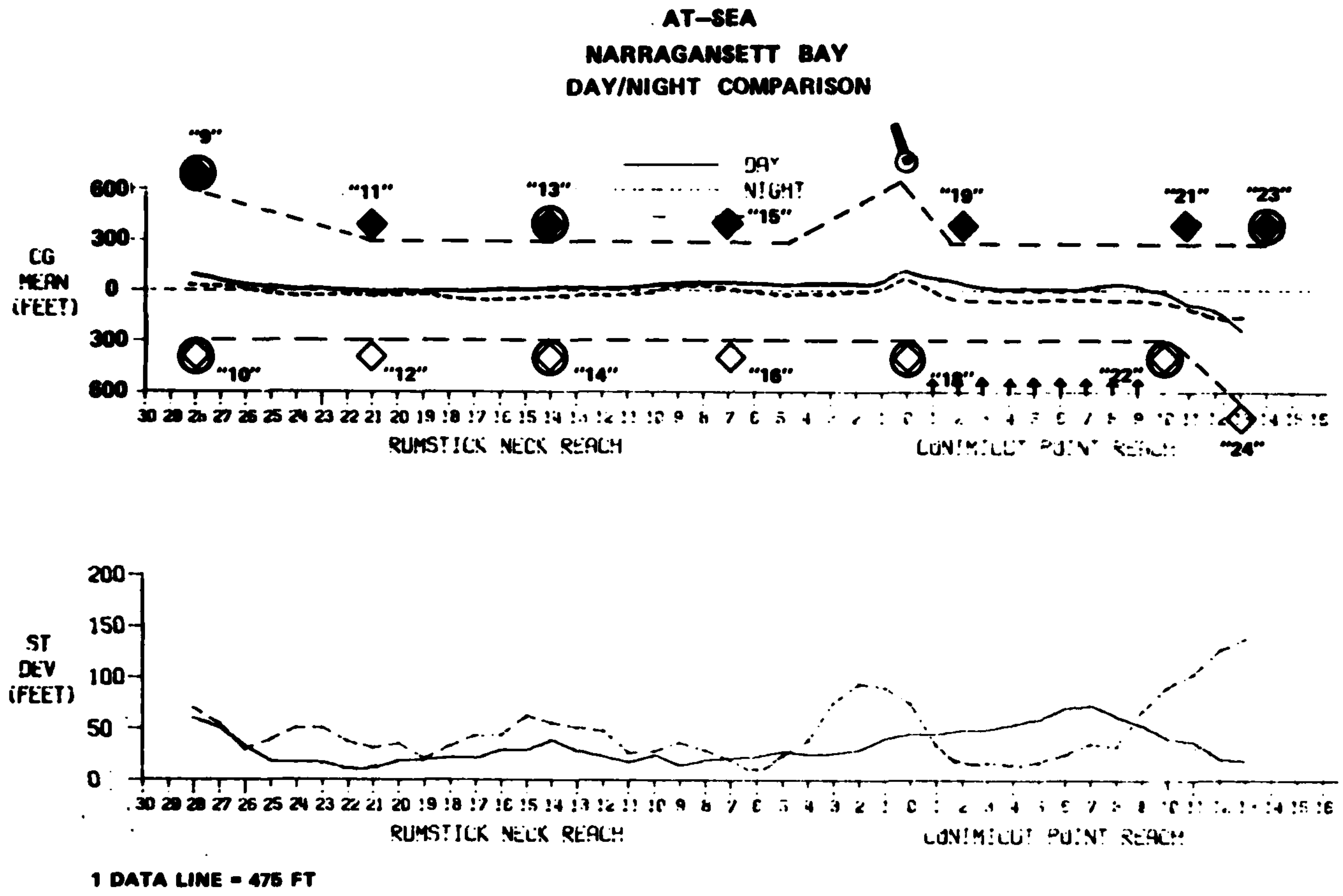
Figure 15. Narragansett Bay Day/Night Comparison: At Sea Lower and Upper Entrance

**SIMULATOR**
NARRAGANSETT BAY
**DAY/NIGHT COMPARISON**



**BUOYS ARE POSITIONED FOR ILLUSTRATION AND MAY NOT APPEAR IN EXACT CHARTED LOCATION**

Figure 16. Narragansett Bay Day/Night Comparison: Simulator Lower and Upper Entrance



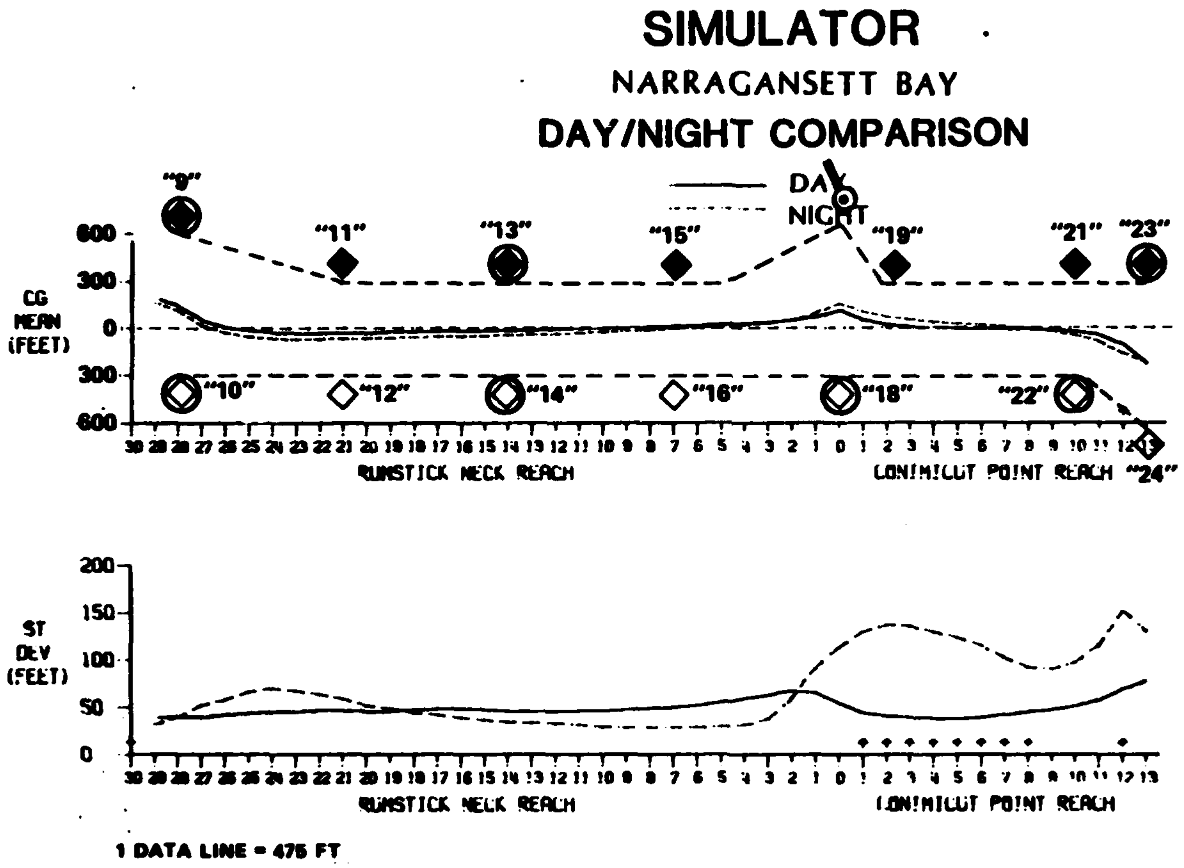
**BUOYS ARE POSITIONED FOR ILLUSTRATION AND MAY NOT APPEAR IN EXACT CHARTED LOCATION**

Figure 17. Narragansett Bay Day/Night Comparison At Sea: Rumstick Neck and Conimicut Point Reaches

convergence of the tendency not to use the cutoff on the simulator and to to use it at night. Simulator performance coming out of the turn is compared for day/night in Figure 18. Again, performance is very similar. The minimal use of the cutoff at night results in a mean showing some slight tendency to make a wider pullout and a slightly increased standard deviation in that region. These differences are not significant.

Day/night performance differences at sea and on the simulator are compared in the Rumstick Neck and Conimicut Point Reaches by Figures 17 and 18. In Rumstick Neck Reach, there are no statistical differences between day and night performance at sea or on the simulator. In Conimicut Point Reach, there are definite day/night differences both at sea and on the simulator. At sea, the means are statistically different throughout the straight leg because daytime tracks are close to the centerline while nighttime tracks are 60 feet right of centerline. During the day unlighted buoys "19 and 21" provide cues as to centerline location and turn initiation. Coming into the turn onto Bullock Point Reach at night, pilots see only lighted buoys "22" and "23". Since the 40-degree turn onto Bullock is to the right and the only lighted buoy marking the straight leg is to the right, pilots keep to the right at night. On the simulator, only the standard deviation is statistically different. At day the standard deviation is smooth, but increases as the turn is initiated. At night the standard deviation is wide at well over 100 feet and increases even further to 150 feet at its highest. Possible reasons for this high standard deviation are discussed in subsection 3.3. During the day, performance was more consistent on the simulator in Conimicut Point Reach, but even so the standard deviation rose to 75 feet.

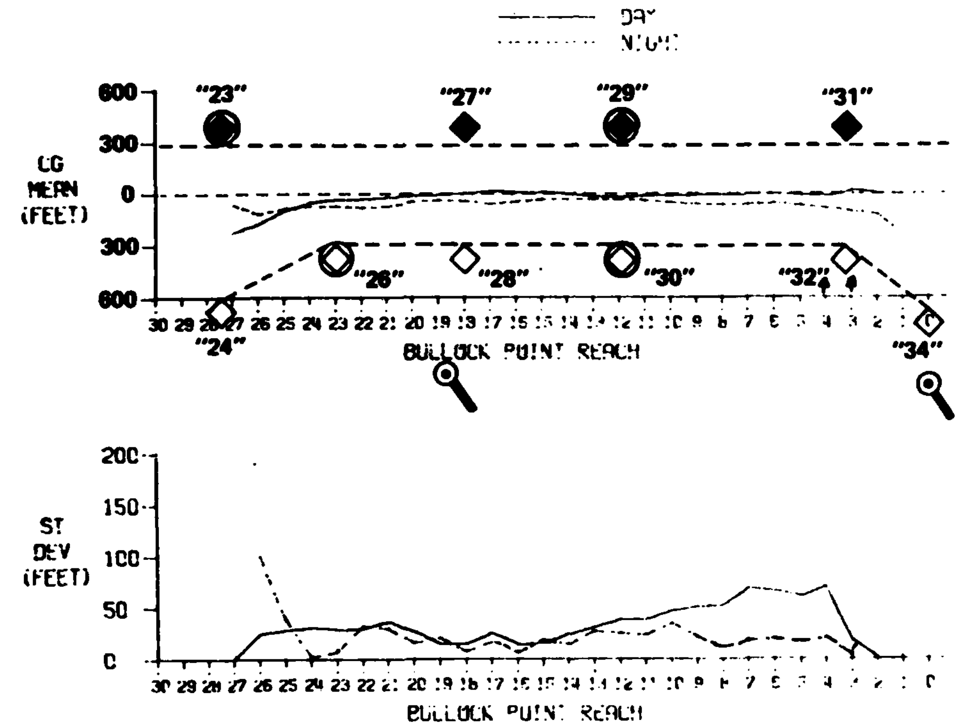
Day/night performance differences at sea and on the simulator are compared in Bullock Point Reach by Figures 19 and 20. At sea, there are no significant differences between the means and the standard deviation. In the day, however, the turn was made sharp onto Bullock as can be seen by the positioning of the mean track at data line 27. At night the turn is also made to the inside but not as sharply, however, the standard deviation is higher at night. Also, at night, the pilots mean track stays to the right side of the channel while at day, the pilots maintain a centerline track. On the simulator, the nighttime turn is made further to the inside than the day, and in addition the standard deviation is high. After recovering from the turn and trackkeeping through Bullock Point Reach, the mean tracks appear identical at day and night and the nighttime standard deviation drops to meet that at day. The mean and standard deviation at night becomes increasingly worse than that during the day as the pilots prepare the turn onto Sabin Point. At night on the simulator, some pilots have problems negotiating the turns as indicated by the significant differences in the means and the standard deviations. This does not occur at sea because the pilots are slowing down to prepare to meet the tugs.



**BUOYS ARE POSITIONED FOR ILLUSTRATION AND MAY NOT APPEAR IN EXACT CHARTED LOCATION**

Figure 18. Narragansett Bay Day/Night Comparison: Simulator Rumstick Neck and Conimicut Point Reaches

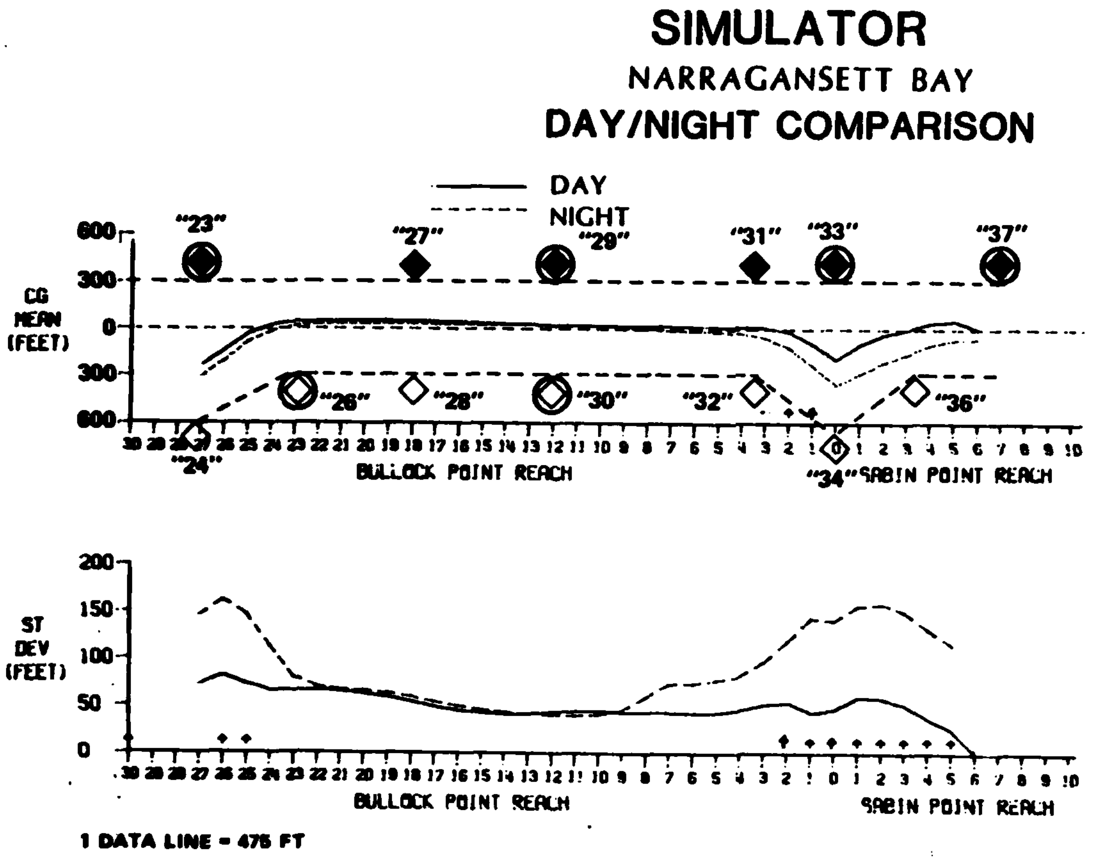
**AT-SEA**
**NARRAGANSETT BAY**
**DAY/NIGHT COMPARISON**



**1 DATA LINE = 475 FT**

**BUOYS ARE POSITIONED FOR ILLUSTRATION AND MAY NOT APPEAR IN EXACT CHARTED LOCATION**

Figure 19. Narragansett Bay Day/Night Comparison At Sea: Bullock Point Reach



**BUOYS ARE POSITIONED FOR ILLUSTRATION AND MAY NOT APPEAR IN EXACT CHARTED LOCATION**

Figure 20. Narragansett Bay Day/Night Comparison: Simulator Bullock Point Reach

## 3.5 SUMMARY

In this section, at-sea data was compared to simulated data collected on the Narragansett Bay. The comparison variable was day versus night.

In the daytime, at-sea and simulator performance were very similar in the straightaways. In the turns, there were apparent differences in pilot strategy in using the cutoff. For those turns that had the extent of the widened area marked by buoys, the at-sea tracks went deep into the widened area. On the simulator the tracks through those turns went centerline to centerline, making less use of the cutoffs. In one turn that was not marked to indicate the cutoff, the at-sea tracks did not go far into the cutoff. In that turn sea and simulator performance were very similar.

At night, the general findings hold. At-sea and simulator performance were very similar in the straightaways. For the turn with the unmarked cutoff, at-sea and simulator performance was very similar as well. There were some interesting differences from sea to simulator at night. At sea, pilots react to unlighted buoys by steering away from them. On the simulator pilots showed no reaction to the unlighted buoys but attempted to steer the centerline course. Generally the mean tracks are similar in the straight legs. One exception occurred in Conimicut Point Reach. Beta analysis shows the mean tracks were not similar. At this point the standard deviation was also different with the simulator standard deviation over 100 feet higher than the at-sea data. This high standard deviation represents a simulation discrepancy which appears for the first time. This difference in standard deviation between sea and simulator is probably due to either: (1) at sea pilots respond to maneuver with a "canned strategy" which may be independent of aids; or (2) cultural lighting is seen at sea in Conimicut Point but not on the simulator.

Day/night differences were more apparent at sea than on the simulator. At sea, pilots tended to maneuver deep into the cutoff turn in the day, but not as deep at night because they were less sure of the channel edges. On the simulator, performance was similar at day and night, although there was a slight tendency to overshoot the turn at night. Regardless of lighting, pilots attempted to maintain a centerline-to-centerline track and did not use much of the cutoff.

## Section 4

## GENERIC PERFORMANCE AT SEA AND ON THE SIMULATOR

### 4.1 INTRODUCTION

The introduction to this paper suggested that validation might have a number of slightly different objectives or forms. The primary objective and form of this validation effort has been the evaluation of the simulator by comparing at-sea transits in a specific channel to transits in a simulation of that same channel. Sections 2 and 3 have described such validations. This section compares at-sea performance to simulator performance taken, not from a dedicated simulation, but from the body of experimental or generic data on which the SRA/RA design manual will be based. These comparisons are further validation of the ability of the USCG/SA simulator to appropriately represent at-sea conditions. They are further a more specific validation of the simulator's ability to provide data to the manual. The data samples used in this section have been selected in the same way that data is selected for the manual.

The methodology for the comparison in this section is illustrated in Figure 21. The experimental objectives as a factor influencing observed performance on the simulator was not a consideration in the earlier comparisons. Here, it means that most of the experimental data is not easily matched to at-sea data. The objectives of the specific experiments and of the manual resulted in the simulation of conditions that would occur at sea with very low frequencies. For example, to ensure differentiation among conditions, the Range Lights experiment[27] evaluated a variety of ranges without any supplementary buoys in the turn or the straightaway. To ensure a conservatism (that is, error in the direction of overestimating risk) in the manual, the turns evaluated in most of the experiments were the worst, realistic case identified by an analysis of major U.S. ports.[28] That turn was the 35-degree noncutoff turn (one not widened by extra dredging). Both to ensure differentiation within experiments and to ensure a conservatism to the manual, most conditions were evaluated with a relatively difficult shiphandling problem: a transit speed of 6 knots with a minimum underkeel clearance and wind and current. The shiphandling problem had its principal effect on performance in the turn pullout and recovery. This effect is described in the SRA Supplemental report[29] and summarized briefly in Appendix C here. The experimental attention that was concentrated on the high-risk maneuvers of turn pullout and recovery means that there the performance was most difficult to match with at-sea performance. Trackkeeping performance, which received less experimental attention, was most naturalistic and easiest to match to at-sea conditions.

---

[27] K.L. Marino, M.W. Smith and W.R. Bertsche, October 1981, op. cit.

[28] W.R. Bertsche and R.T. Mercer. "Aids to Navigation Configurations and the Physical Characteristics of Waterways in 32 Major U.S. Ports." CG-D-7-80, U.S. Coast Guard, Washington, D.C., October 1979. NTIS AD-1083612.

[29] K.L. Marino, M.W. Smith and J.D. Moynehan, November 1983, op. cit.

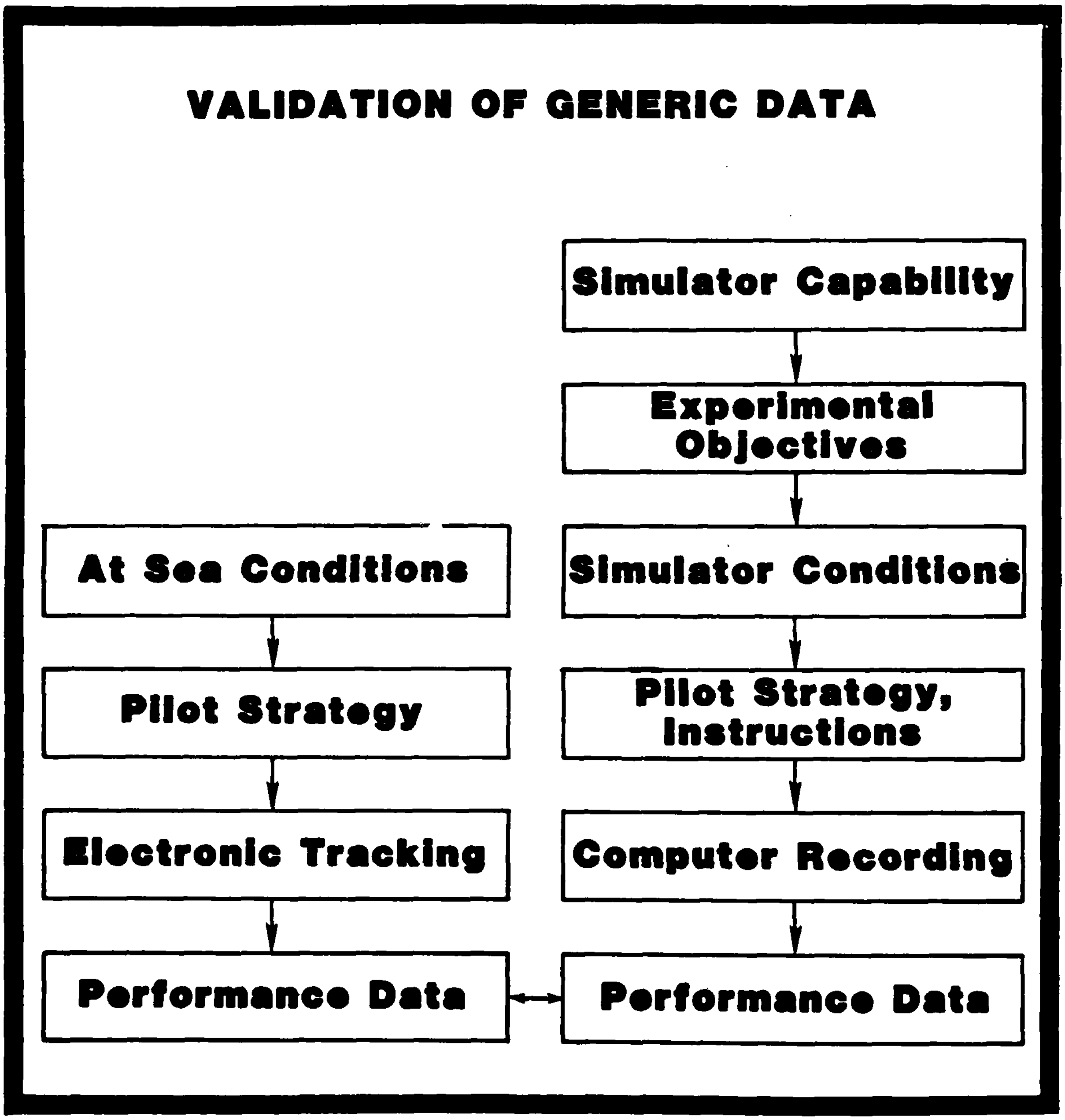


Figure 21. The Validation Methodology for Using Generic Data

This section evaluates the simulator's ability to provide the following samples of generic performance data to the manual:

- trackkeeping on a range
- ship size differences in trackkeeping on a range
- trackkeeping in a gated channel
- day/night differences in trackkeeping in a gated channel
- day/night differences in the turn pullout (a qualitative comparison)

## 4.2 TRACKKEEPING ON A RANGE

The Chesapeake Bay data collection tracked ships on the approach to Baltimore through two channel segments marked by ranges. Performance at sea is compared to performance in a simulation of that channel in Section 2. Here, a sample of that data is compared to trackkeeping on a range in the Range Lights experiment. Samples of at-sea and simulator performance were selected to afford the best possible match in conditions. The conditions and the resulting performance are summarized in Table 8.

The Chesapeake Bay conditions are thoroughly described in Section 2 and in the earlier reports listed there. The comparison plot from which the data was selected appears in that section; the combined plot appears in

TABLE 8. TRACKKEEPING ON A RANGE

Conditions

| Chesapeake Bay | Range Lights |
|---|---|
| Craighill Channel | Scenario 1, Leg 1 |
| gated buoys | no buoys |
| 800 feet wide | 500 feet wide |
| sensitivity: ±15.8 feet | sensitivity: ±6.3 feet |
| daytime | no flashing |
| small ships | 30,000 dwt tanker |
| 12-14 knots | 7+ knots over ground |
| after first gate | after maneuver to centerline |

Performance

| MN (feet) | SD (feet) | N | MN (feet) | SD (feet) | N |
|---|---|---|---|---|---|
| 5.37L | 23.96 | 4 | 1.32R | 11.97 | 8 |

Statistical Tests

Probability that difference in MNs is chance ≥0.10
Probability that difference in SDs is chance ≥0.10
Probability that MNs are same within 20 feet = 0.85
Probability that MNs are same within 30 feet = 0.88

Appendix D. To summarize conditions, the channel included long, 800-foot wide segments marked by both gated buoy and highly sensitive ranges. "Highly sensitive" means that a crosstrack displacement of the ship is easily detected by an observer. Where the data was selected, the sensitivity of the range is such that 68 percent of observers would be expected to detect a crosstrack displacement of ±15.8 feet. (The procedures for quantifying sensitivity from the geometry of a range are described in the appendices of the Range Lights report.) Performance for the small ships was selected as most similar to the 30,000 dwt tanker used in the Range Lights experiment. The ships transited at approximately 12-14 knots with a minimum wind and current. Values were chosen after the first gate to ensure that the effects of the turn were past.

In the Range Lights experiment Scenario 1 also had a highly sensitive range: 68 percent of observers would be expected to detect a crosstrack displacement of ±6.3 feet. (The channel was only 500 feet wide but the quantification of sensitivity takes this into account.) Conditions were equivalent to daytime in the sense that the range "lights" did not flash. The 30,000 dwt tanker was used in this experiment. The values were taken in Leg 1 for trackkeeping on the range between a maneuver to the centerline and the beginning of the turn. In Leg 1 the difficult shiphandling problem has the least effect, the wind and current are following, serving only to increase the ship's speed over the ground. The data line selected is the same one that appears in the draft SRA/RA manual[30] on page C-30 representing trackkeeping on a high sensitivity range with no crosscurrent.

Performance for the two conditions is summarized in Table 8. The means of 5.37 feet left of centerline at sea and 1.32 feet right on the simulator need no t-test to demonstrate that their difference is not statistically significant. By a consideration of both alpha and beta error, it was calculated that the probability that they are the same within 20 feet is 0.85, within 30 feet, 0.88. (This statistical procedure is described in Appendix F.) The standard deviations were compared as variances by a F-test. The difference approaches but does not reach statistical significance at the 0.10 level. Possibly the slight difference in sensitivity of the ranges accounts for the slight differences in the standard deviation of transits. Since the simulated channel with no buoys shows the slightly better performance, it has not suffered from its lack of buoys, at least in trackkeeping. Even without the support of the statistical tests, it can be inferred that performance is very similar. Both the real and the simulated ranges support highly precise trackkeeping and the differences between them are trivial in absolute numbers, compared to the dimensions of the ships and channels. This comparison justifies confidence in the simulator's ability to simulate a highly sensitive range and in the representativeness of the values in the manual.

---

[30] W.R. Bertsche, M.W. Smith, K.L. Marino and R.B. Cooper, February 1982, op. cit.

## 4.3 SHIP SIZE DIFFERENCES IN TRACKKEEPING ON A RANGE

In Chesapeake Bay performance data was collected for trackkeeping on a range with both a group of smaller, and a group of larger ships. These two groups were compared to a dedicated simulation of those two conditions in Section 2. The plotted data appears in Section 2. Generic trackkeeping on a range was evaluated in the Range Lights experiment for the 30,000 dwt ship. This performance is compared to small ship performance at sea in the preceding Section 4.2. Ranges were never evaluated with a larger ship in any AN experiment. But the SRA/RA manual will provide correction factors to "correct" performance with the 30,000 dwt ship to represent that of the 80,000 dwt ship. The present section is an evaluation of this correction as it applies to trackkeeping on a range.

The relevant data is summarized in Table 9. At-sea performance is represented by the same values used in Section 4.2 for the small ships. The means of the transits is 5.37 feet to the left of the centerline of the channel and the standard deviation 17.19 feet. The corresponding large ship performance is relatively less precise with a larger mean and standard deviation, although it is absolutely good, given the ship and channel dimensions. The mean there is 46.63 feet right of the centerline and the standard deviation is 52.62 feet. The Chesapeake Bay simulation performance is included in the table for reference. The third set of data in the table is taken from the data summary presented in the SRA-Supplemental report for the preparation of the ship size correction factor. It represents trackkeeping with following wind and current in a channel marked by short-spaced gates. The mean of the transits is the same for both ship sizes, 16 and 14 feet to the left of centerline. The standard deviation for the large ship is 43 feet, approximately double that for the smaller, which is 26 feet. The SRA/RA manual data for the 30,000 dwt ship is the same Range Lights experimental performance used in the preceding Section 4.2. The values for the 80,000 dwt tanker were obtained by applying the change with ship size observed in the SRA-Supplemental data to the performance observed for the 30,000 dwt tanker in the range experiment: 12 + (43 - 26). In other words, a "correction factor" was applied.

The at-sea performance showed a considerable increase in both mean and standard deviation with the increase in ship size. The manual "data" shows an increase in the standard deviation, but it is not as large. The increase in risk predicted by the manual is less than could be realistically expected. As concluded in Section 2, the 80,000 dwt tanker has less random movement during trackkeeping than the large ships on which at-sea data was collected. However, performance in the gated scenarios, on which the correction factor will be based, did show some change with increase in ship size. This change will be reflected in the body of manual data.

## 4.4 TRACKKEEPING IN A CHANNEL WITH GATED BUOYS

The Narragansett Bay at-sea performance described in Section 3 included precise trackkeeping in a channel with gated buoys. Rumstick Neck Reach in a straight segment 2.2 nm long marked, in the daytime, with gates spaced at approximately 0.6 nm. There, the pilots demonstrated a precision to be expected of a highly sensitive range, that is, the mean was on the

## TABLE 9. SHIP SIZE DIFFERENCES IN TRACKKEEPING ON A RANGE

| | | | | | |
|---|---|---|---|---|---|
| Chesapeake Bay: at sea | | | | | |
| Small Ships | | | Large Ships | | |
| MN (feet) | SD (feet) | N | MN (feet) | SD (feet) | N |
| 5.37L | 17.19 | 4 | 46.63R | 52.62 | 4 |
| Chesapeake Bay: simulator | | | | | |
| 30,000 dwt Tanker | | | 80,000 dwt Tanker | | |
| MN (feet) | SD (feet) | N | MN (feet) | SD (feet) | N |
| 3.84R | 32.71 | 8 | 3.66R | 36.44 | 8 |
| Data from SRA Supplemental Report | | | | | |
| 30,000 dwt Tanker | | | 80,000 dwt Tanker | | |
| MN (feet) | SD (feet) | N | MN (feet) | SD (feet) | N |
| 16L | 26 | 8 | 14L | 43 | 8 |
| SRA/RA Manual | | | | | |
| 30,000 dwt Tanker | | | 80,000 dwt Tanker | | |
| MN (feet) | SD (feet) | N | MN (feet) | SD (feet) | N |
| 1.32R | 11.97 | 8 | 1+R | 29 | NA |

centerline and the standard deviation was very small. In Section 3.2 it was reported that performance in the simulation of this condition was not so precise, showing large standard deviations. There, the large standard deviation on the simulator was attributed to the pilots' strategy or intention. At sea, with the risk of rocky ledges in that area, they took advantage of the marking to stay on the centerline; on the simulator, "close is good enough." However, there are two possible explanations for the greater standard deviation on the simulator. Either the pilots' intended precision is less or the simulator does not allow the precision possible at sea. Generic data is available to eliminate the later possibility.

The at-sea conditions in Rumstick Neck Reach are summarized in Table 10 along with those for the generic performance that provides the best match. The recent SRA-Supplemental experiment included a scenario with more representative shiphandling conditions than the difficult shiphandling conditions that had been used generally in the experiments to provide differentiation and conservatism. As was the case in Rumstick Neck Reach, the simulator performance was that of the Northeast Marine Pilots, Inc., on a 30,000 dwt tanker, in the "daytime", running at 10 knots, with no wind and current, with an underkeel clearance greater than the 1 foot used in the experiments. The performance data used in the comparison was taken after the first gate in both cases to ensure that the effects of the preceding turn is past. The turns in both cases were 35 degrees, but they were dredged and marked differently and the track through them was quite differ-

TABLE 10. TRACKKEEPING IN A CHANNEL WITH GATED BUOYS

Conditions

| | |
|---|---|
| Rumstick Neck Reach | SRA-S Sc. 10 |
| Northeast Marine Pilots, Inc. | Northeast Marine Pilots, Inc. |
| 30,000 dwt tankers (approx.) | 30,000 dwt tanker |
| Day | Day |
| 11 knots (approx.) | 10+ knots |
| Nominal wind/current | No wind/current |
| 5 feet underkeel (approx.) | 10 feet underkeel |
| After first gate | After first gate |
| Gate spacing: 0.6 nm (approx.) | Gate spacing: 1-1/4 nm |

Performance

| MN (feet) | SD (feet) | N | MN (feet) | SD (feet) | N |
|---|---|---|---|---|---|
| 0.61L | 19.17 | 4 | 9.8R | 16.04 | 8 |

Statistical Tests

Probability that difference in MNs is chance $\geq$ 0.10
Probability that difference in SDs is chance $\geq$ 0.10
Probability that MNs are same within 20 feet = 0.84
Probability that MNs are same within 30 feet = 0.88

ent. Notice that the gate spacing is shorter at sea. If the gate spacing is to result in poorer performance, it will be on the simulator. This is a conservative evaluation of the simulator's capability.

Performance in the two domains is summarized in Table 10. The Rumstick Neck Reach transits show trackkeeping with a mean 0.61 feet to the left of centerline with a standard deviation of 19.17 feet, while simulator performance shows a mean 9.8 feet to the right and a standard deviation of 16.04 feet. Neither the difference between the means nor the standard deviations is statistically significant. An analysis shows that the probability that the means are the same within 20 feet is 0.84; within 30 feet, 0.88. The similarity is such that the statistical tests are really unnecessary to support it.

This comparison supports the general conclusion that the simulator is able to support trackkeeping performance in a gated channel similar to that to be expected at sea. It also supports the specific conclusion that the greater standard deviation for Rumstick Neck Reach on the simulator than at sea is the result of differences in pilots' strategy rather than any intrinsic limitation of the simulator.

## 4.5 DAY/NIGHT DIFFERENCES IN TRACKKEEPING IN A CHANNEL WITH GATED BUOYS

Generic trackkeeping in a channel with gated buoys under daytime conditions was evaluated by comparison to at-sea performance in Rumstick Neck Reach in Section 4.4. The generic conditions provided by SRA-Supplemental Scenario 10 provide a good match in conditions and the resulting similarity in performance was impressive. The generic nighttime conditions that provide the best available match to Rumstick Neck Reach at night are outlined in Table 11. The SRA-Supplemental Scenario 9 was a nighttime transit in a channel with gated buoys. There, the conditions differed from those in SRA-Supplemental Scenario 10 and from the at-sea transits in having the difficult experimental shiphandling problem of slow speed, wind and current, and minimum underkeel clearance. The data line chosen to represent trackkeeping is far down in the second leg of the scenario where the current is at a minimum, but there is potentially some residual effect.

A comparison between nighttime performance at sea and on the simulator shows a similiarity between the two. The at-sea performance has a mean of 25.28 feet to the right of centerline and standard deviation of 35.81; the simulator performance has a mean of 33.38 feet to the right and a standard deviation of 40.23 feet. The slightly poorer simulator performance may be the result of the poorer shiphandling conditions. However, neither the means nor the standard deviations show a significant difference at 0.10. An analysis found the probability that the means are the same within 40 feet was 0.74, within 50 feet was 0.876. It is not strictly accurate to say that the simulator performance is a better match during the day than at night. The absolute level of the standard deviation or variances involved limits the degree of similarity between the means that can be expressed. (See Appendix F.)

## TABLE 11. DAY/NIGHT DIFFERENCES IN TRACKKEEPING IN A CHANNEL WITH GATED BUOYS

Rumstick Neck Reach: At Sea

| Day<br>Gate spacing: 0.6 nm (approx.) | | | Night<br>Gate spacing: 0.6 nm (approx.) | | |
|---|---|---|---|---|---|
| MN (feet) | SD (feet) | N | MN (feet) | SD (feet) | N |
| 0.61L | 19.17 | 4 | 25.28R | 35.81 | 4 |

Generic Simulator Performance

| SRA-S Sc 10<br>Representative shiphandling<br>Gate spacing: 1-1/4 nm | | | SRA-S Sc 9<br>Difficult shiphandling<br>Gate spacing: 1-1/4 nm | | |
|---|---|---|---|---|---|
| MN (feet) | SD (feet) | N | MN (feet) | SD (feet) | N |
| 9.8R | 16.04 | 8 | 33.38R | 40.23 | 8 |

Statistical Tests on Difference Between Sea and Simulator

| | Day | Night |
|---|---|---|
| Probability that difference in MNs is chance | ≥0.10 | ≥0.10 |
| Probability that difference in SD is chance | ≥0.10 | ≥0.10 |
| Probability that MNs are same | Within 20 feet, 0.84 | Within 40 feet, 0.74 |
| | Within 30 feet, 0.88 | Within 50 feet, 0.876 |
| | | Within 60 feet, 0.88 |

The sampled performance data can also be considered a comparison of the difference between day and night in the two domains. Both sea and simulator performance shows a greater degree of displacement from the centerline in trackkeeping, approximately one-third of the 85-foot beam of the ship. Both show an approximate doubling of the standard deviation with this displacement. During trackkeeping, under the conditions sampled, the performance and the performance differences are very similar. The next section analyzes day/night differences in the more critical turn maneuver.

## 4.6 DAY/NIGHT DIFFERENCES IN THE TURN PULLOUT (A QUALITATIVE ANALYSIS)

It was pointed out in the introduction to this Section, 4.1, that the at-sea and generic turns available do not match. This is true even in Narragansett Bay where the channel and turn dimensions are more similar to the experimental conditions. The at-sea turns are all "cutoff", or widened by dredging. At the generic turns evaluated on the USCG/SA simulator are "noncutoff", or abrupt, and most have been run with difficult shiphandling conditions. However, the turn maneuver has been identified as critical in the AN process and the day/night differences have been identified as particularly important in the turn. Any possible comparison of available data is worthwhile.

A striking finding in the AN experiments was that pilots use a nighttime strategy for turns. At least for the smallest ship evaluated, the 30,000 dwt tanker; they turn harder at night, presumably to avoid the uncertain outside edge of the channel. This effect is described on the Turn Lights[31] and SRA-Supplemental[32] reports. A summary table from the SRA-Supplemental report is reproduced here as Table 12 for reference. Notice that for the 30,000 dwt tanker the mean of the transits exists the turn closer to the centerline at night than in the daytime, but with a higher standard deviation at night. It would have been a valuable component of the validation effort to have been able to observe that switch in strategy at sea. There was no opportunity to do so, but there seems to be an analogous switch in Narragansett Bay.

The difference in day and night performance at sea is described in Section 3.4. There, it was observed that the pilots take advantage of the cutoff region in the daytime, but do not do so at night. Again, the best explanation is that they are not sure of the extent of the cutoff at night. The turns in upper Narragansett Bay Channel are not marked to indicate the room available.[33] A summary of data appears in Table 13. The comparison plot from which these data are taken is in Section 3; the combined plots are in Appendix E. The means in the table are the distance from a centerline continued from the straight segment, as if there were no widened area. Notice that in the daytime the mean of transit moves to the right in a righthand cutoff much farther than at night. As an example, the pullout

---

[31] J. Multer and M.W. Smith, February 1983, op. cit.

[32] K.L. Marino, M.W. Smith and J.D. Moynehan, December 1983, op. cit.

[33] Memo, Smith to Ridley, July 22, 1983.

## TABLE 12. GENERIC PERFORMANCE IN A 35-DEGREE NONCUTOFF TURN

| | Day | | | | | | Night | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Experiment | Scenario | Data Line | Mean | Standard Deviation | Relative Risk Factor | Experiment | Scenario | Data Line | Mean | Standard Deviation | Relative Risk Factor |
| **30,000 dwt ship** | | | | | | | | | | | | |
| **Three buoys** | One-Side | 1 | 3 | 59R | 34 | 0.0001 | Turn Light | 1 | 2 | 8R | 56 | 0.0012 |
| **Two buoys** | One-Side | 6 | 3 | 94R | 33 | 0.0035 | Turn Light | 8,9 | 3 | 28R | 67 | 0.0110 |
| **One buoy** | Ship Variables | 2 | 3 | 72R | 45 | 0.0068 | SRA Supplemental | 9 | 3 | 6R | 73 | 0.0126 |
| **80,000 dwt ship** | | | | | | | | | | | | |
| **Three buoys** | Ship Variables | 7 | 3 | 107R | 54 | 0.1841 | SRA Supplemental | 8 | 3 | 79R | 70 | 0.1363 |
| **Two buoys** | SRA Supplemental | 7 | 3 | 81R | 45 | 0.0485 | Turn Light | 10 | 3 | 187R | 58 | 0.7019 |
| **One buoy** | Ship Variables | 6 | 3 | 137R | 65 | 0.3859 | | | | | | |

## TABLE 13. TURN PULLOUTS IN NARRAGANSETT BAY

| | Day | | Night | |
|---|---|---|---|---|
| Data Line | MN (feet) | SD (feet) | MN (feet) | SD (feet) |
| **Popasquash Turn: Cutoff is to Right** | | | | |
| 0 | 179.26R | 32.65 | 72.14R | 57.05 |
| 1 | 146.64R | 51.35 | 55.20R | 61.86 |
| 2 | 128.28R | 55.38 | 50.20R | 68.48 |
| **Onto Rumstick Neck Reach: Cutoff is to Left** | | | | |
| -28 | 99.58L | 60.67 | 35.77L | 70.21 |
| -27 | 62.08L | 51.51 | 27.52L | 54.80 |
| -26 | 36.70L | 33.23 | 21.37L | 29.91 |
| **Onto Conimicut Reach: Cutoff is to Left** | | | | |
| 0 | 143.41L | 46.15 | 85.46L | 74.75 |
| 1 | 71.84L | 45.64 | 24.18L | 34.43 |
| 2 | 31.27L | 49.43 | 59.90L | 11.76 |
| **Onto Bullock Reach: Cutoff is to Right** | | | | |
| -26 | 173.13R | 25.23 | 119.29R | 100.33 |
| -25 | 98.20R | 28.91 | 88.06R | 39.03 |
| -24 | 48.20R | 31.08 | 69.45R* | 2.60* |

*incomplete data

from Popasquash is 179.26 feet to the right in the daytime and only 72.14 feet to the right at night. The shift in strategy is similar for the other turns. At night they appear to avoid an uncertain edge, making a more difficult turn to do so.

This comparison between the simulator and at-sea turns is obviously inexact and qualitative. It is comparison by analogy. If the pilots make a more abrupt turn at night to avoid the edge of the cutoff that they cannot see in one type of turn, it is more credible that they would do so in another type of turn. This comparison supports statements about day/night differences in turns made in the experiments and the manual.

## 4.7 SUMMARY

In this section, comparisons were made between selected samples of at-sea data and selected samples of simulator data taken, not from dedicated simulations, but from the body of available generic data from AN experiments. Quantitative comparisons demonstrated close similarities between at-sea and generic samples for trackkeeping on a highly sensitive range, trackkeeping in a channel with gated buoys in the daytime, and trackkeeping in a channel with gated buoys at night.

Generic measures of trackkeeping on a range with a larger, 80,000 dwt ship, a condition that had never been run in an experiment, were produced by applying a ship size correction factor to data for the smaller, 30,000 dwt ship. The resulting measures showed performance deteriorating with ship size but not to the same extent it did in Chesapeake Bay. The SRA/RA manual will predict some increase in risk with ship size in the low-risk trackkeeping region.

Because of the importance of the turn maneuver to the project and because of the importance of day/night differences to the turns, a qualitative comparison was made of a day/night shift in pilot strategy on the simulator and at sea. On the simulator it has been observed that the pilots make a harder turn at night, coming out of the turn closer to the center of the channel, avoiding the unseen outside edge. At sea in Narragansett Bay, they use less of the unseen cutoff area at night. By analogy this at-sea behavior is a validation of that observed on the simulator.

While only a sample of the body of AN data can be compared to at-sea data, that is the nature of validation.

# Section 5

# GENERAL SUMMARY AND CONCLUSIONS

## 5.1 INTRODUCTION

A general methodology was planned for the AN project validation. Conditions as similar as possible at sea and on the simulator were arranged and the resulting performance data in the two domains were compared. The degree of similarity or difference in performance is a measure of simulator "validity". If a dissimilarity in conditions exists, or in pilot behavior, or in data collection, it is difficult to interpret differences in performance as reflecting the validity of the simulator. There were differences from sea to simulator that had to be considered before interpretations could be made about simulator validity.

One objective of simulator use, the one originally chosen for validation, was to reproduce the performance observed in a specific waterway. The method for this validation was the comparison of at-sea performance with simulator performance data taken from a simulation prepared to reproduce the specific waterway as closely as assumed necessary. The goal of such a simulation is not necessarily to reproduce the conditions in the waterway, but to reproduce piloting performance observed in that waterway. It was the plots of ship tracks collected in both domains that were compared. When the validation task was begun, it was understood that the pilots would play a considerable role in the data match. Therefore, pilots with "local knowledge" in the waterways selected were brought to the USCG/SA simulator to represent their groups. However, the extent to which they could influence the matches was not appreciated until after the fact. Where their strategies or intentions differed from sea to simulation, it was impossible to evaluate the effectiveness of the simulator. Where their intentions coincided, it was possible to consider the comparison an evaluation of the simulator.

The reproduction of performance in a specific waterway was not a central objective of the AN project. The simulator was developed to generate quantitative data on which to base the manual. In evaluating the ability of the simulator to produce this generic data, some of the issues are different. "Local knowledge" is not a factor and the pilots' intentions have less of an effect. However, the experimental design and procedures have effects on the observed performance. The experiments tended to concentrate on worse case events or difficult shiphandling tasks to ensure differentiation among conditions and conservatism in performance data. Where possibilities that would be expected with low frequency at sea were introduced to meet the objectives of the experiment or of the manual, no comparison can be made with available at-sea data. Where the simulator conditions are more representative of real world conditions, comparisons can be made that are evaluations of the simulator.

## 5.2 SIMULATION FOR A SPECIFIC WATERWAY

Two separate simulations of specific waterways are described in this report. First, performance was observed in Chesapeake Bay in the approach to Baltimore, Maryland. A simulation of this waterway was run on the USCG/SA simulator by members of the Association of Maryland Pilots, Inc. Later, performance was observed in Narragansett Bay in the approach to Providence, Rhode Island. A simulation of this waterway was run by members of the Northeast Marine Pilots, Inc. To the extent that different conditions were observed in the two waterways, the two sets of data allow the evaluation of the effectiveness of the simulator in producing a variety of effects. Regardless of the similarities or differences in the conditions observed, two separate and independent evaluations of the effectiveness of the simulator are more convincing than one.

The effects of the pilots' strategy on the plots of ship tracks are most obvious in the Narragansett Bay waterway with its larger number and variety of turns. The turns in that channel are cutoff, or widened by dredging, but are not necessarily marked to correspond to that dredging. Apparently, the local pilots have developed distinctive ways of making each turn. To simplify the findings of Section 3, the crosstrack means of the set of transits at sea in the daytime generally go deep into the cutoff, making a more gradual turn by taking advantage of the available space. In one turn for which the extent of the cutoff is not marked, the tracks do not go as deep. At night they do not make as great a use of any of the cutoffs, marked or not. It was inferred that this shift in tracks is made because the pilots are less sure of the exact location and extent of the cutoff at night. Day or night, after each turn the mean of the transits comes to the channel centerline and stays on it for some distance before beginning the maneuver for the next turn.

The simulator replication of this waterway was run in combination with the Implementation experiment, a comparison of performance under a number of alternative aid arrangements for the channel. The pilots were instructed to stay on the centerline for two reasons. First, they reported that they used the channel centerline at sea. Second, it was necessary for the purposes of the Implementation experiment. In order to find differences between experimental conditions, it is necessary to minimize the variability within a condition. The pilots in the experiment were very cooperative. The emphasis on the simulator was different from that at sea. On the simulator the mean of the transits stayed closer to the centerline (a continuation of the straight segment centerline, as if the turns were not cutoff) through the turn and returned to the centerline in the next segment sooner than at sea. This track did not allow for much difference from day to night on the simulator. The simulator tracks provided the best match to at-sea tracks in the straightaways and in the turn with the unmarked cutoff. Differences in pilots' strategy from sea to simulator are interesting but are not an evaluation of the simulator.

The Narragansett Bay evaluation identified the importance of the pilots' strategy or intentions in matching sea performance on the simulator. This experience suggests that the pilots should be encouraged to transfer their at-sea strategies to the simulator in future attempts to reproduce at-sea performance on the simulator. Ideally, at-sea data should be collected and inspected before the simulation to understand the strategies. The next best preparation would be having an experimenter ride at sea with a pilot (preferably more than one), observing his behavior and discussing it with him as close to the time and place as would not interfere. At the very least, there should be thorough discussion with a representative pilot (preferably more than one). The last two steps were followed in this case, but without the sensitivity to the issues that comes from experience. Instructions to the pilots for a simulation intended to reproduce at-sea performance should encourage the pilots to transfer their behavior, forgetting any instructions from previous experiments or any preconceived assumptions about what is wanted on a simulator. (Note that the instructions would be quite different when the objectives of the simulation are different.)

Where the pilots' intended tracks at sea and on the simulator coincided, generally in the straightaways and in one turn, a quantitative comparison of the degree of similarity is possible. Over much of the transit, the statistical analysis shows the crosstrack means of the sets of transits in the two domains are the same within errors 50 or 60 feet of each other with a probability of 0.88. (The statistical analysis is described in Appendix F.) Those values were obtained in a situation where the ship's beam is 85 feet and the width of the channel is 600 feet. The amount of difference is meaningless given the dimensions involved.

The Narragansett Bay nighttime evaluation produced data demonstrating that the objective of reproducing a specific waterway, or the performance in a specific waterway, is the most demanding one for a simulator. There was a region of the channel where the simulator performance was conspicuously different from at-sea performance. (The relevant data is illustrated in Figure 11 in Section 3.) Halfway up the channel in the wide turn onto Conimicut Point Reach, the standard deviation of the transits is very large, both at sea and on the simulator. At sea as the transits pass an unlighted buoy, the mean pulls away from it and the standard deviation drops sharply. On the simulator the mean of the transits does not pull away and the standard deviation does not drop but continues to rise, producing one of the highest standard deviations seen in the AN project. This unusual performance needs discussion.

There are a number of possible explanations for the nighttime difference in Conimicut that vary in how easily they are remedied. One possibility is that the at-sea maneuver is well-rehearsed to avoid the unlighted buoy and takes place relatively independently of aids or other visual cues. This possibility is the most likely. It is consistent with nighttime performance in Popasquash turn where the mean pulled away from an unlighted buoy at sea but not on the simulator. If this is the case, the remedy is the general one of encouraging the pilots to transfer their at-sea strategies to the simulator. Another possibility is that they see the unlighted buoy at sea. This is less likely. If they could see the buoy, it is unlikely that they would pull away from it more at night than in the daytime. It is possible they see it only when it is too close to be useful or they see it only in some fraction of the transits. These possibilities could be simulated by having the unlighted buoy appear as a dark object against a black sky or appear in only some fraction of the transits. A third possibility that they are using visual cues other than the aids in the channel at sea. If these are lights, these could be reproduced in the nighttime simulation. The explanation that is most critical for the simulator is a difference in the distance cues available. On the simulator only the geometry of the lights provides a cue for distance: lights further away are higher on the screen. There is no decrease in intensity with distance. But is unlikely that a paucity of distances cues should be a problem in some sections of the transit and not others. Only controlled experimentation could identify the reason with certainty. But the first three possibilities can be dealt with without certainty. While it would be a painstaking process, it should be possible to bring the simulation and, thus, the performance closer to the at-sea case.

The objective of the simulation of a specific waterway is not always the reproduction of the at-sea performance. If the objective is the evaluation of the aids, as it was for the Implementation task, there is less of a demand on the simulator. Inspection of the nautical chart (reproduced in Figure 8 in Section 3) shows that the turn onto Conimicut was made with a minimum of lighted aids in the vicinity of the turn. Conimicut Point Light is on the middle of the turn, off the channel edge and passes abaft the beam early in the turn. Only the outside apex of the turn has a lighted aid on the channel edge. The high standard deviation at sea indicates that this marking is less than optimal. Only the large width of the channel relative to the sizes of the ships that make that transit makes it workable. The real question at sea is why the standard deviation comes down before the next turn. It is not in response to close, lighted aids. (Possible reasons are listed in the preceding paragraph.) The availability of ship track data for the evaluation of aids in a waterway is almost unique. Generally, it will not be available. Then, a simulation that consists only of the aids, both lighted and unlighted for day and only lighted for night, is an appropriate, conservative test of the aids. Poor performance in such a simulation can be considered indicative of poor aids. For such an objective there need not be, and should not be, any attempt to simulate the most subtle cues available to the pilot. This assumption was made early in the AN project when it was decided that aid arrangements would be evaluated using a simple visual scene. The validation effort has supported that beginning assumption.

The Chesapeake Bay conditions described in Section 2 differed from those in Narragansett Bay. There was a different pilot group; a greater variety of ship types and sizes was tracked; the channel and turn dimensions were greater; and the aids included highly sensitive ranges as well as gated buoys. As in Narragansett Bay, there was some difference in pilot strategy through the turns from sea to simulator. However, ranges at sea or on a simulator dictate a strategy. On the ranges, the statistical analysis shows that the crosstrack means of the sets of transits in the two domains are the same within error of 30, 40, or 50 feet of each other with probabilities close to and including 0.88. Those values were obtained in a situation where the ships' beams were approximately 70 to 100+ feet and the width of the channel is 800 feet. The amount of difference is meaningless, given the dimensions involved.

The Chesapeake Bay comparison included ship size as a variable. The performance of a group of smaller ships was well matched by that with the simulated 30,000 dwt tanker. However, the performance of a group of larger ships was not as well matched by that with the simulated 80,000 dwt tanker.

While trackkeeping on the range, the 80,000 dwt tanker showed a mean as close to the centerline and a standard deviation as small as that of the 30,000 dwt tanker, while at sea the larger ships showed a larger crosstrack error and larger standard deviation than did the smaller ships. If a limitation of the simulation was identified by the Chesapeake Bay validation, it was the too-accurate trackkeeping performance of the 80,000 dwt tanker. This is presumably the result of a lack of random movement in the ship model during trackkeeping. However, there are only very limited consequences to the 80,000 dwt tankers' precision in trackkeeping for a simulation of performance in a specific waterway. It is more relevant to the collection of generic data and is discussed in this context in Section 5.3.

## 5.3 SIMULATION FOR GENERIC DATA

Most of effort on the AN project involved collecting generic data on the simulator. This data collection was designed to meet the objectives of an experiment or of the manual. For example, for the purposes of the manual, the experiments concentrated on higher-risk, higher-angle noncutoff turns which do not occur with a very high frequency in the real world. Another conspicuous example is the difficult shiphandling conditions that were imposed on the experimental scenarios to ensure differentiation among conditions for the experiments and to ensure a conservatism to estimates of risk for the manual. The conservatism that was introduced in this way has been evaluated and discussed in the SRA Supplemental,[34] Radar I,[35] and Chesapeake Bay[36] reports. Those findings are summarized here in Appendix C. For these and other reasons most of the generic data is not amenable to comparison to at-sea data. However, some of the generic performance data, some scenarios, or some parts of scenarios are representative of real world conditions. Unfortunately, these tend to be the low-risk trackkeeping portions of the channels. Only a qualitative evaluation of generic turn performance is possible. However, the availability of the at-sea data presents a unique opportunity to validate some portion of the generic data. Validation of any sample of performance on the simulator increases confidence in the whole. The following is a very brief review of the comparisons described in Section 4 and the implications of these comparisons.

In Chesapeake Bay the ships were tracked on a highly sensitive range. This performance was compared to generic data for trackkeeping on a highly sensitive range taken from the Range Lights experiment.[37] Trackkeeping performance was very accurate in both domains. A statistical analysis showed that the means were the same within an error of 20 feet with a probability of 0.85, within an error of 30 feet with a probability of 0.88. When pilot strategy or differences in shiphandling conditions do not mask effects, the simulator produced valid generic performance, here, for trackkeeping on a range.

---

[34] K.L. Marino, M.W. Smith and J.D. Moynehan, December 1983, op. cit.

[35] J. Multer and M.W. Smith, December 1983, op. cit.

[36] G.E. Grant, J.D. Moynehan and M.W. Smith, February 1983, op. cit.

The variety of ships tracked in Chesapeake Bay were divided into smaller and larger ships. The Range Lights generic data collected with the 30,000 dwt tanker model provided a good match to the performance of group of smaller ships. No generic evaluation was ever run of ranges with the 80,000 dwt tanker model. Instead, the ship size correction factor proposed for the SRA/RA design manual was applied to the data for the 30,000 dwt tanker to generate an approximation. This approximation was compared to at-sea data for the larger ships. The generic values for the 80,000 dwt ship showed more accurate trackkeeping than the at-sea values. This latter comparison is described in Section 4.3. There is no plan to adjust the generic data for the manual on the basis of this data. Trackkeeping has proven to be a low-risk maneuver and has been de-emphasized in the AN project and in the manual. There is no question that the 80,000 dwt tanker is a sufficiently difficult ship in the higher-risk turn, pullout, and recovery maneuvers. Both its inherent and piloted controllability have been evaluated in the Ship Variables[38] and SRA Supplemental[39] reports.

In Narragansett Bay, the strategy of the pilots was most restricted while trackkeeping in Rumstick Neck Reach, a straight segment with close-spaced gated buoys under daytime conditions. Performance was compared to trackkeeping in a gated channel in the SRA Supplemental experiment.[40] A statistical analysis showed that the means of the two sets of transits were the same within an error of 20 feet with a probability 0.84, within an error of 30 feet with 0.88. When pilot strategy or differences in shiphandling condition do not mask effects, the simulator produced valid generic performance, here, for trackkeeping in a channel with gated buoys.

The runs in Narragansett Bay were split into day and nighttime transits. The best available match among generic conditions had the difficult shiphandling problem, which might have degraded performance more than night alone. Despite this, there was still reasonable similarity between the sea and simulator performance. The means of transits in the two domains were the same within an error of 50 feet with a probability of 0.87, within an error of 60 feet with a probability of 0.88. Even with some bias by difficult shiphandling conditions, the simulator produced valid generic performance, here for trackkeeping in a channel with gated buoys at night.

Differences in turn geometry, pilot strategy, and shiphandling conditions from sea to simulator combined to make it impossible to make direct, quantitative comparisons for the higher-risk more critical turn region. AN experiments have shown that the turns are higher risk and that this is especially true at night. A striking finding was a nighttime strategy for turns discussed in the Turn Lights[41] and SRA Supplemental[42] reports. At

---

[37] K.L. Marino, M.W. Smith and W.R. Bertsche, October 1981, op. cit.

[38] W.R. Bertsche, D.A. Atkins and M.W. Smith, October 1981, op. cit.

[39] K.L. Marino, M.W. Smith and J.D. Moynehan, December 1983, op. cit.

[40] Ibid.

[41] J. Multer and M.W. Smith, December 1983, op. cit.

[42] K.L. Marino, M.W. Smith and J.D. Moynehan, December 1983, op. cit.

night in the noncutoff turns evaluated in the experiments, the pilots turned harder exiting closer to the centerline; presumably to avoid the uncertain outside edge of the channel. The at-sea day-to-night shift in turn strategy observed in the cutoff turns in Narragansett Bay seems analogous. There, the pilots stayed away from the edge of the cutoff at night. The at-sea finding supports the simulator finding. It increases confidence that the nighttime turn strategy seen on the simulator is a valid reflection of what the pilots would do at sea in such a turn and not an artifact of the simulation. The "validation" of such an unexpected and subtle performance effect increases confidence in the validity of the simulator as a whole.

## 5.4 SUMMARY OF FINDINGS

Two sets of at-sea performance data from Chesapeake and Narragansett Bays were used for comparison with simulator performance data. The simulator data was taken both from simulations of the specific waterways and from the available body of generic performance data collected for the experiments and for the manual. The "validity" of the simulator is a subjective total of these comparisons. The following is a summary of the findings from the comparisons.

- For a variety of conditions performance in the straightaways provided good matches. Statistical tests demonstrated that the crosstrack means of the sets of tracks from sea and from the simulator were the same within errors of 20 to 80 feet for much of the transits. These data came from situations in which the ships' beams were 85 to 125 feet and the channel width was 500 to 800 feet. Performance was the same, given the dimensions involved. A high degree of similarity was observed in buoyed channels, in channels marked by ranges, for daytime transits, and for nighttime transits.

- Turn performance at sea showed distinctive strategies that did not always transfer to the simulator. When they did transfer, the differences between the means of the transits at sea and on the simulator were, again, very small. Statistical tests demonstrated that the means were the same within errors of 50 to 80 feet through the turn. These data came from a situation in which the ship's beam was approximately 85 feet and the channel turn area was widened beyond 600 feet. Performance was the same, given the dimensions involved.

- The 30,000 dwt tanker model provided a very good match to the performance of a group of small ships tracked in Chesapeake Bay. The 80,000 dwt tanker model was too accurate in trackkeeping to match the performance of a group of larger ships.

- When at-sea performance was not strictly dependent on the aids but on well-rehearsed pilot strategies or on visual cues other than the aids, simulator performance did not match. Poorer performance on the simulator, where subtle cues were missing, can be interpreted as identifying potential weaknesses in the aid system.

- An increase in the cautiousness of turn strategy from day to night was observed at sea in Narragansett Bay and in the generic turn data. (In both cases the pilots avoid the uncertain outside edge by turning harder at night.) Differences in turn geometry made a quantitative comparison impossible, but the observation of this unexpected behavior both at sea and on the simulator supported the general validity of turn performance on the simulator.

## 5.5 SUMMARY OF CONCLUSIONS

The findings of the validation effort support conclusions on the future use of the USCG/SA simulator for several purposes and on the use of the available performance data already collected.

- The use of the USCG/SA simulator for the experimental or generic evaluation of aid arrangements is the least demanding task for the simulator. In a generic waterway there is little expectation that very subtle cues be present or that the pilots transfer very specific behaviors. A visual scene that includes only simple representations of the aids is a conservative and appropriate test of generic conditions. The validation findings support confidence in the simulator for this use, in the data already collected, and in the use of these data for the development of the design manual.

- The use of the USCG/SA simulator for the evaluation of aid systems in specific waterways is also a relatively undemanding task for the simulator. A visual simulation that consists only of the aids in the waterway simply presented is, again, a conservative test of the aids. Poor performance on the simulator identifies those regions of the waterways where the aids are potentially weak, where they function at sea because they are supplemented by other factors like cultural objectives or well-rehersed pilot strategies relatively independent on visual cues. The present simulator capability and experimental methodology is appropriate for this use.

- The use of the USCG/SA simulator to reproduce the unique performance to be expected at sea in a specific waterway is the most demanding task for the simulator. Such reproduction will occur only if the pilots transfer accustomed behavior to the simulator and if conditions on the simulator are sufficiently similar to allow them to implement that behavior. Any limitations to reproduction that were observed have consequences only for the validation effort itself. They have no consequences for the projects other uses of simulator performance data.

# APPENDICES

# APPENDIX A

# CHARACTERISTICS OF THE USCG/SA SIMULATOR

The simulator used in the experiments is located at Ship Analtyics, Inc. in North Stonington, Connecticut. Its visual capability was developed for the U.S. Coast Guard for the Performance of Aids to Navigation Project. The Components of the simulator are illustrated in Figure A-1 and consist of the following:

1. The ship's bridge
2. Standard ship's controls
3. Ship's indicators
4. An advanced "radio aided" navigation display unit
5. Computer generated visual system
6. Host computer with requisite interface equipment
7. Postexercise data reduction facility

## A.1 THE SHIP'S BRIDGE

The bridge is 15 feet 9 inches wide and 15 feet 6 inches deep with windows for viewing the visual scene. Additional facilities include a chart table with a ten drawer chart storage. The lighting on the bridge can be controlled, and total darkness can be achieved.

## A.2 SHIP'S CONTROLS

The control mechanisms found in the bridge simulator are tied directly to the host computer, providing the proper inputs for ship's controls with resultant ship's motion incorporated in the visual image. These control mechanisms include the following:

1. A ship's wheel and helm unit
2. An engine order telegraph which provides control of the ship's engines both ahead and astern. Propeller rpm and ship acceleration are determined by ownship's dynamics programmed for the computer for each specific ship size.

## A.3 SHIP'S INDICATORS

The indicators are also tied to the host computer to provide information to the pilot. They include the following:

1. Two gyro repeaters, one on the steering stand and one mounted with an azimuth circle
2. A shaft rpm indicator
3. A rudder angle indicator
4. A ship's clock which has been modified to show scenario time

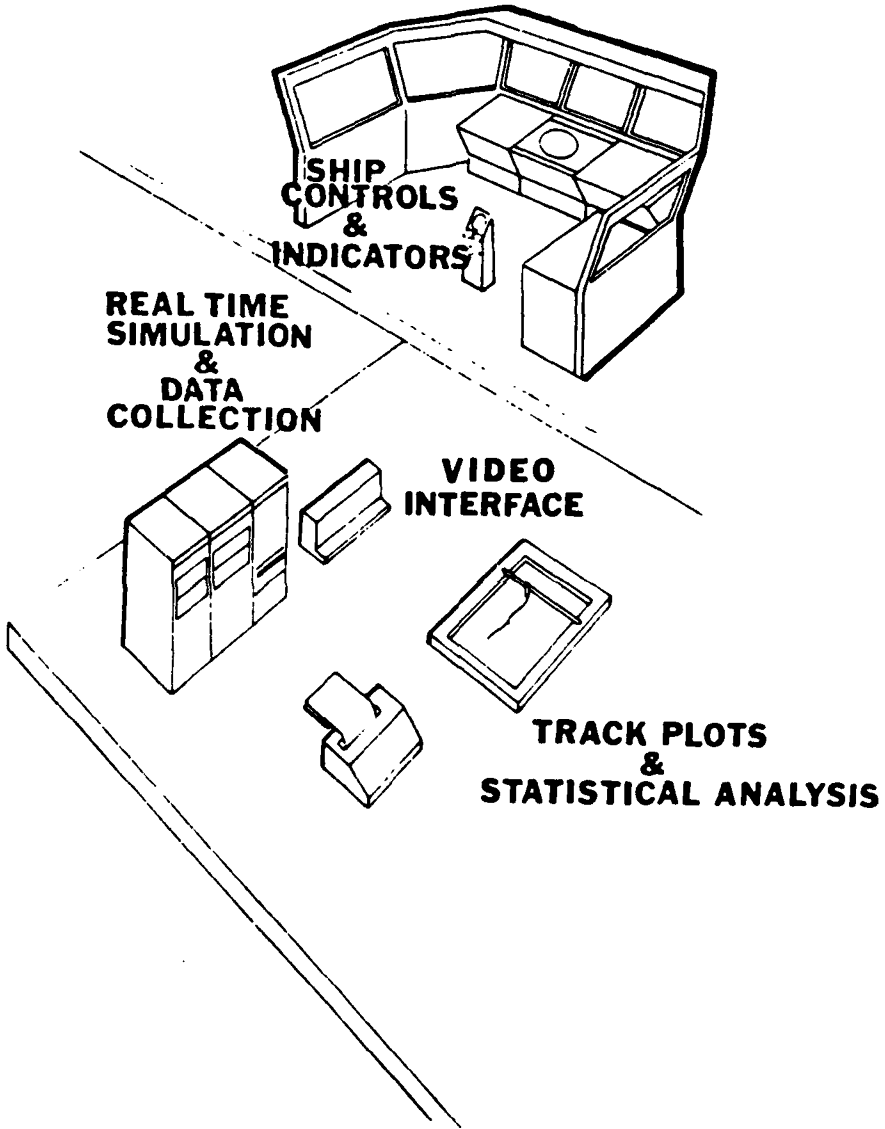


Figure A-1. USCG/EA Simulator

## A.4 RADAR PPI

A 16-inch PPI simulates a generic 3 cm radar.

## A.5 NAVIGATION DISPLAY UNIT

The navigation display unit presents a variety of information displays to the pilot. It was used for radio aids experiments included in the U.S. Coast Guard Aids to Navigation Project, Phase II.

## A.6 VISUAL SYSTEM

The visual system provides a 182-degree horizontal and a 20-degree vertical field of view. The dynamic scene for daytime conditions includes ownship's bow, the sky, water, and visible aids. The nighttime scene translates the aids into appropriate lights.

## A.7 THE HOST COMPUTER

The host computer provides processing for the visual system consistent with ownship's characteristics, including maneuverability. The visibility conditions, the hydrodynamic model, and individual scenario topographical conditions are part of the initial conditions.

## A.8 THE DATA REDUCTION CAPABILITY

Computer facilities are available to provide postexercise data reduction, analysis, and hard copy for individual scenarios or groups of scenarios.

# APPENDIX B

## COMPONENTS OF VALIDATION TASK

The planning and data collection for the validation task consists of a large number of components and reports. The principal reports were:

### Planning

"Aids to Navigation Presimulation Report for Validation: Validation For A Simulator-Based Design Project" by G.E. Grant and M.W. Smith. Technical Memorandum written for U.S. Coast Guard, Washington, D.C., September 1982.

> This report describes a plan for simulator validation in support of the United States Coast Guard's Performance of Aids to Navigation Project, addendum. This experimental plan identifies simulated scenarios which pilots will transit to compare to piloted performance in actual at-sea transits in the Upper Chesapeake Bay and Upper Narragansett Bay.

### Chesapeake Bay Validation

"Aids to Navigation Preliminary Data Analysis for the At-Sea Data Collection Program" by Eclectech Associates. Technical Memorandum written for U.S. Coast Guard, Washington, D.C., August 1981.

> This technical memorandum consists of performance track plots of at-sea data collected in the Craighill and Upper Craighill channels of the Chesapeake Bay.

"At-Sea Data Collection for the Validation of Piloting Simulation" by R.B. Cooper, R.C. Cook and K.L. Marino. Interim Report written for U.S. Coast Guard, Washington, D.C., September 1981.

> This report analyzes piloted performance of data collected at sea in the Upper Chesapeake Bay under a variety of conditions. The data was evaluated to determine the effect of (1) traffic, (2) direction of travel, (3) ship characteristics, and (4) wind and current on piloted performance.

"Aids to Navigation Validation: Preliminary Observations and Data Analysis" by J.D. Moynehan and G.E. Grant. Technical Memorandum written for the U.S. Coast Guard, Washington, D.C., January 1983.

> This technical memorandum contains performance track plots and pilots' comments pertaining to the simulation experiment which included scenarios modeled after the Chesapeake Bay. The experiment subjects were pilots from the Maryland Pilots Association. (The experiment included Narragansett Bay scenarios as an evaluation of local knowledge and difficult shiphandling conditions.)

"Aids to Navigation Validation Draft Principal Findings Report: Simulation I: Chesapeake Bay" by G.E. Grant, J.D. Moynehan, and M.W. Smith. Draft Interim Report written for U.S. Coast Guard, Washington, D.C., February 1983.

This report compares piloted differences between performance observed at sea in the Upper Chesapeake Bay and performance observed on the simulator in scenarios modeled after the Chesapeake Bay conditions. The data collected in this experiment demonstrated qualitatively that the simulation of visual aids provided performance comparable to at-sea aids.

## Narragansett Validation

"Aids to Navigation Validation II/Implementation: Preliminary Observations and Data Analysis" by G. Grant and J. Moynehan. Technical Memorandum written for U.S. Coast Guard, Washington, D.C., May 1983.

This technical memorandum consists of performance track plots of and pilots' comments pertaining to the simulation experiment of the Narragansett Bay. The experiment subjects were pilots from the Northeast Marine Pilots Association.

"Aids to Navigation Implementation At-Sea Preimplementation Draft Principal Findings" by J.D. Moynehan. Draft Interim Report written for U.S. Coast Guard, Washington, D.C., September 1983.

This report describes the at-sea data collection tracked on 30,000 dwt tankers inbound on Narragansett Bay.

# APPENDIX C

# VALIDATION RELATED ISSUES

Validation related evaluations have been reported in a number of earlier reports. These evaluations include: (1) simulator comparison evaluation, (2) range light simulation evaluation, and (3) design condition evaluation. These reports are summarized below.

## C.1 SIMULATOR COMPARISON EVALUATION

Two simulators, one in the Maritime Administration's Computer Aided Operations Research Facility (CAORF) and one in the USCG/Ship Analytics' facility, have been used to collect data for the Aids to Navigation Project. A discussion and comparison of piloted performance on the simulators appears in the Channel Width Principal Findings Report and in the Ship Variables Principal Findings Report. A summary of these discussions is presented below.

1. "Aids to Navigation Principal Findings Report on the Channel Width Experiment: The Effects of Channel Width and Related Variables on Piloting Performance" (Section 3) by M.W. Smith and W.R. Bertsche. CG-D-54-81, U.S. Coast Guard, Washington, D.C., January 1981. NTIS-AD-A111337.

   A purpose of the Channel Width experiment was to compare piloted performance of similar scenarios on both simulators to determine if similar conclusions could be drawn. It was found that the two simulations showed similar patterns of differences when in the straight legs. The magnitude of the standard deviation, however, was generally larger in Leg 2 for the USCG/SA simulation than for CAORF. It was determined that there were several possible reasons for this difference: (1) the wind was from a different direction in the USCG/SA simulation, (2) the visual information was poorer in the USCG/SA simulation, and (3) the bow image in the USCG/SA scenarios provided less perceptual reference than the CAORF ship. It was recommended that the next experiment, the Ship Variables experiment, test these hypotheses and further compare simulator performance.

2. "Aids to Navigation Principal Findings Report on the Ship Variables Experiment: The Effect of Ship Characteristics and Related Variables on Piloting Performance" (Section 3) by W.R. Bertsche, D.A. Atkins, and M.W. Smith.

   One purpose of the Ship Variables experiment was to implement changes identified in the Channel Width experiment to determine if piloted performance on the USCG/SA simulator would be more similar to that observed on the CAORF simulator. As a result of changes to wind direction and bow image, performance on the CAORF and the USCG/SA simulators was the same, not only in relative differences among conditions but also in absolute magnitude of effects for conditions simulated.

## C.2 RANGE LIGHT SIMULATION EVALUATION

Range lights were first simulated on the USCG/SA simulator in the Range Light experiment. They have subsequently been simulated in the Narragansett Bay validation scenarios. The validity in the use of the range lights for shiphandling is discussed in the Range Light Principal Findings Report and is summarized here.

1. "Aids to Navigation Principal Findings Report: Range Light Characteristics and Their Effect on Performance" (Appendices A and B) by K.L. Marino, M.W. Smith and W.R. Bertsche. CG-D-66-81, U.S. Coast Guard, Washington, D.C., January 1981. NTIS-AD-A109716

> The United States Coast Guard and International Association of Lighthouse Authorities (IALA) have published equations derived from laboratory and field tests describing the relationship between the vertical separation of the lights and the detectability of the horizontal separation (representing crosstrack movement of the ship). Prior to the conduct of the range light experiments, perception tests were conducted on the simulator to establish whether or not the perception of the horizontal separation of vertical lights would be equivalent to the previously tested in both laboratory and field tests. The simulator results showed that the piloting performance using range lights may be evaluated on the USCG/SA shiphandling simulator if the vertical visual angle separating the lights is greater than 25 minutes of arc. Perception of the horizontal separation of such range lights will be equivalent to behavior at sea.

## C.3 DESIGN CONDITION EVALUATION

All the Aids to Navigation experiments have been run with difficult shiphandling conditions including slow speed, low underkeel clearance, and crosswind and crosscurrent on the assumption that only difficult shiphandling conditions would reveal differences in the performance of alternative aid arrangements. An evaluation of these "design conditions" is included in three reports. These reports are identified below with a summary of the design condition evaluation.

1. "Aids to Navigation Radar I Principal Findings: Performance in Limited Visibility of Short Range Aids With Passive Reflectors" (Section 5) by J. Multer and M.W. Smith. CG-D-79-83, U.S. Coast Guard, Washington, D.C., December 1983.

> The Radar I experiment included one scenario with the ship transiting at a slow speed with low underkeel clearance in a channel without wind and current to evaluate their effects. It was found that wind and current increased the difficulty of shiphandling particularly in that part of the transit where the ship had the greatest crosstrack velocity, the turn pullout and recovery. It also concluded that the design conditions of wind and current contribute conservatism to the estimates of risk in the highest risk portion of the transit. They do this differentially among conditions, increasing the estimates more for such higher-risk conditions as low buoy density.

2. "Aids to Navigation SRA Supplemental Experiment Draft Principal Findings: Performance of Short Range Aids Under Varied Shiphandling Conditions" (Section 6) by K.L. Marino, M.W. Smith, and J.D. Moynehan. U.S. Coast Guard, Washington, D.C., December 1983.

The SRA Supplemental Experiment included a scenario with "representative" shiphandling conditions to test the assumption that more difficult shiphandling conditions build a degree of conservatism into the data. The scenario consisted of the ship transiting at 10 knots with a 10-foot underkeel clearance in a channel with no wind and current. It was found that performance was more precise under the more representative shiphandling conditions. This further verifies that without difficult conditions built into scenarios, performance differences between variables would be minimal. Also there is more conservatism in the data because of the low frequency environmental or shiphandling events.

3. "Aids to Navigation Validation II/Implementation: Preliminary Observations and Data Analysis" (Section 4) by G. Grant and J. Moynehan. U.S. Coast Guard, Washington, D.C., May 1983.

The validity of the simulator and its experimental results depend on the design of the experimental conditions under which the simulator data was collected. One purpose of the validation simulation scenarios was to examine the impact of the difficult shiphandling "design" conditions on experimental results. It was found that the artificiality imposed on the conditions increased the difficulty of the shiphandling problem. This indicates that there is conservatism in the "design conditions" that are reflected in the SRA/RA manual.

# APPENDIX D

# CHESAPEAKE BAY SIMULATOR AND AT-SEA PILOTED PERFORMANCE TRACK PLOTS

This appendix contains piloted performance track plots of the at-sea and simulated data collected in the Upper Chesapeake Bay. Figure D-1 shows the chart and location for the at-sea data collection. The at-sea data plotted shows performance for a group of four ships smaller than 45,000 dwt and four ships larger than 45,000 dwt. Figure D-2 shows the simulated scenario chart. The simulator data plotted shows performance with eight transits of the 30,000 dwt tanker model and eight transits of the 80,000 dwt tanker model.

Table D-1 identifies the conditions represented by the track plots. The plots are grouped by location: Craighill Entrance, Craighill Channel, Craighill Upper Range.

There are two types of track plots in this section. A combined plot is illustrated by Figure D-3. It consists of a series of three plots for one of the two channel legs. The axis for the abscissa is scaled so that one unit of alongtrack distance represents 475 feet (5/64 nm). The top plot displays the crosstrack mean of the center of gravity of the ships as they transit the channel and the middle plot displays the crosstrack standard deviation. The bottom plot is a combined plot showing the crosstrack mean and an envelop encompassing two standard deviations to either side, an area within which performance is expected to occur 95 percent of the time. A comparison plot is illustrated by Figure D-7. For each comparison there are two sets of axes, one showing the means for the two conditions and one showing the crosstrack standard deviation as the performance measures. Data is plotted as a continuous unbroken line and a dotted line to distinguish the experimental conditions from each other.

Statistical tests were used to test the differences in performance at each data line, to determine if any differences between conditions were statistically significant. The means were compared using a t-test. The symbols along the axis of the mean plot indicate a difference at the 0.10 level of significance. The standard deviations were compared as variances using an F test. The symbols along the axis of the standard deviation plot also indicate a difference at the 0.10 level of significance.

NOTE: On the plots, buoys are positioned for the purpose of illustration and may not appear in their exact charted location.

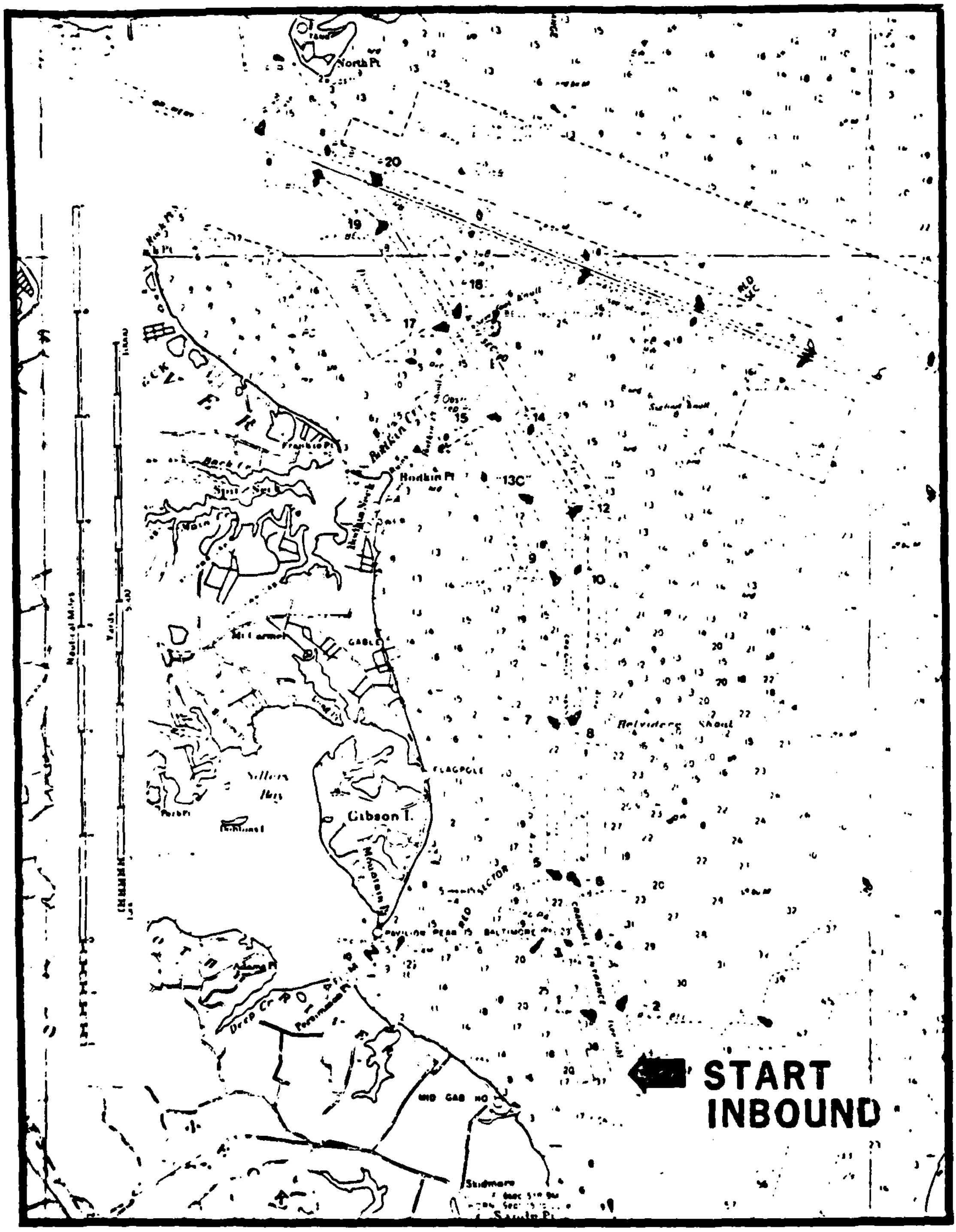
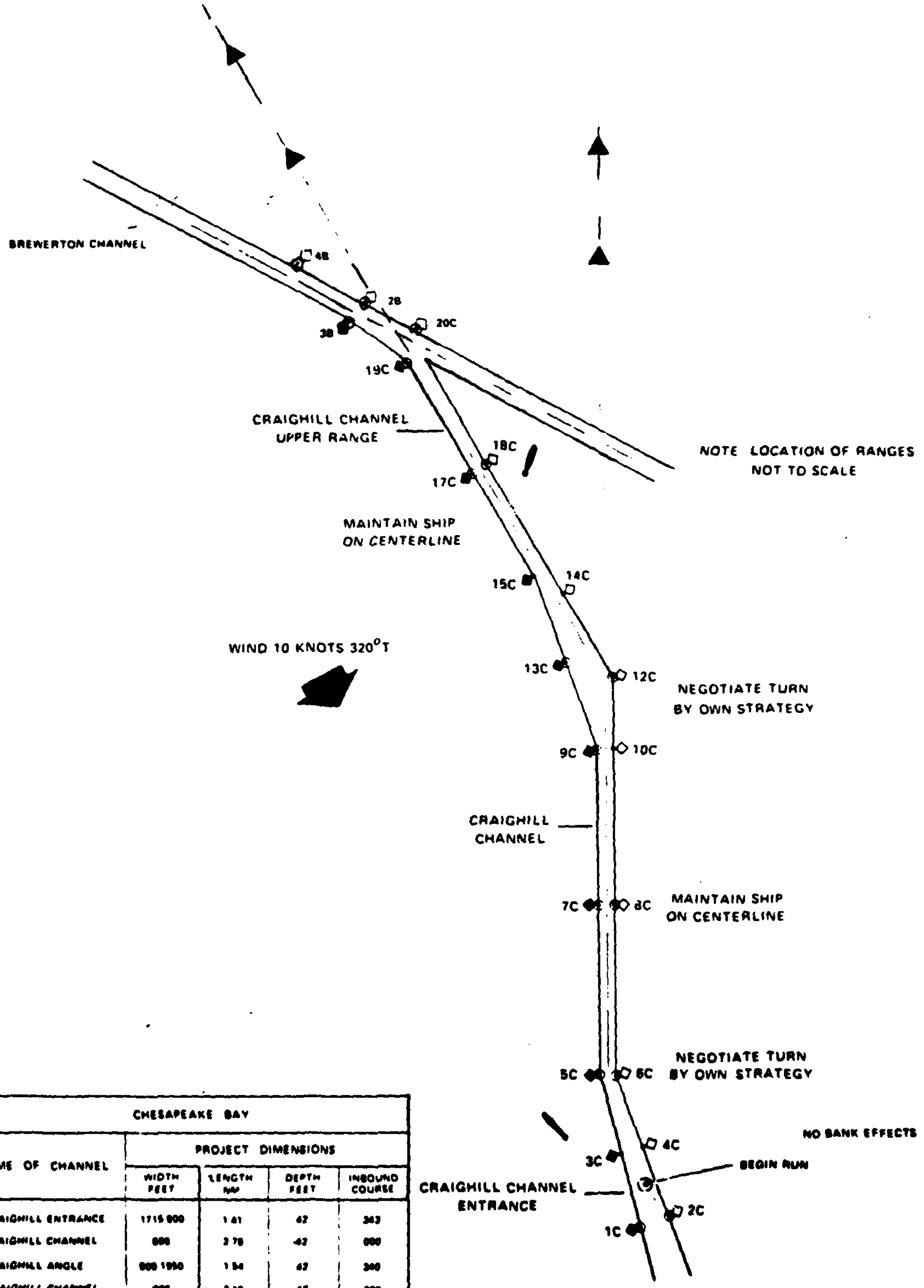


Figure D-1. Upper Chesapeake Bay Where At-Sea Data Was Collected
Chart No. 12273 of April 17, 1982



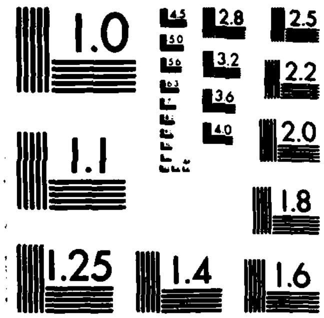
| CHESAPEAKE BAY | | | | |
|---|---|---|---|---|
| NAME OF CHANNEL | PROJECT DIMENSIONS | | | |
| | WIDTH FEET | LENGTH NM | DEPTH FEET | INBOUND COURSE |
| CRAIGHILL ENTRANCE | 1715 800 | 1 41 | 42 | 343 |
| CRAIGHILL CHANNEL | 800 | 2 78 | 42 | 000 |
| CRAIGHILL ANGLE | 800 1950 | 1 54 | 42 | 340 |
| CRAIGHILL CHANNEL UPPER RANGE | 800 | 2 18 | 42 | 329 |
| BREWERTON CHANNEL | 800 | | 42 | 291 |

SCENARIOS 4 5 AND 6

Figure D-2. Simulated Chesapeake Bay Channel Chart Scenario

AD-A146 789   AIDS TO NAVIGATION PRINCIPAL FINDINGS REPORT:   2/2
VALIDATION FOR A SIMULATOR-..(U) ECLECTECH ASSOCIATES
INC NORTH STONINGTON CT  M W SMITH ET AL. JUL 84
UNCLASSIFIED   EA-84-U-203 USCG-D-06-84 DOT-CG-835285-A   F/G 13/10   NL

END
DATE
FILMED
11 84
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

## TABLE D-1. CHESAPEAKE BAY TRACK PLOTS

| Figure | Type of Data Collected | Ships | Channel | Plot |
|---|---|---|---|---|
| D-3 | At Sea | Small | Craighill Channel | Combined |
| D-4 | At Sea | Large | Craighill Channel | Combined |
| D-5 | At Sea | Small | Craighill Channel Upper Range | Combined |
| D-6 | At Sea | Large | Craighill Channel Upper Range | Combined |
| D-7 | At Sea | Large/Small | Craighill Channel | Comparison |
| D-8 | At Sea | Large/Small | Craighill Channel Upper Range | Comparison |
| D-9 | Simulator | 30,000 | Craighill Entrance | Combined |
| D-10 | Simulator | 80,000 | Craighill Entrance | Combined |
| D-11 | Simulator | 30,000 | Craighill Channel | Combined |
| D-12 | Simulator | 80,000 | Craighill Channel | Combined |
| D-13 | Simulator | 30,000 | Craighill Channel Upper Range | Combined |
| D-14 | Simulator | 30,000 | Craighill Channel Upper Range | Combined |
| D-15 | Simulator | 30,000/80,000 | Craighill Entrance | Comparison |
| D-16 | Simulator | 30,000/80,000 | Craighill Channel | Comparison |
| D-17 | Simulator | 30,000/80,000 | Craighill Channel Upper Range | Comparison |

AT–SEA

# SMALL SHIPS

## CHESAPEAKE BAY

## CRAIGHILL CHANNEL

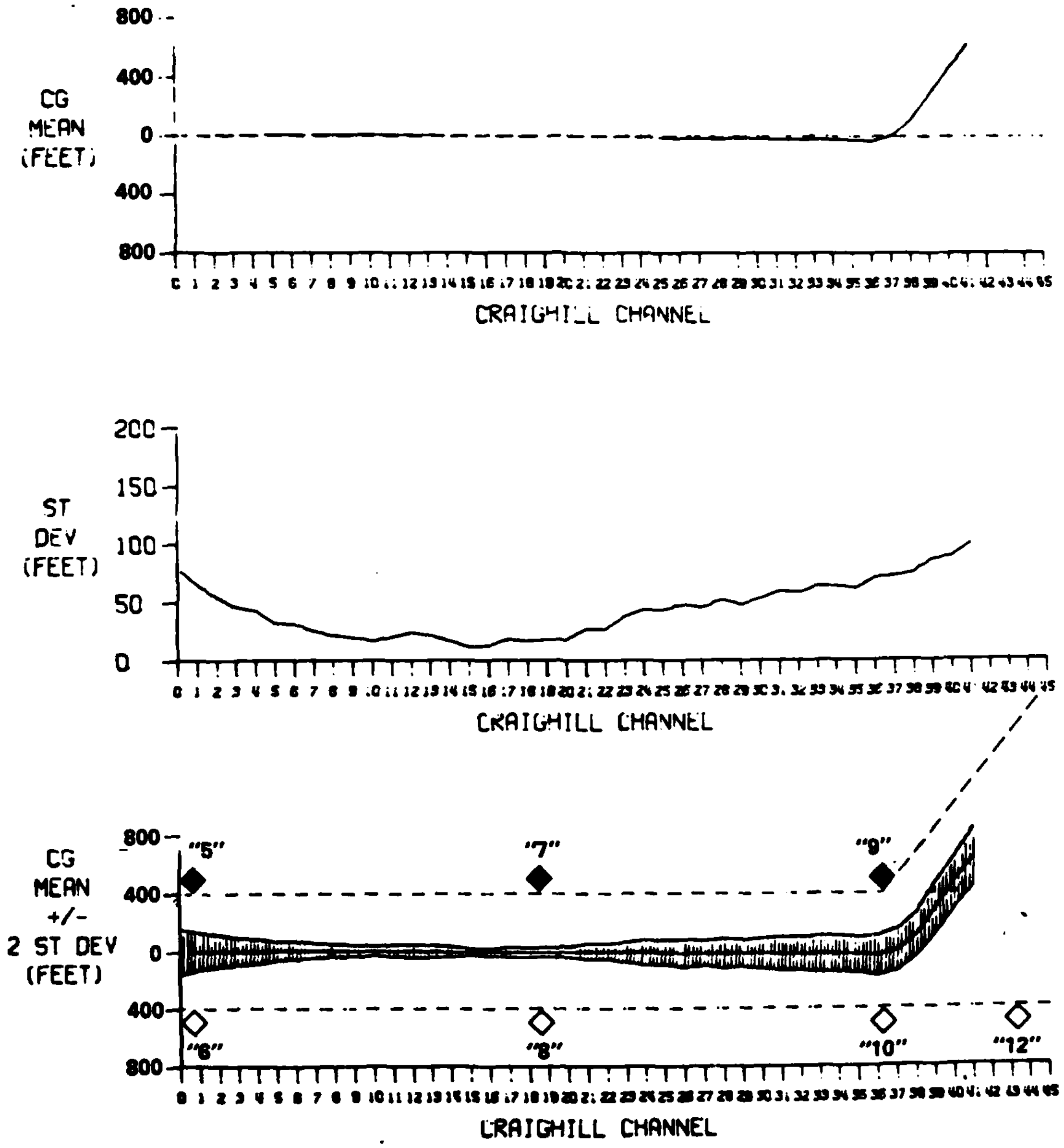


Figure D-3

### AT-SEA

# LARGE SHIPS

## CHESAPEAKE BAY

## CRAIGHILL CHANNEL

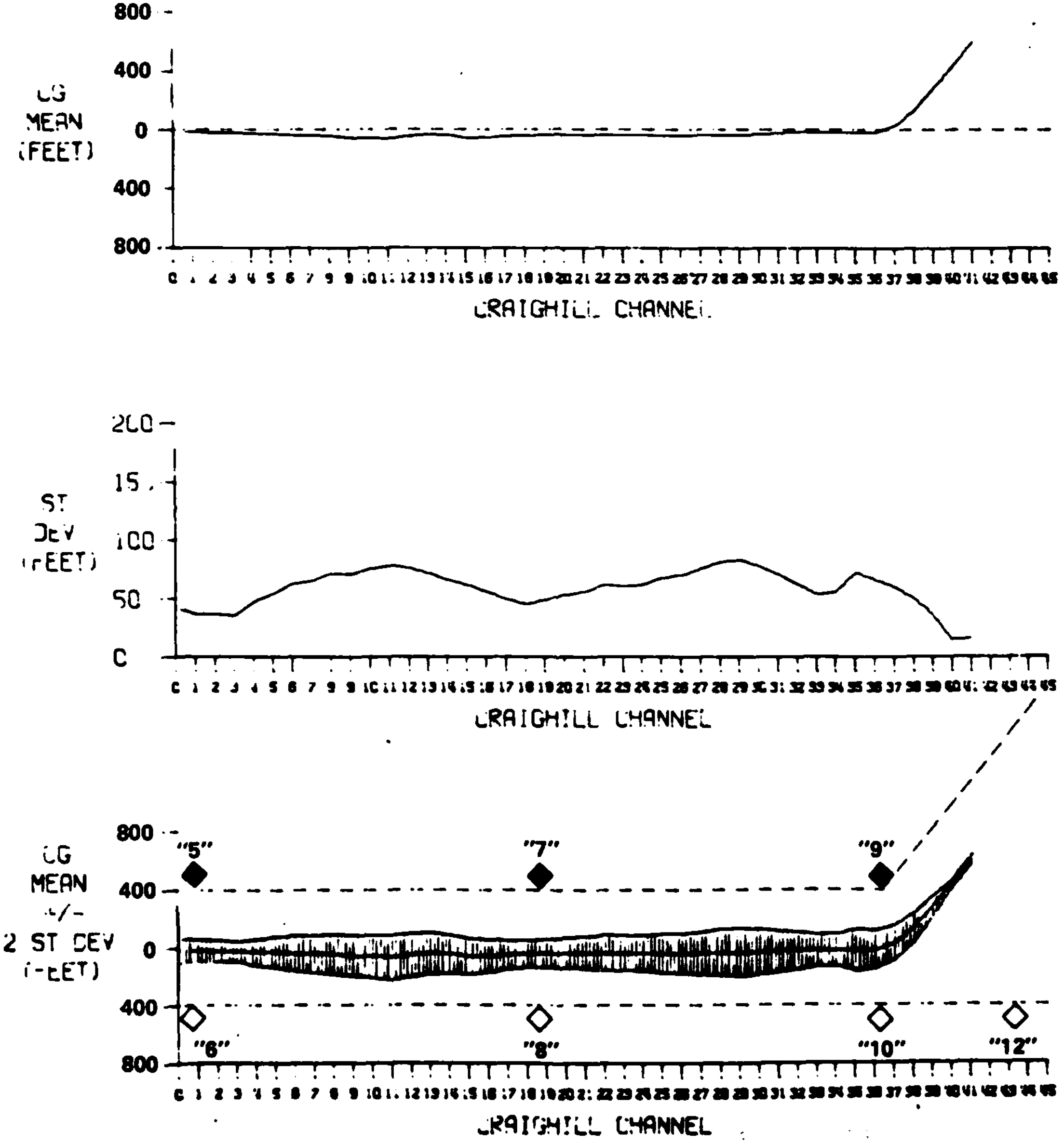


Figure D-4

AT-SEA

## SMALL SHIPS

CHESAPEAKE BAY

CRAIGHILL CHANNEL UPPER RANGE

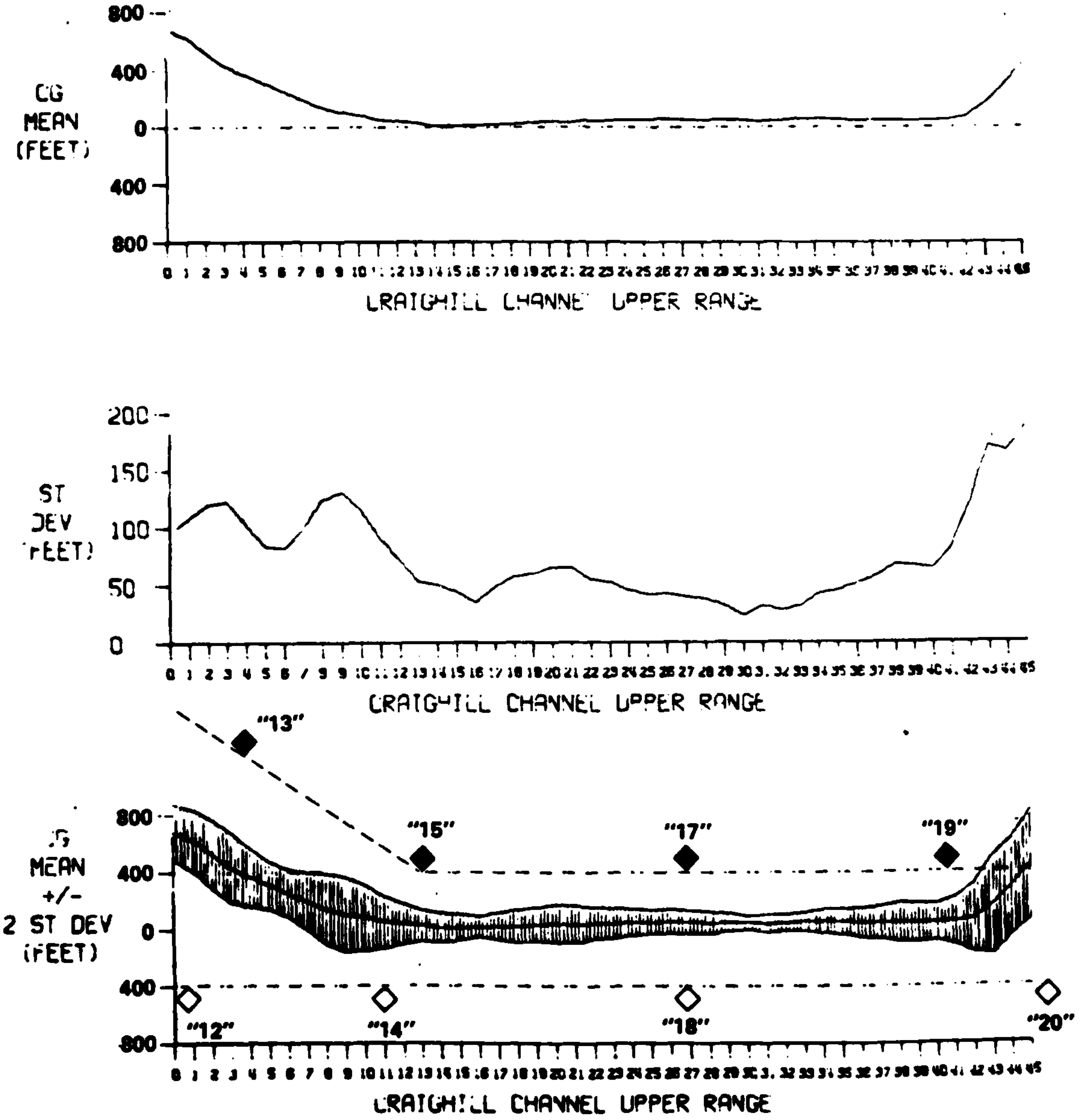


Figure D-5

### AT-SEA

# LARGE SHIPS

### CHESAPEAKE BAY

### CRAIGHILL CHANNEL UPPER RANGE

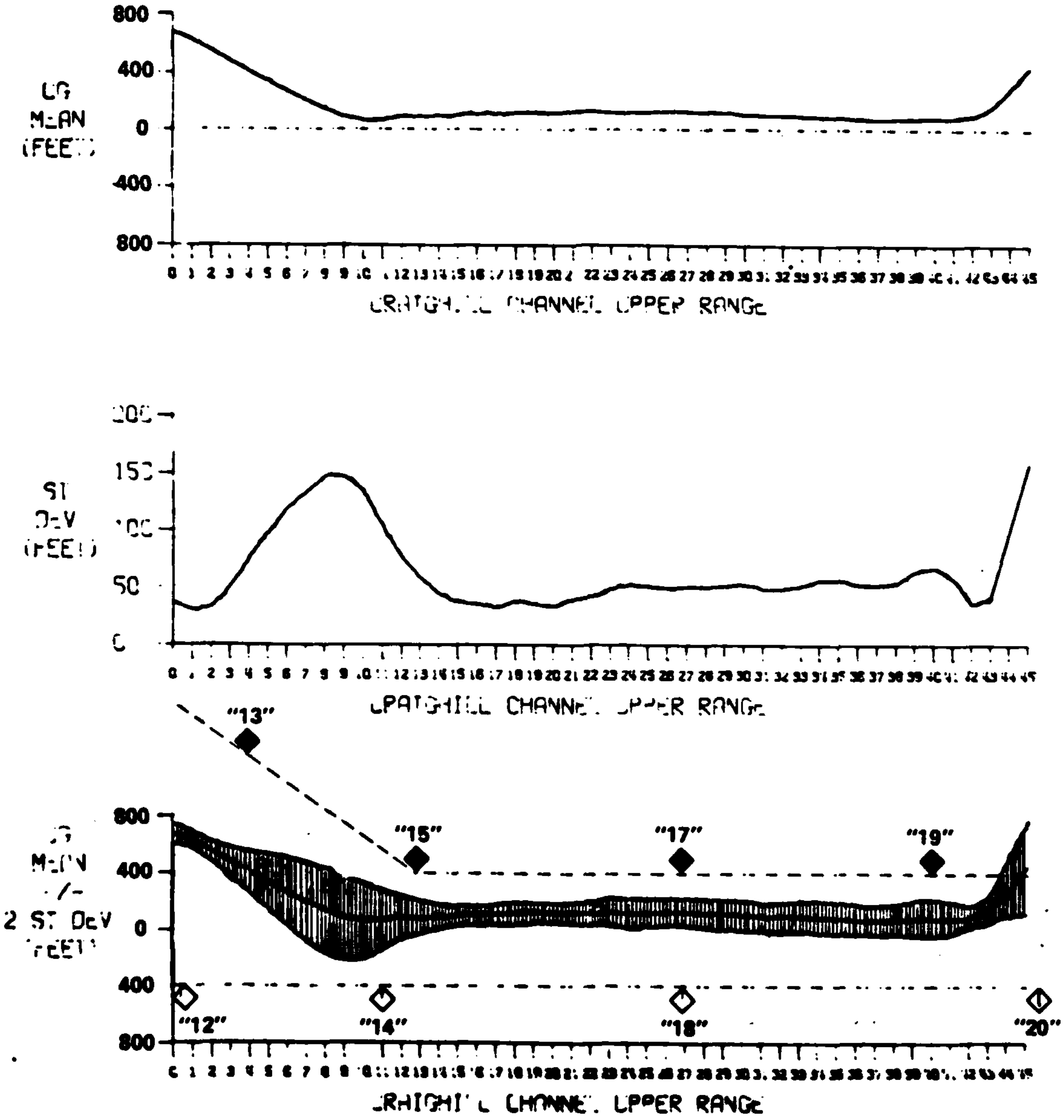


Figure D-6

### AT-SEA

# LARGE VS SMALL SHIPS

### CHESAPEAKE BAY
### CRAIGHILL CHANNEL

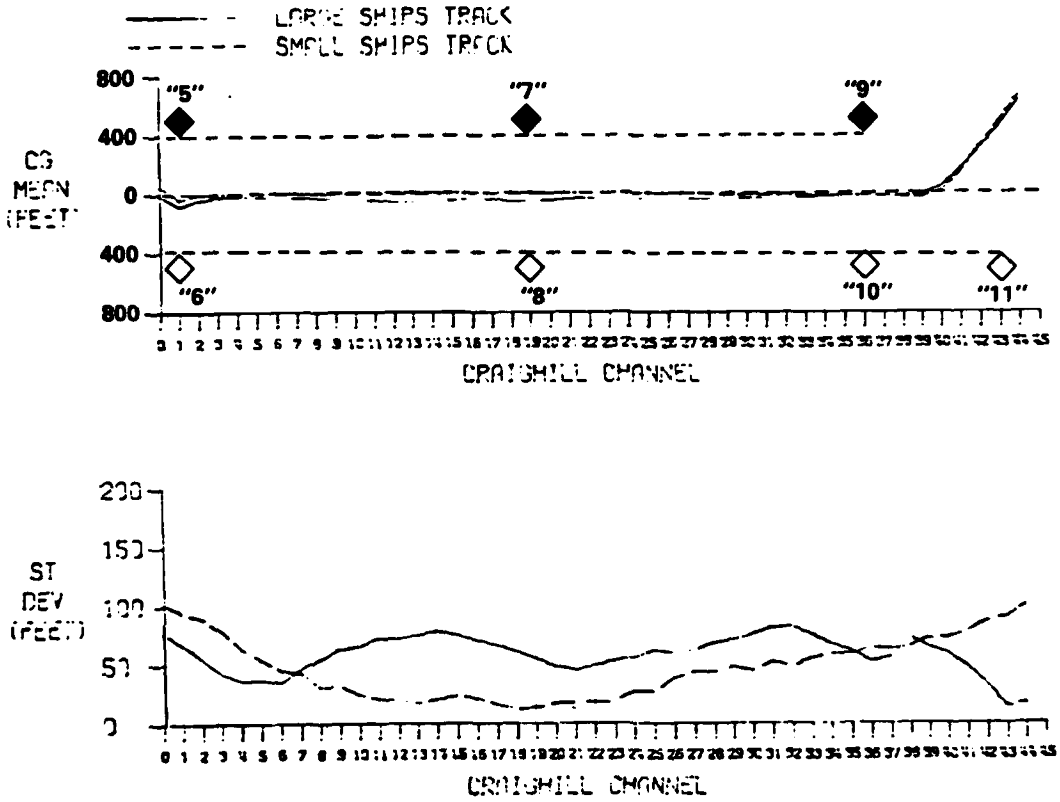


Figure D-7

### AT–SEA

# LARGE VS SMALL SHIPS

## CHESAPEAKE BAY
## CRAIGHILL CHANNEL UPPER RANGE

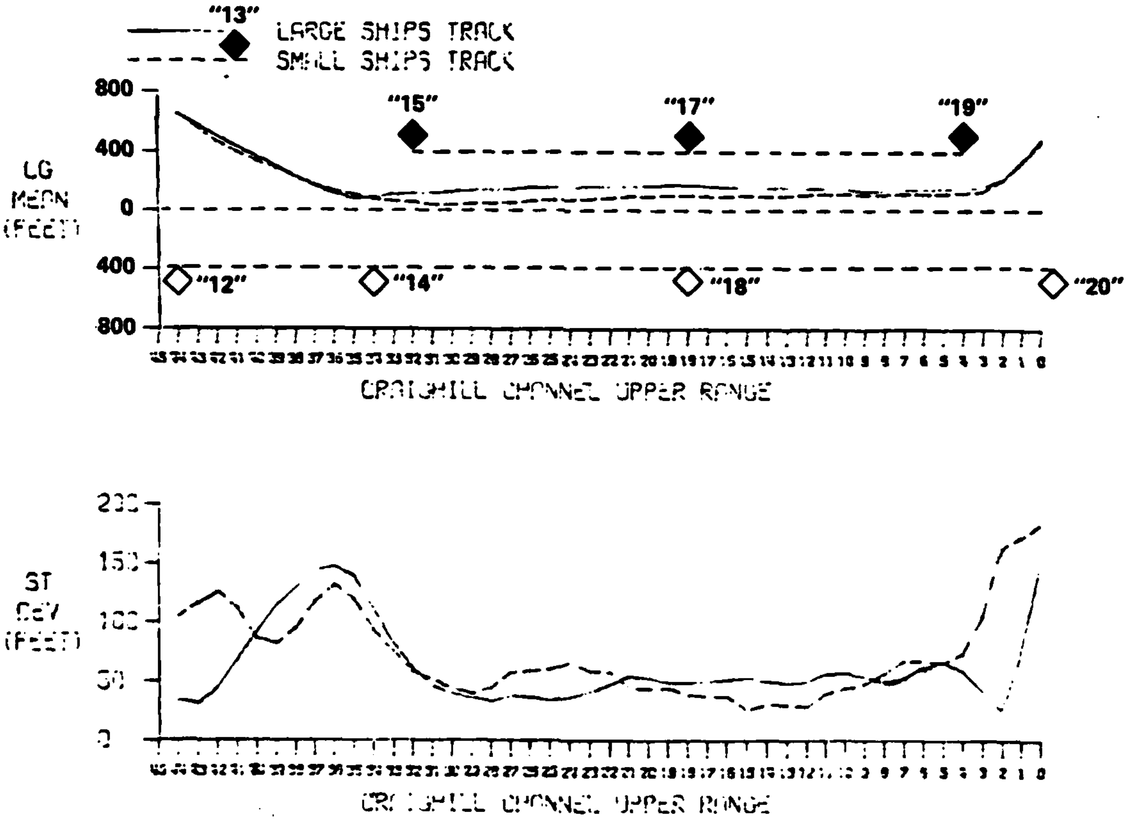


Figure D-8

### SIMULATOR

# 30,000 DWT TANKER

## CHESAPEAKE BAY

## CRAIGHILL ENTRANCE LEG

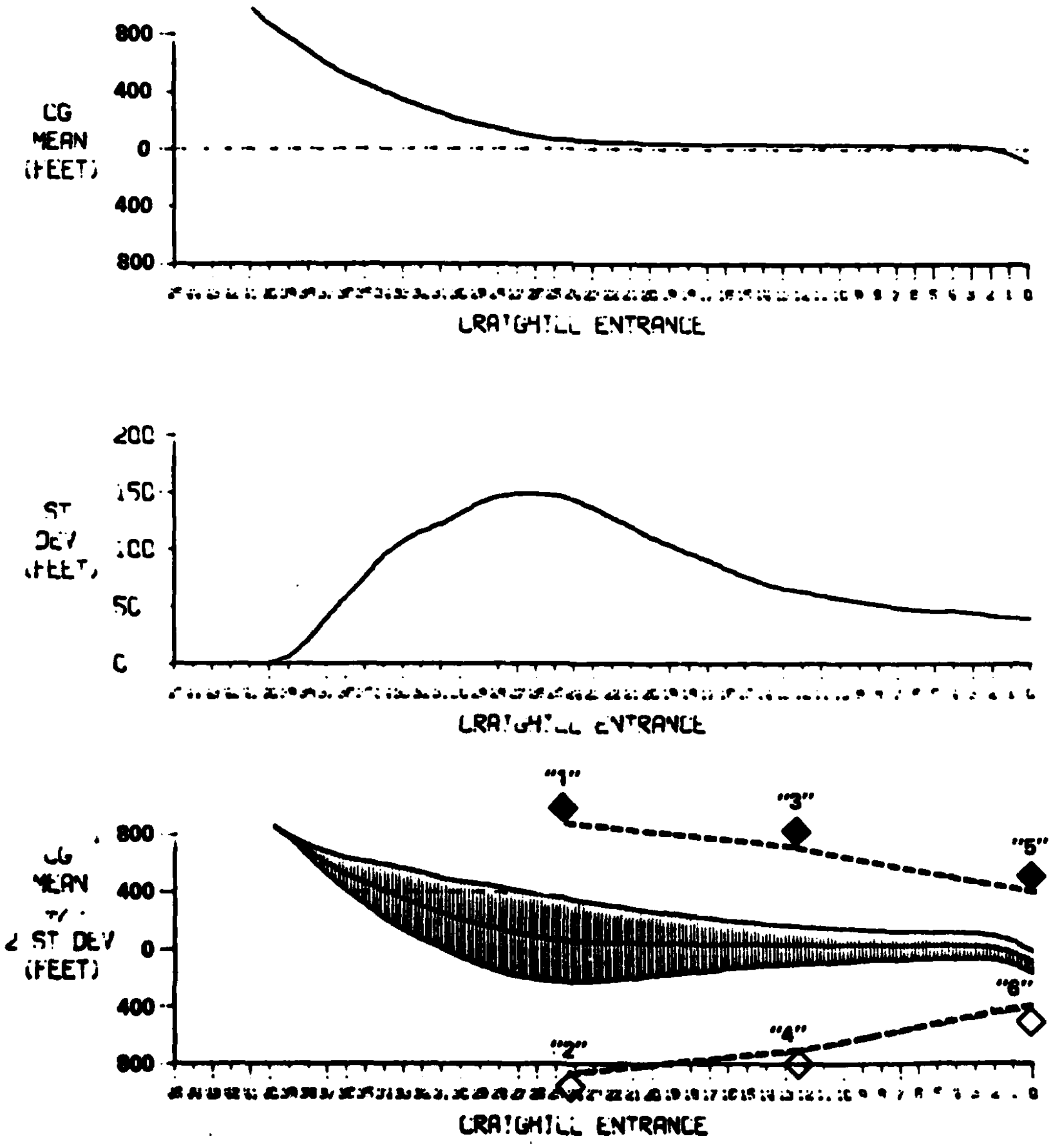


Figure D-9

## SIMULATOR

# 80,000 DWT TANKER

### CHESAPEAKE BAY

### CRAIGHILL ENTRANCE LEG

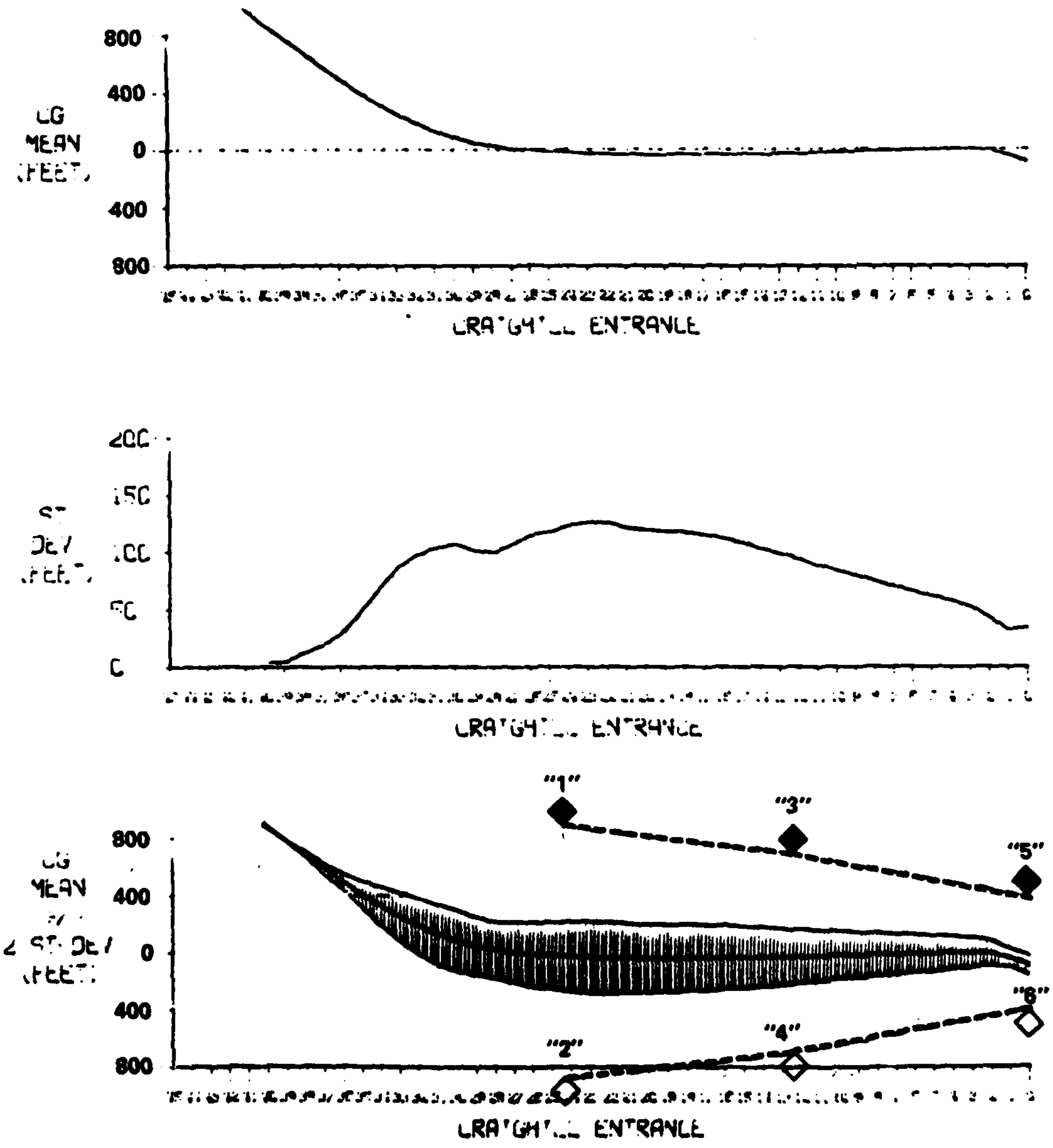


Figure D-10

SIMULATOR

# 30,000 DWT TANKER

CHESAPEAKE BAY

CRAIGHILL CHANNEL

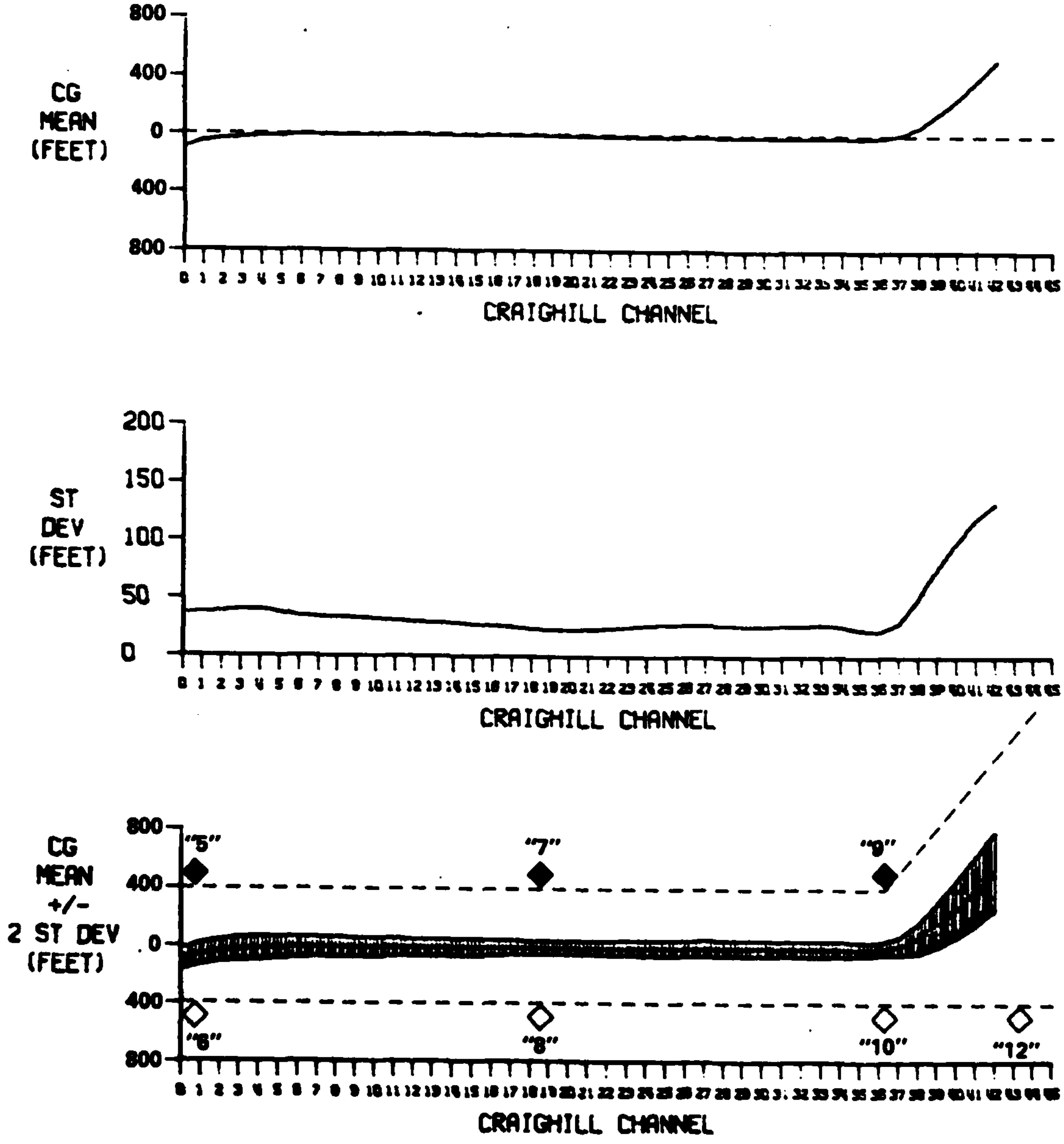


Figure D-11

### SIMULATOR

# 80,000 DWT TANKER

### CHESAPEAKE BAY
### CRAIGHILL CHANNEL

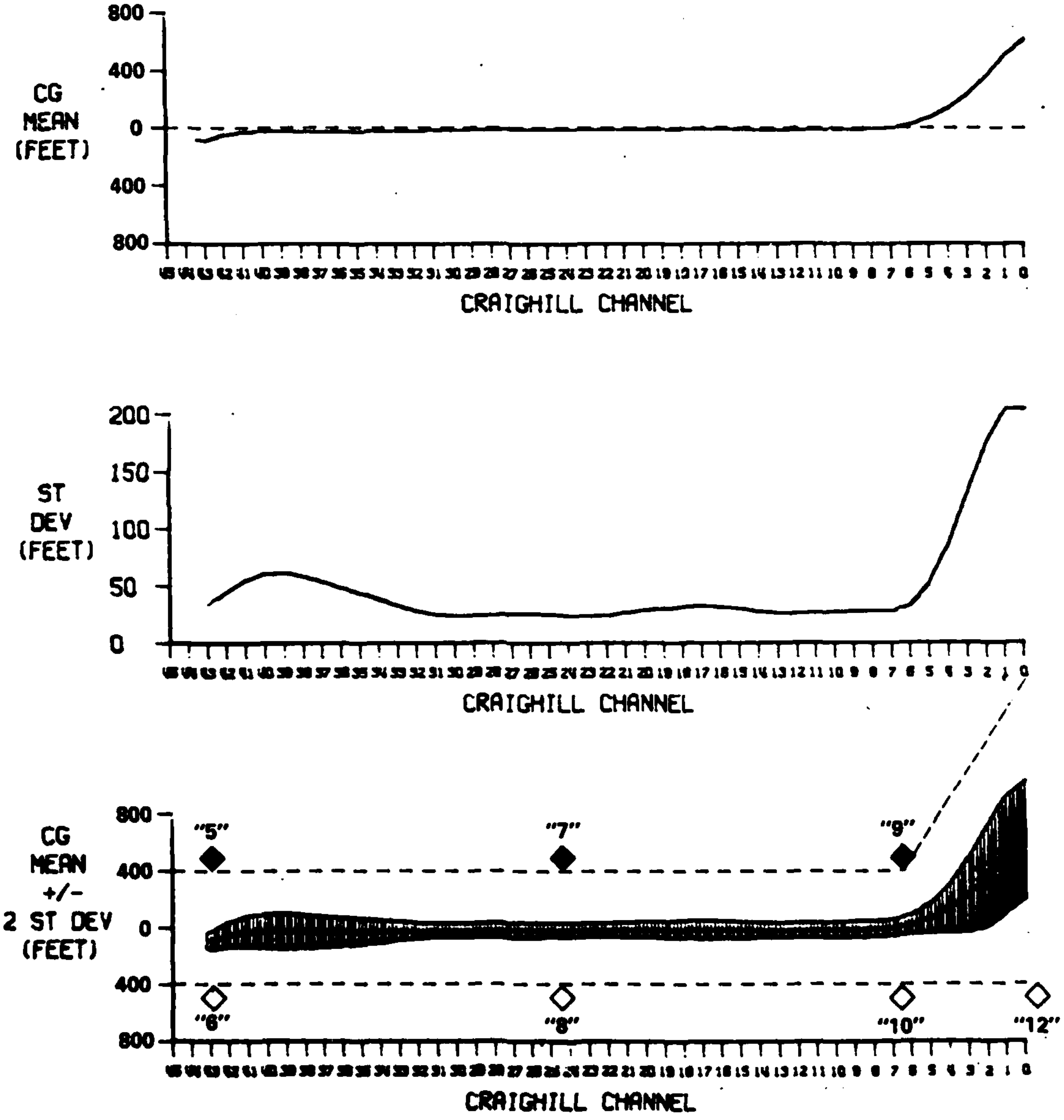


Figure D-12

SIMULATOR

# 80,000 DWT TANKER

CHESAPEAKE BAY

CRAIGHILL CHANNEL UPPER RANGE

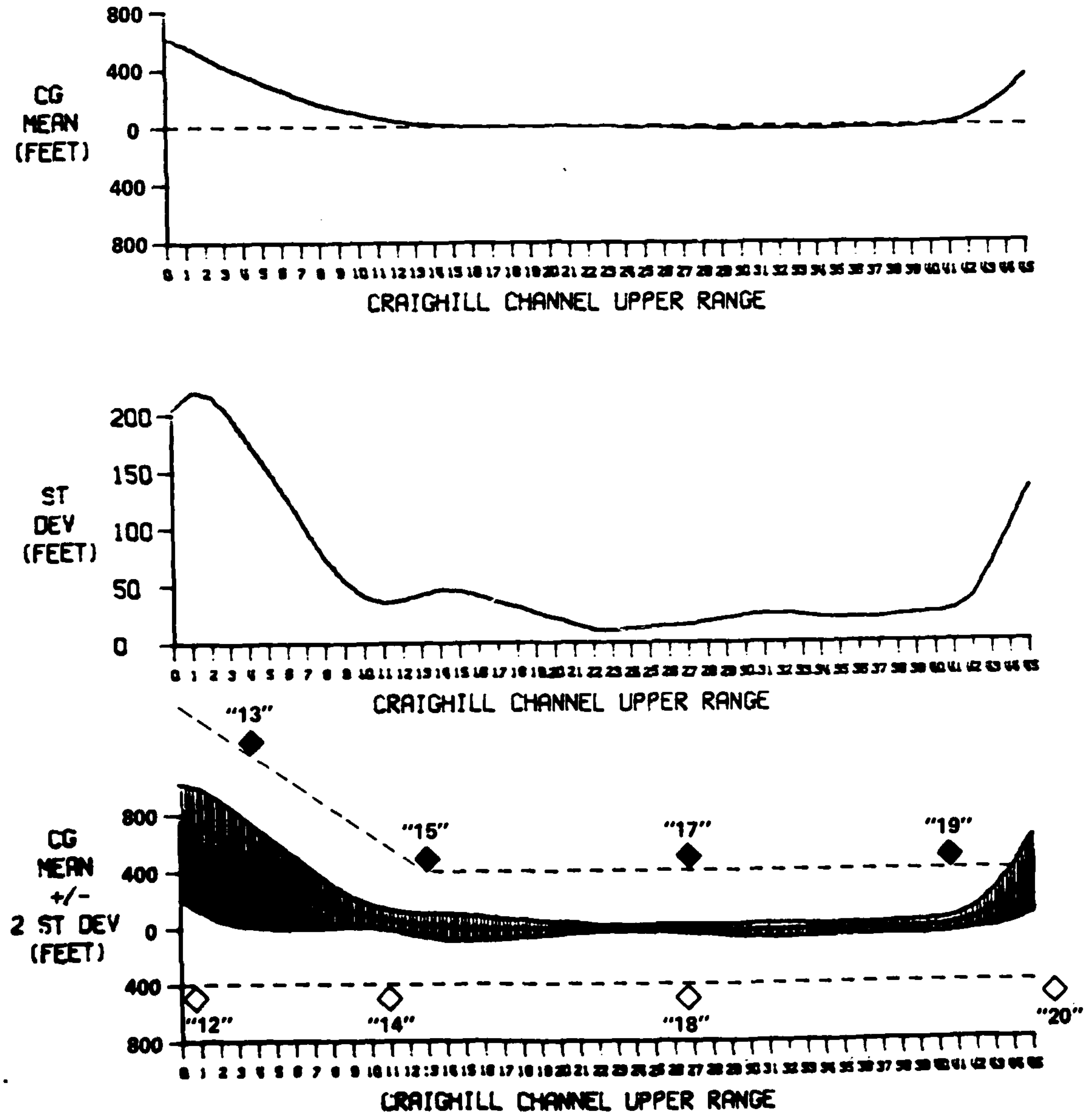


Figure D-14

### SIMULATOR

# 30,000 DWT TANKER

### CHESAPEAKE BAY

### CRAIGHILL CHANNEL UPPER RANGE

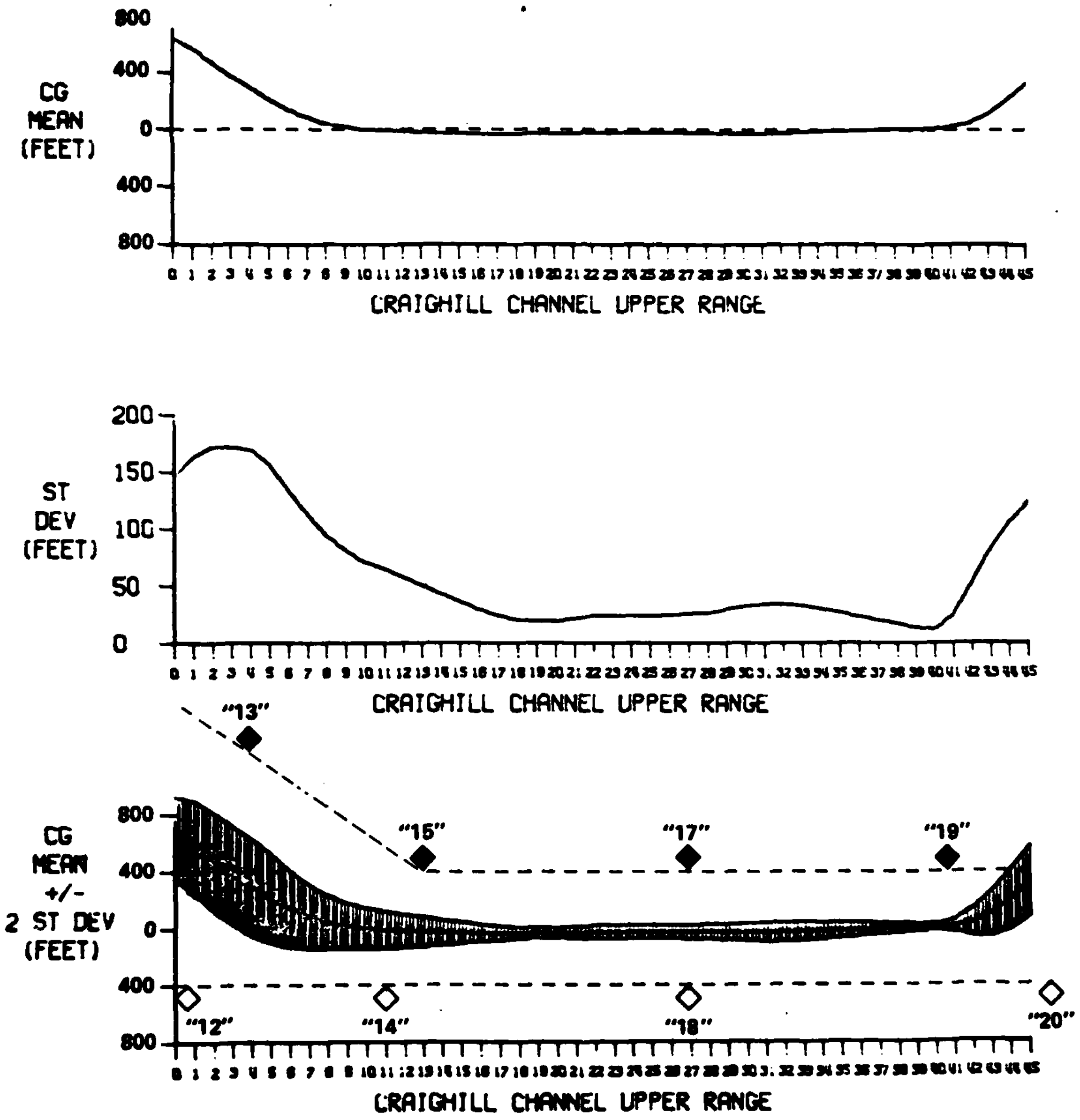


Figure D-13

### SIMULATOR

## 80,000 DWT TANKER

### CHESAPEAKE BAY

### CRAIGHILL CHANNEL UPPER RANGE

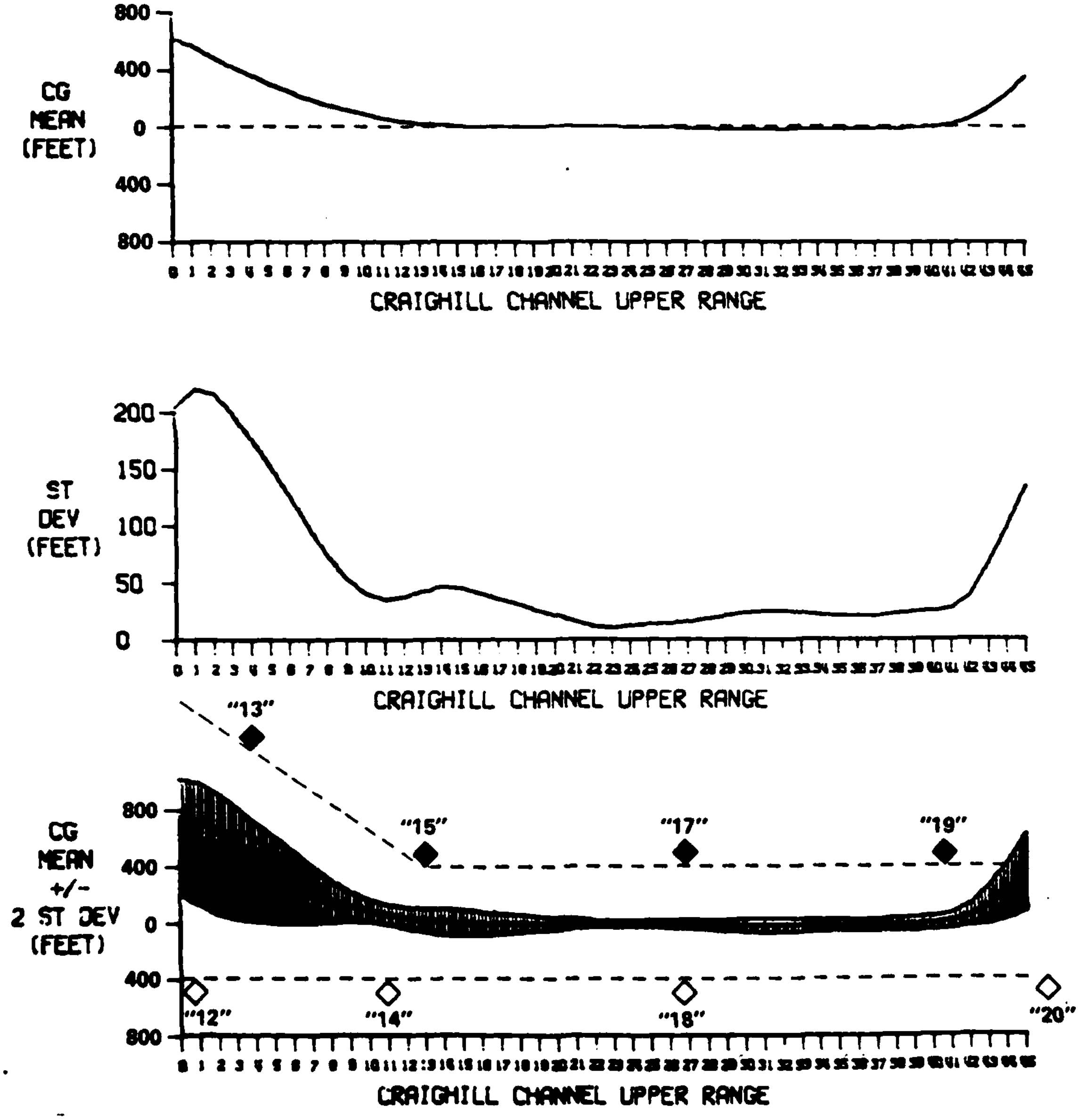


Figure D-14

### SIMULATOR

# 30,000 DWT VS 80,000 DWT TANKER

## CHESAPEAKE BAY
## CRAIGHILL ENTRANCE LEG

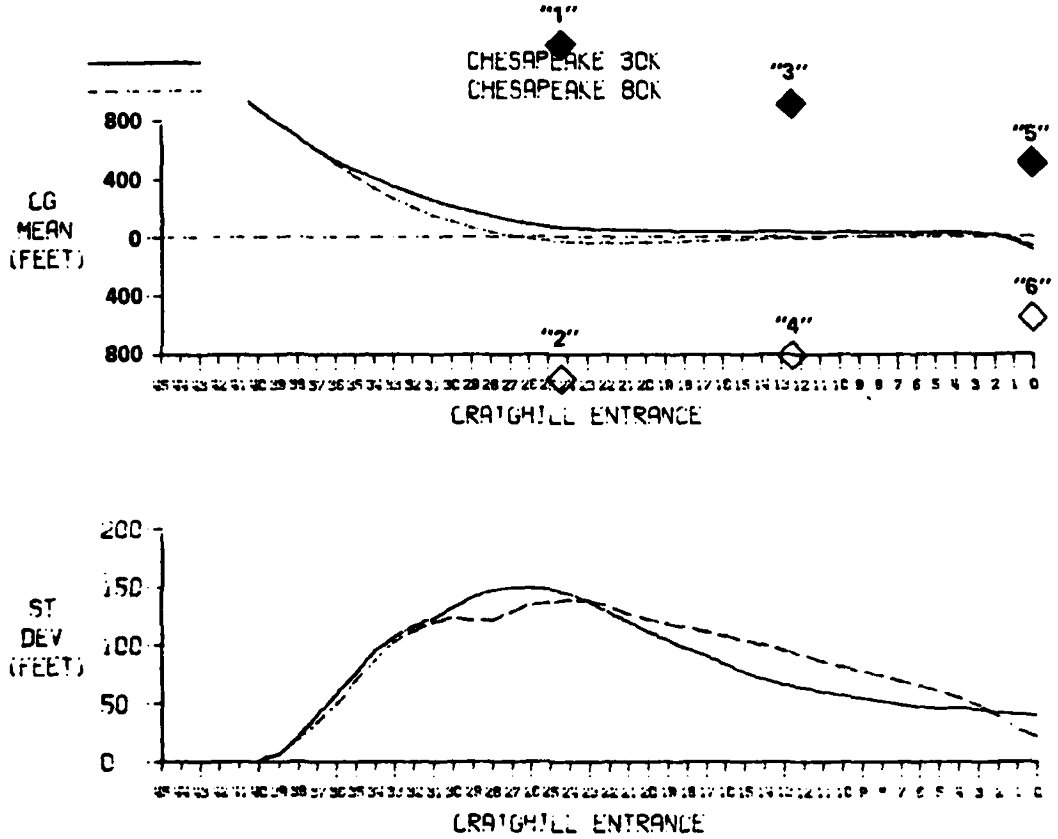


Figure D-15

SIMULATOR

# 30,000 VS 80,000 DWT TANKER

## CHESAPEAKE BAY

## CRAIGHILL CHANNEL

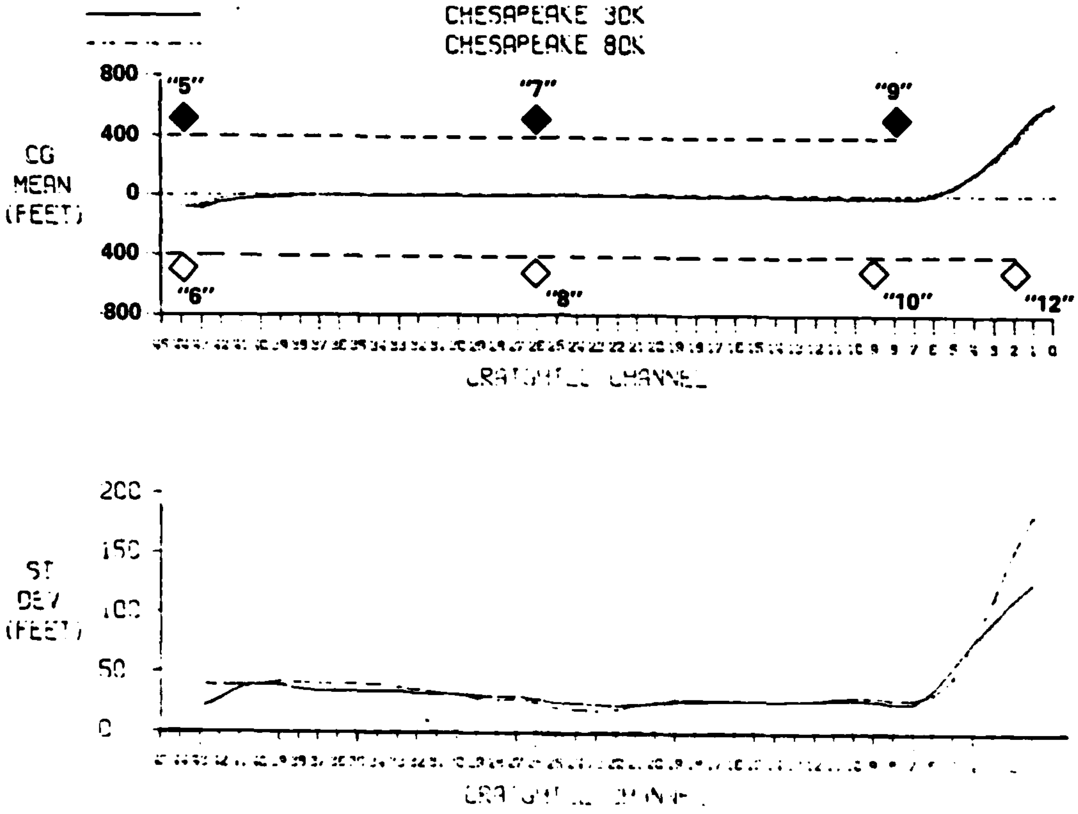


Figure D-16

SIMULATOR

# 30,000 VS 80,000 DWT TANKER

## CHESAPEAKE BAY
## CRAIGHILL CHANNEL UPPER RANGE

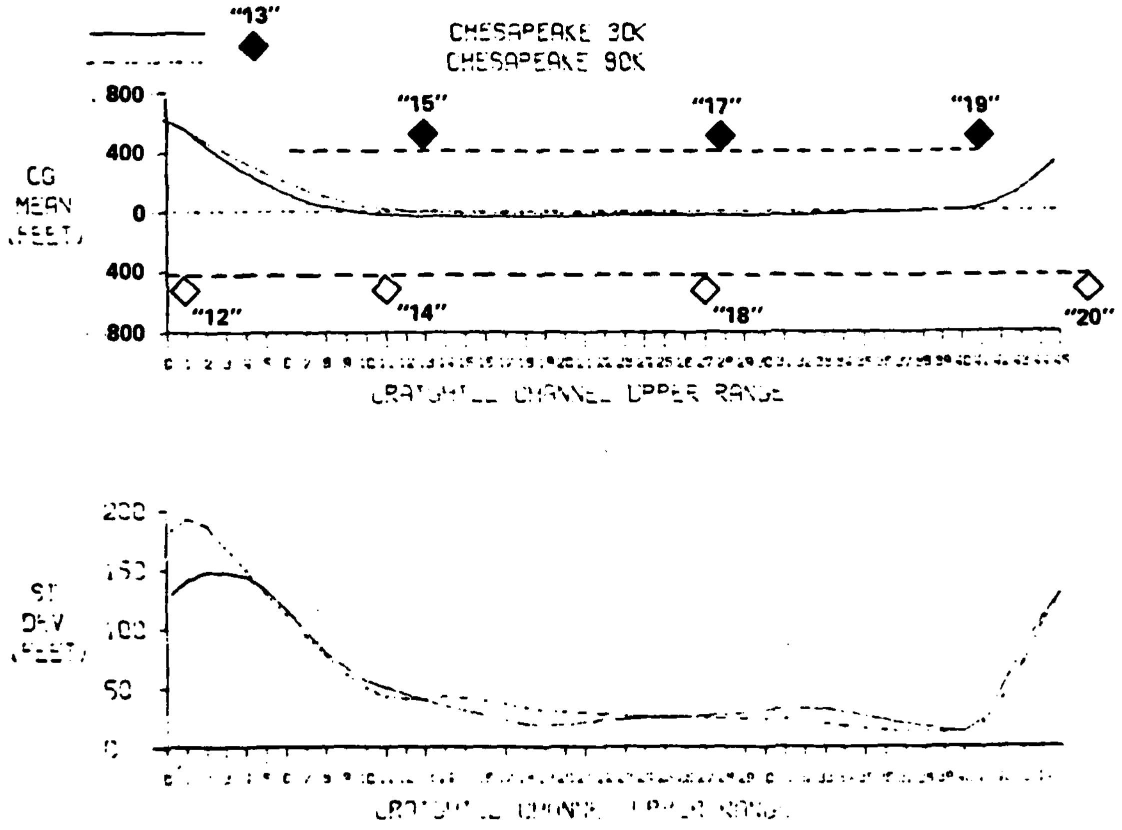


Figure D-17

# APPENDIX E

## NARRAGANSETT BAY SIMULATOR AND AT-SEA PILOTED PERFORMANCE TRACK PLOTS

This appendix contains performance track plots of the at-sea and simulated data collected in the Narragansett Bay. Figures E-1 and E-2 show the charts and the location of the at-sea data collection and Figures E-3 and E-4 show the charts of the simulated scenarios. The at-sea data plotted shows performance for four daytime and four nighttime runs with tankers of approximately 30,000 dwt. The simulator data plotted shows performance for eight daytime and eight nighttime runs with the 30,000 dwt tanker model.

Table E-1 identifies the conditions represented by the track plots. The plots are grouped by location of data.

There are two types of track plots in this section. A combined plot is illustrated by Figure E-5. It consists of a series of three plots for one of the two channel legs. The axis for the abscissa is scaled so that one unit of alongtrack distance represents 475 feet (5/64 nm). The top plot displays the crosstrack mean of the center of gravity of the ships as they transit the channel and the middle plot displays the crosstrack standard deviation. The bottom plot is a combined plot showing the crosstrack mean and an envelop encompassing two standard deviations to either side, an area within which performance is expected to occur 95 percent of the time. A comparison plot is illustrated by Figure E-15. For each comparison there are two sets of axes, one showing the means of two conditions and one showing the crosstrack standard deviation as the performance measures. Data is plotted as a continuous unbroken line and a dotted line to distinguish the experimental conditions from each other.

Statistical tests were used to test the differences in performance at each data line, to determine if any differences between conditions were statistically significant. The means were compared using a t-test. The symbols along the axis of the mean plot indicate a difference at the 0.10 level of significance. The standard deviations were compared as variances using an F test. The symbols along the axis of the standard deviation plot also indicate a difference at the 0.10 level of significance.

Plot Notes:

1. On the plots, buoys and lights are position for the purpose of illustration and may not appear in their exact charted location.

2. Aids to Navigation symbols

   - ◇ nun buoy
   - ◆ can buoy
   - ◇⃝ red, lighted buoy
   - ◆⃝ black, lighted buoy
   - ⊙ light beacon

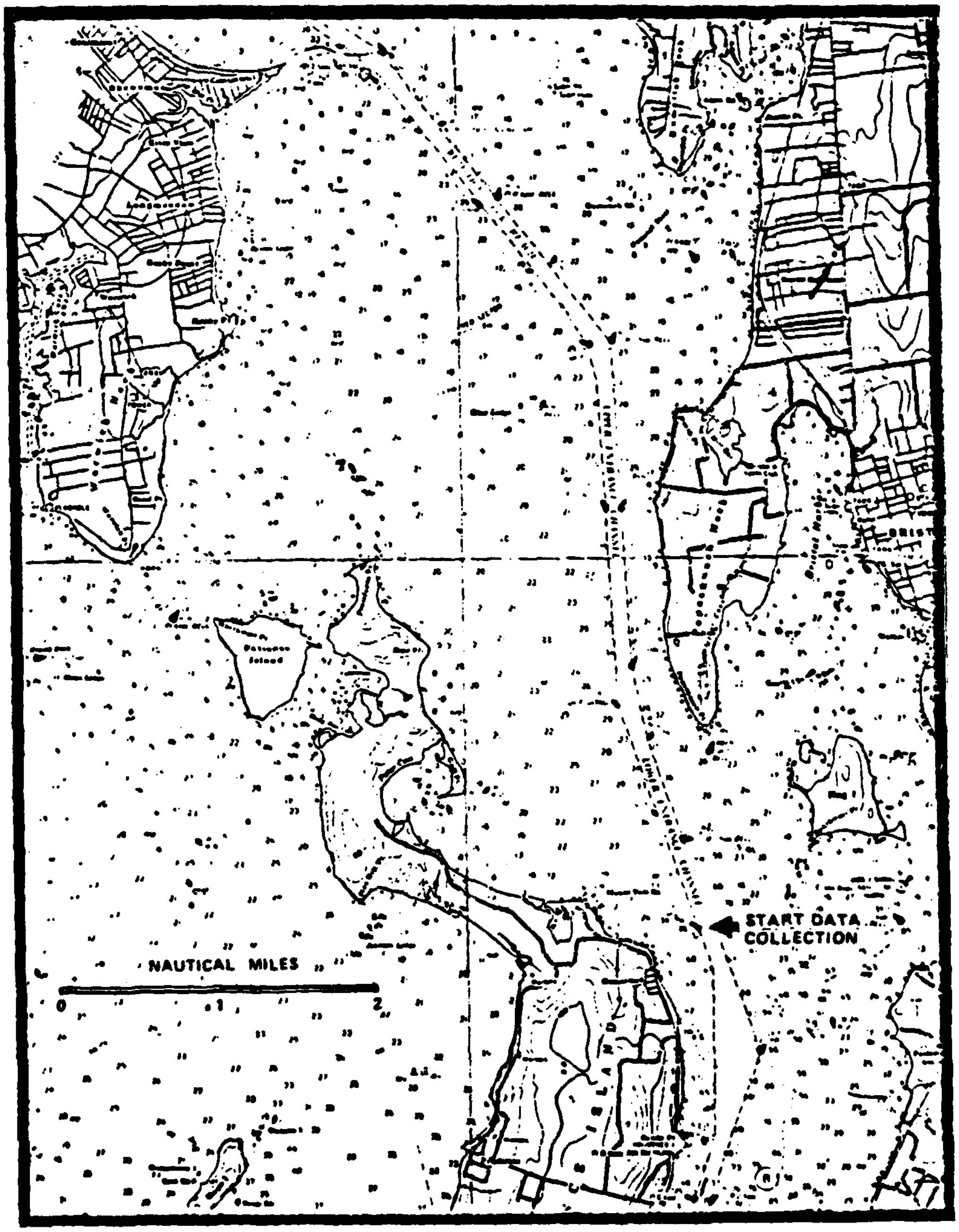


Figure E-1. Narragansett Bay Where At-Sea Data Was Collected
Chart No. 13221 of March 28, 1981

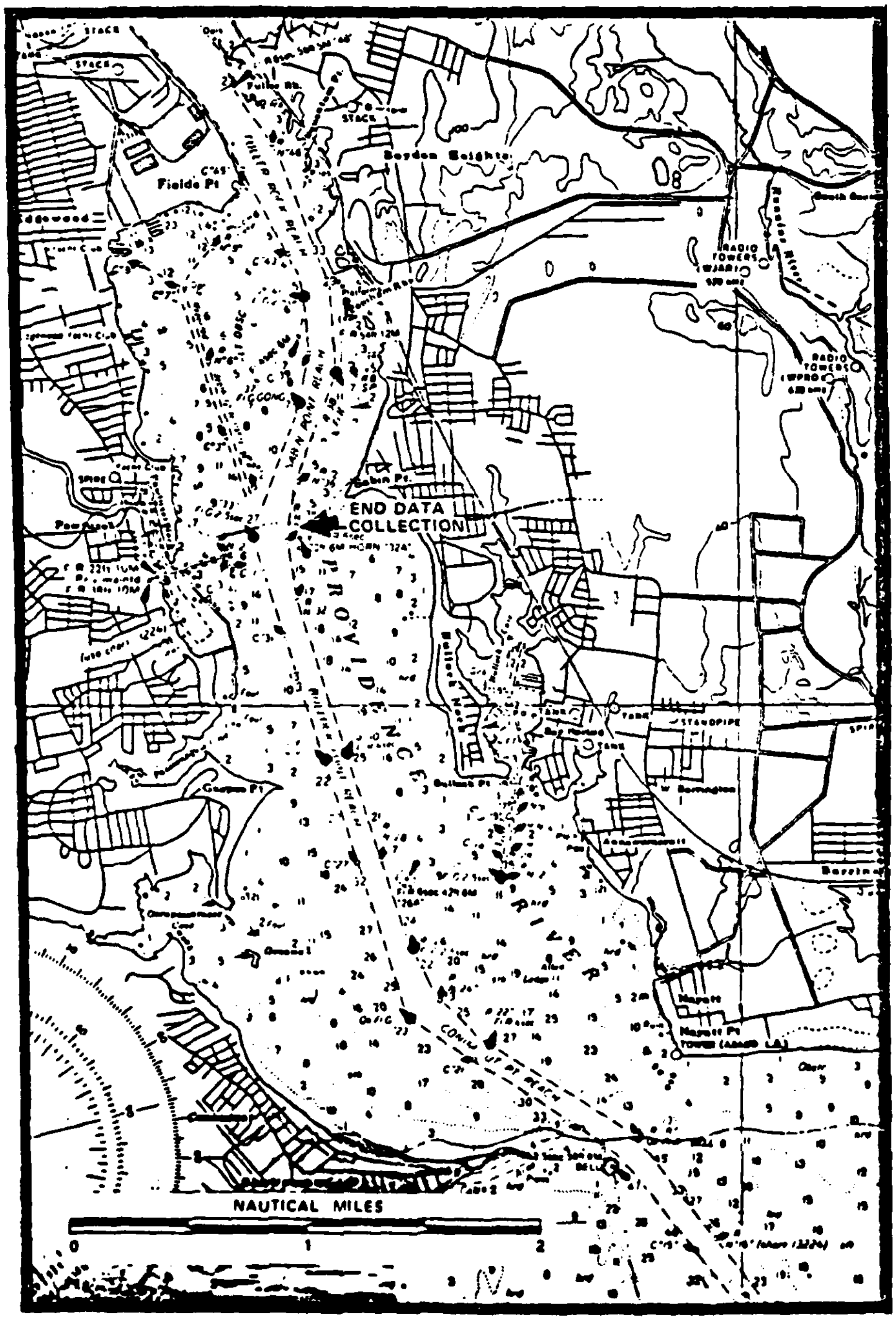
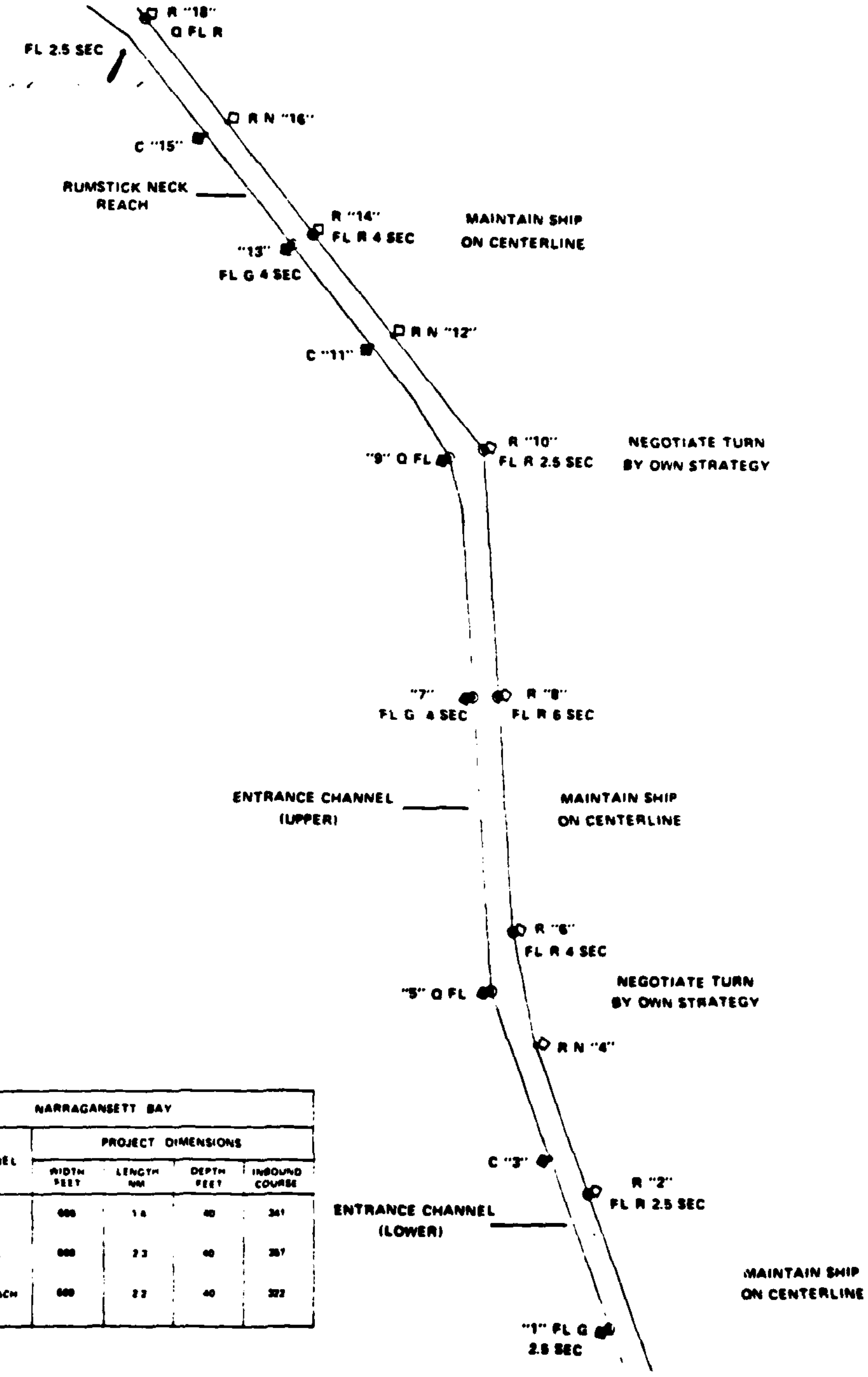
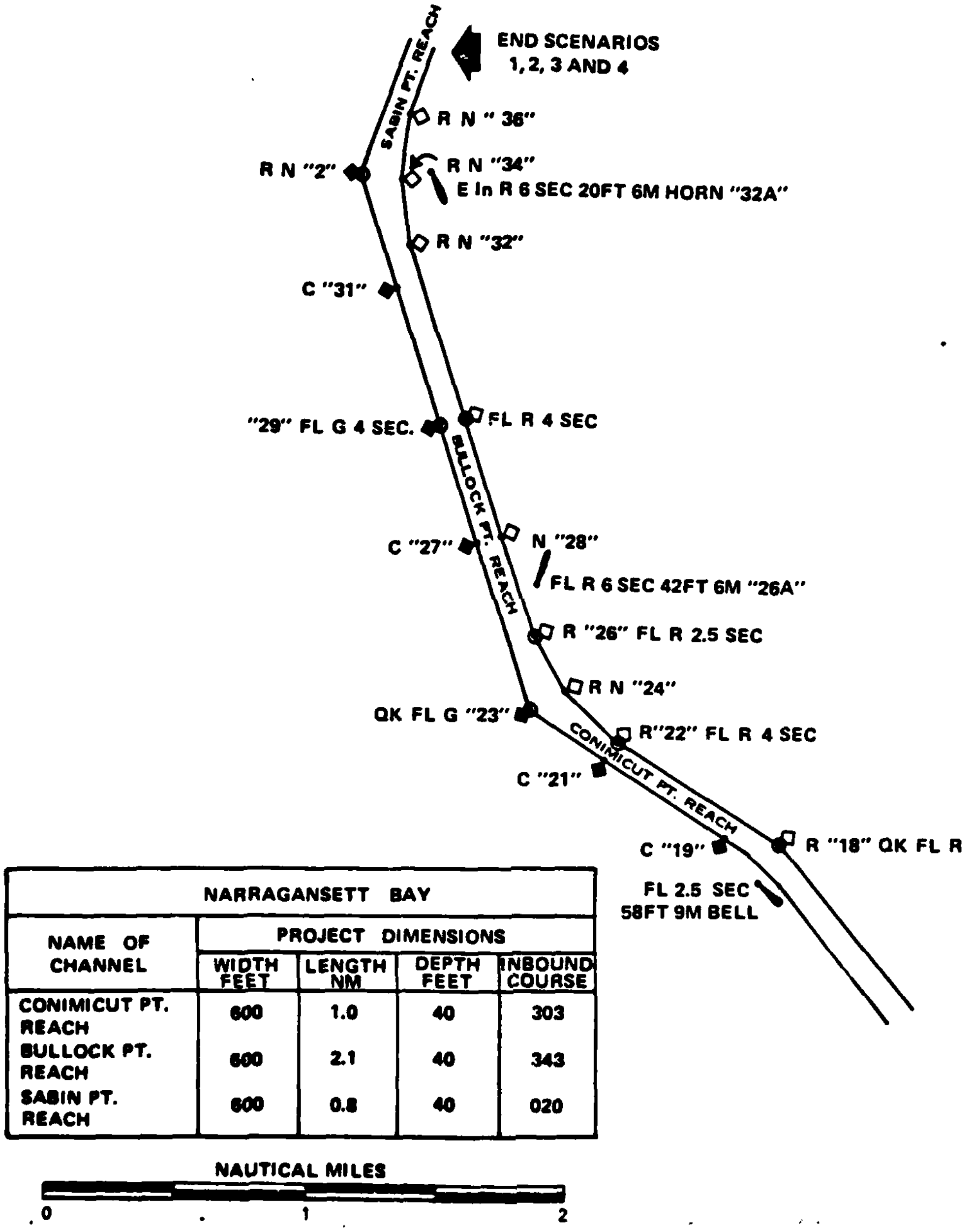


Figure E-2. Narragansett Bay Where at Sea Data Was Collected Chart No. 13221 of March 28, 1981 (Continued)



| NARRAGANSETT BAY | | | | |
|---|---|---|---|---|
| NAME OF CHANNEL | PROJECT DIMENSIONS | | | |
| | WIDTH FEET | LENGTH NM | DEPTH FEET | INBOUND COURSE |
| ENTRANCE CHANNEL LOWER | 600 | 1.6 | 40 | 341 |
| ENTRANCE CHANNEL UPPER | 600 | 2.3 | 40 | 357 |
| RUMSTICK NECK REACH | 600 | 2.2 | 40 | 322 |

Figure E-3. Lower Narragansett Channel Chart for Familiarization and Validation (Scenarios 1, 2, 3, and 4)



| NARRAGANSETT BAY | | | | |
|---|---|---|---|---|
| NAME OF CHANNEL | PROJECT DIMENSIONS | | | |
| | WIDTH FEET | LENGTH NM | DEPTH FEET | INBOUND COURSE |
| CONIMICUT PT. REACH | 600 | 1.0 | 40 | 303 |
| BULLOCK PT. REACH | 600 | 2.1 | 40 | 343 |
| SABIN PT. REACH | 600 | 0.8 | 40 | 020 |

Figure E-4. Upper Narragansett Channel Chart for Familiarization and Validation (Scenarios 1, 2, 3, and 4)

TABLE E-1. NARRAGANSETT BAY TRACK PLOTS

| Figure | Type of Data | Light | Channel | Plot |
|---|---|---|---|---|
| E-5 | At Sea | Day | Lower Entrance | Combined |
| E-6 | At Sea | Night | Lower Entrance | Combined |
| E-7 | At Sea | Day | Upper Entrance | Combined |
| E-8 | At Sea | Night | Upper Entrance | Combined |
| E-9 | At Sea | Day | Rumstick Neck Reach | Combined |
| E-10 | At Sea | Night | Rumstick Neck Reach | Combined |
| E-11 | At Sea | Day | Conimicut Point Reach | Combined |
| E-12 | At Sea | Night | Conimicut Point Reach | Combined |
| E-13 | At Sea | Day | Bullock Point Reach | Combined |
| E-14 | At Sea | Night | Bullock Point Reach | Combined |
| E-15 | At Sea | Day/Night | Lower and Upper Entrance | Comparison |
| E-16 | At Sea | Day/Night | Rumstick Neck & Conimicut Point | Comparison |
| E-17 | At Sea | Day/Night | Bullock Point Reach | Comparison |
| E-18 | Simulator | Day | Lower and Upper Entrance | Combined |
| E-19 | Simulator | Night | Lower and Upper Entrance | Combined |
| E-20 | Simulator | Day | Rumstick Neck & Conimicut Point | Combined |
| E-21 | Simulator | Night | Rumstick Neck & Conimicut Point | Combined |
| E-22 | Simulator | Day | Bullock Point and Sabin Point Reaches | Combined |
| E-23 | Simulator | Night | Bullock Point and Sabin Point Reaches | Combined |
| E-24 | Simulator | Day/Night | Lower and Upper Entrance | Comparison |
| E-25 | Simulator | Day/Night | Rumstick Neck & Conimicut Point | Comparison |
| E-26 | Simulator | Day/Night | Bullock Point and Sabin Point Reaches | Comparison |

## AT–SEA
## DAYTIME TRANSIT
## NARRAGANSETT BAY
## LOWER ENTRANCE CHANNEL

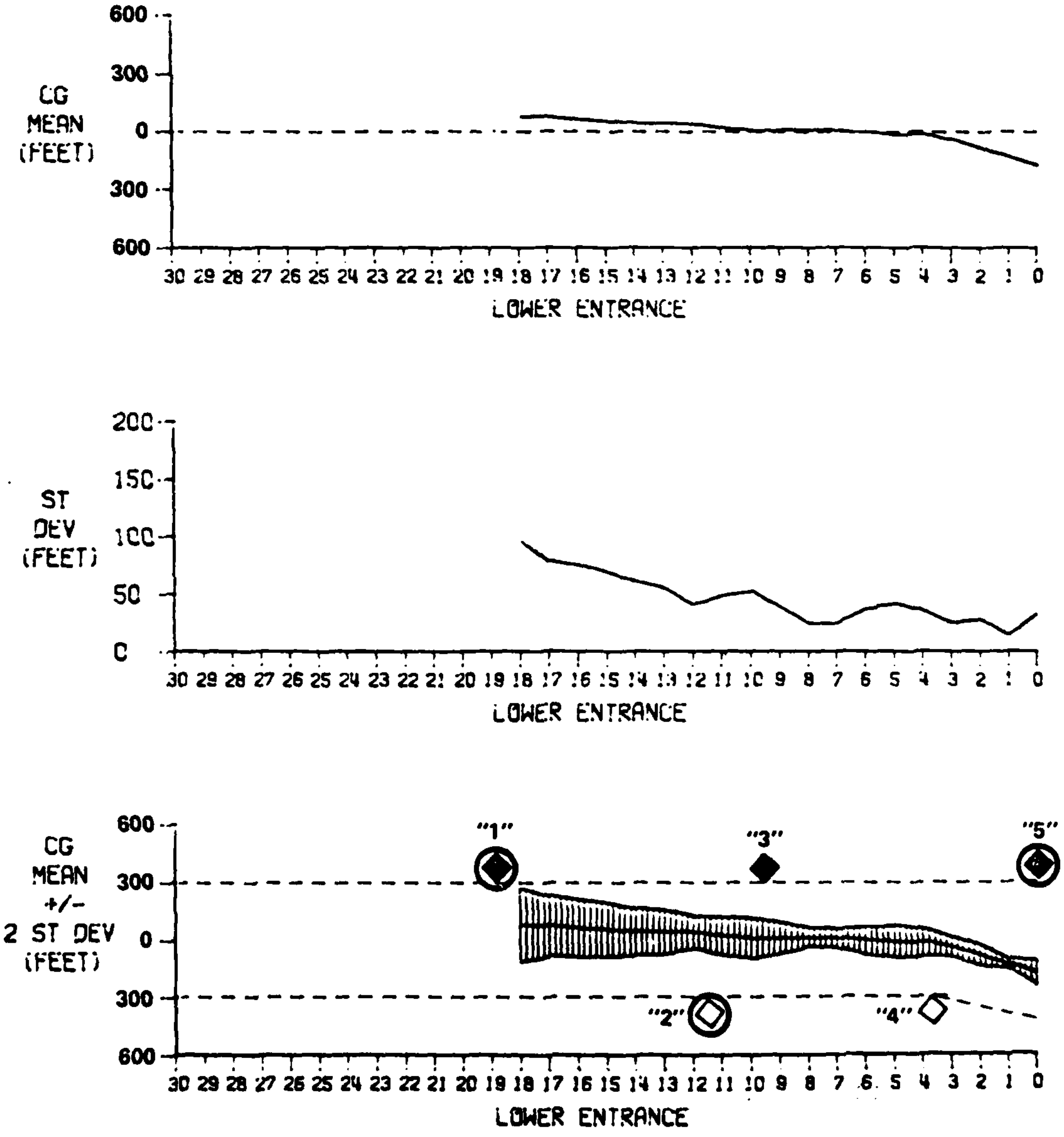


Figure E-5

## AT–SEA
## NIGHTTIME TRANSIT
## NARRAGANSETT BAY
## LOWER ENTRANCE CHANNEL

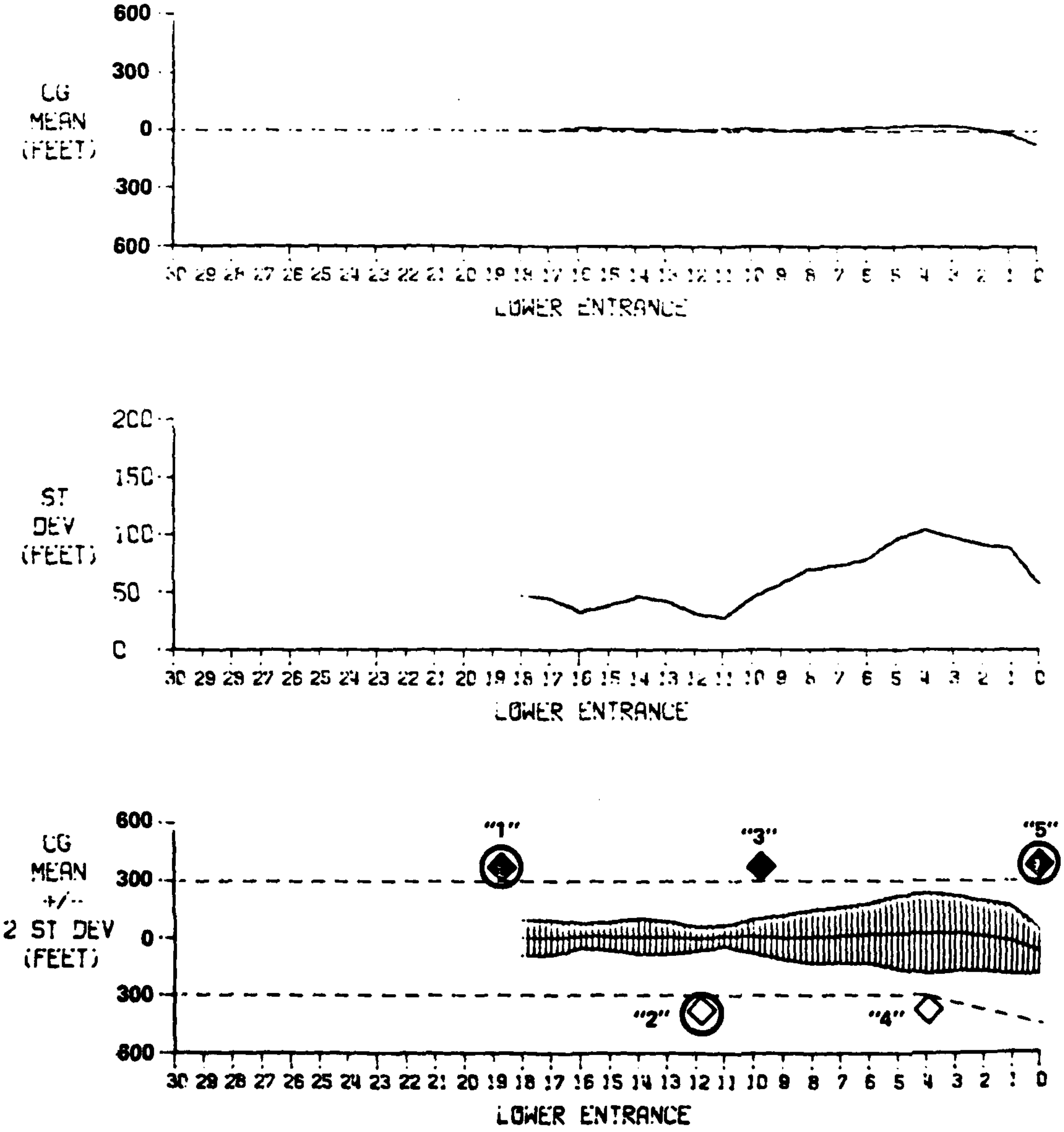


Figure E-6

# AT–SEA
# DAYTIME TRANSIT
# NARRAGANSETT BAY
# UPPER ENTRANCE CHANNEL

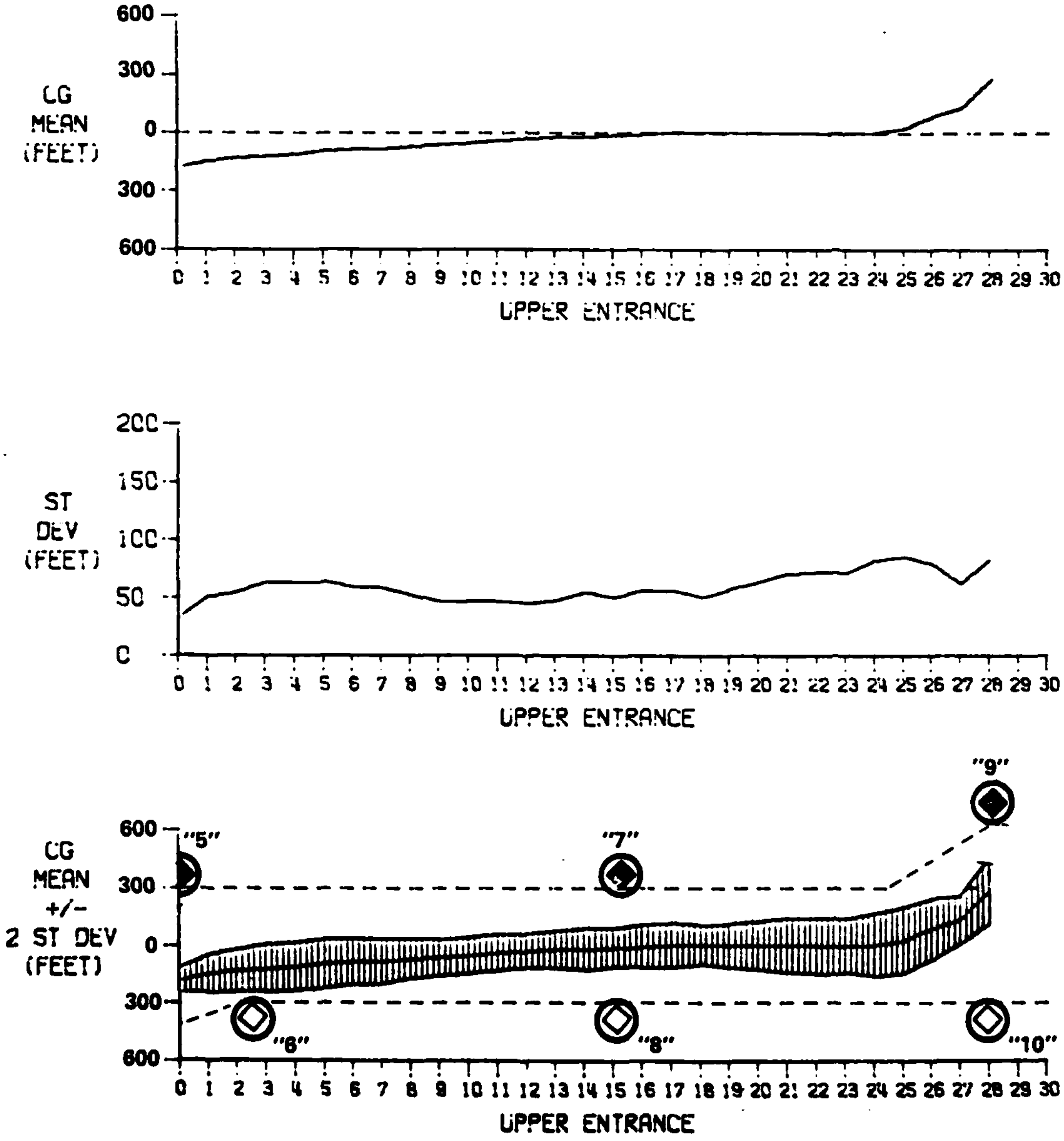


Figure E-7

# AT–SEA
# NIGHTTIME TRANSIT
# NARRAGANSETT BAY
# UPPER ENTRANCE CHANNEL

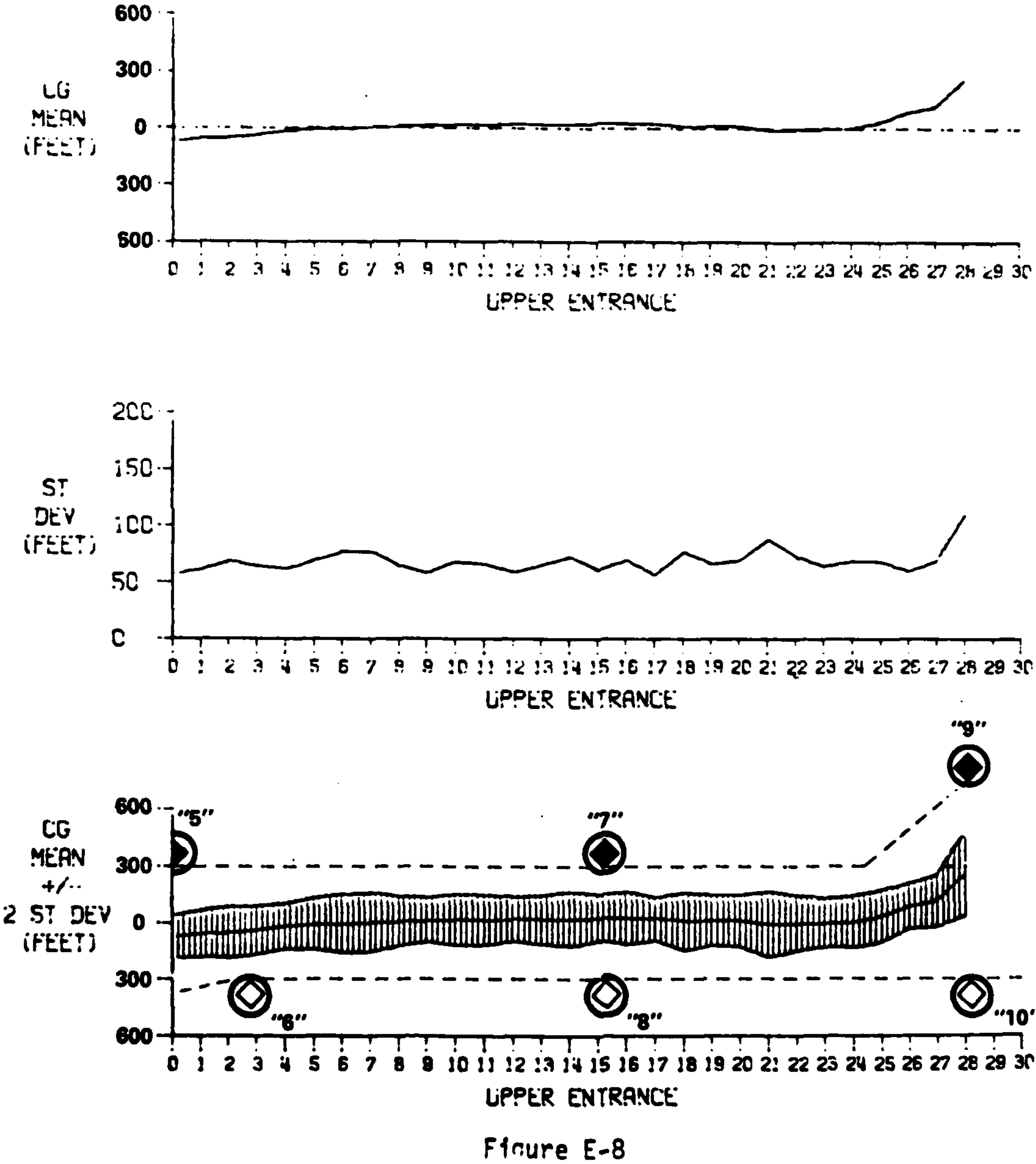


Figure E-8

**AT–SEA**

**DAYTIME TRANSIT**

**NARRAGANSETT BAY**

**RUMSTICK NECK REACH**

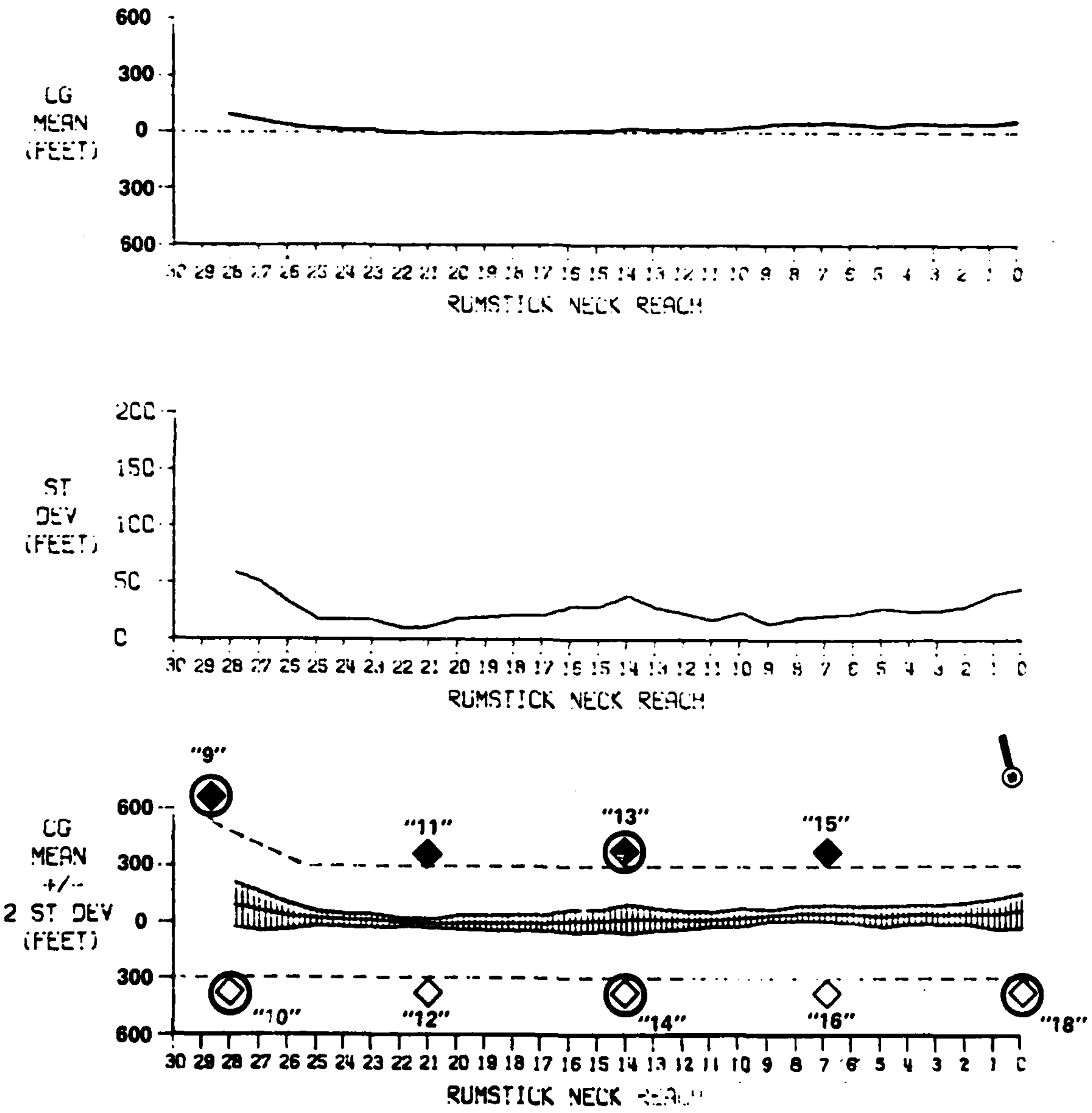


Figure E-3

# AT–SEA
# NIGHTTIME TRANSIT
# NARRAGANSETT BAY
# RUMSTICK NECK REACH

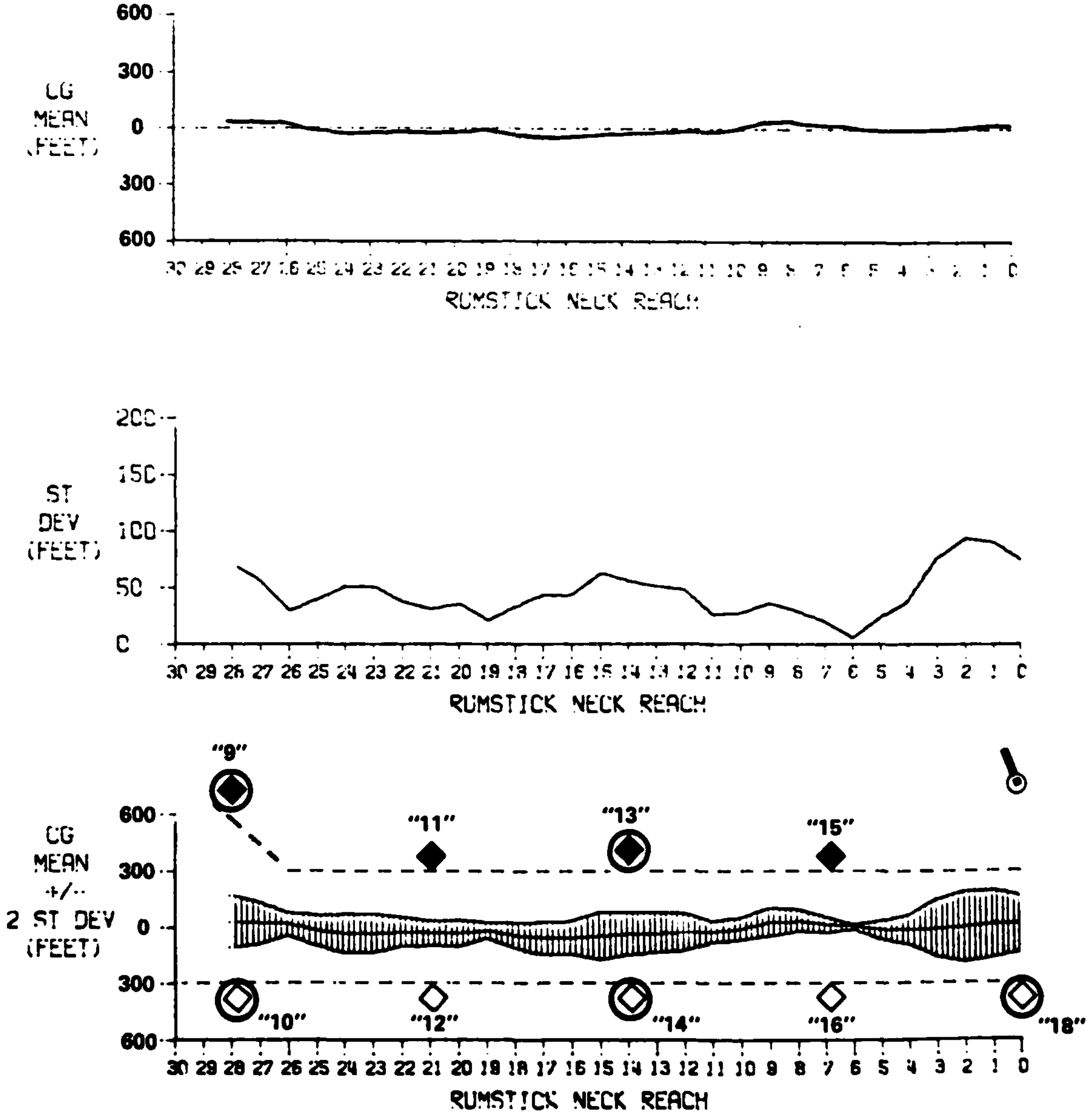


Figure E-10

# AT–SEA
# DAYTIME TRANSIT
# NARRAGANSETT BAY
# CONIMICUT POINT REACH

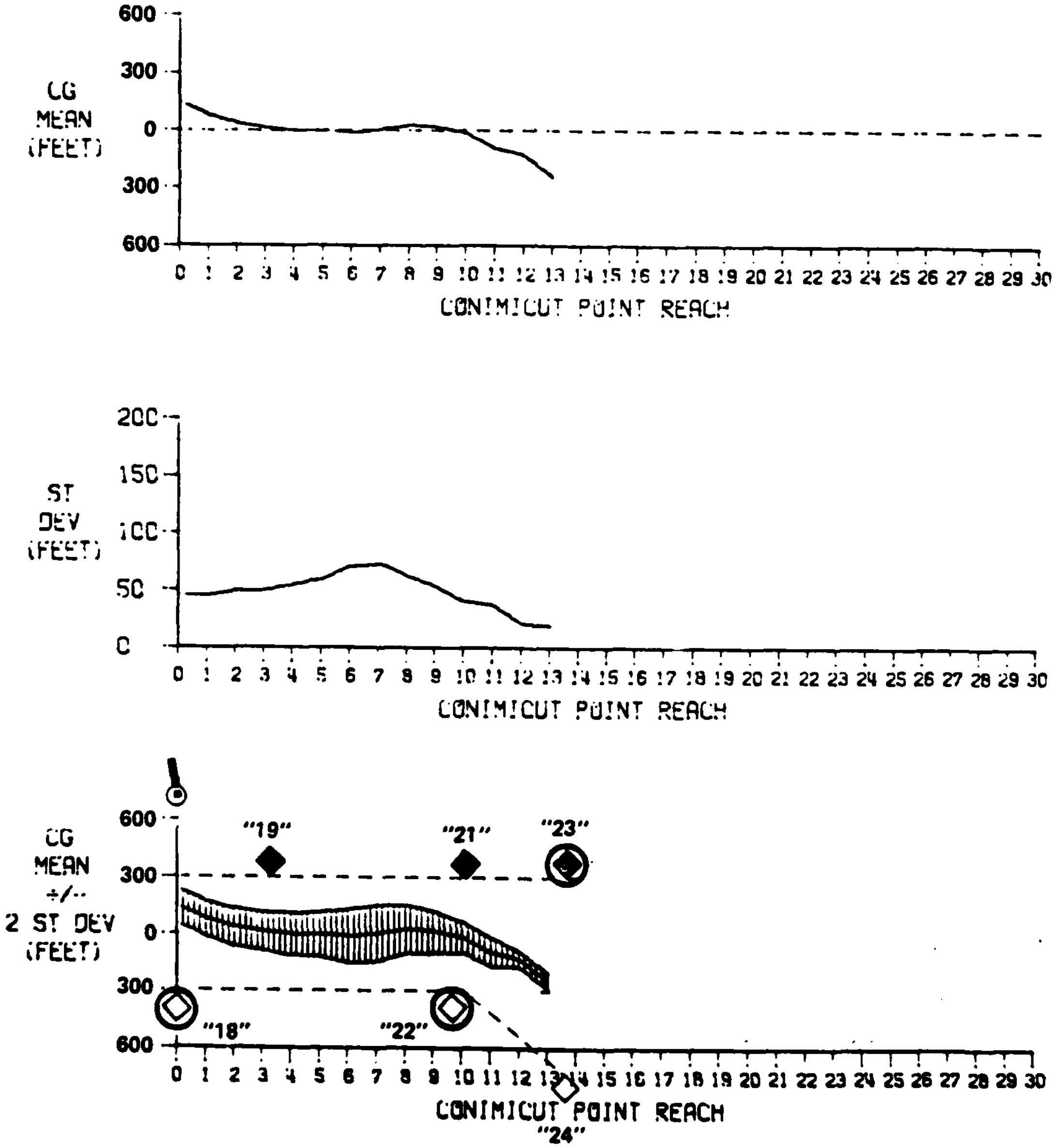


Figure E-11

## AT–SEA
## NIGHTTIME TRANSIT
## NARRAGANSETT BAY
## CONIMICUT POINT REACH

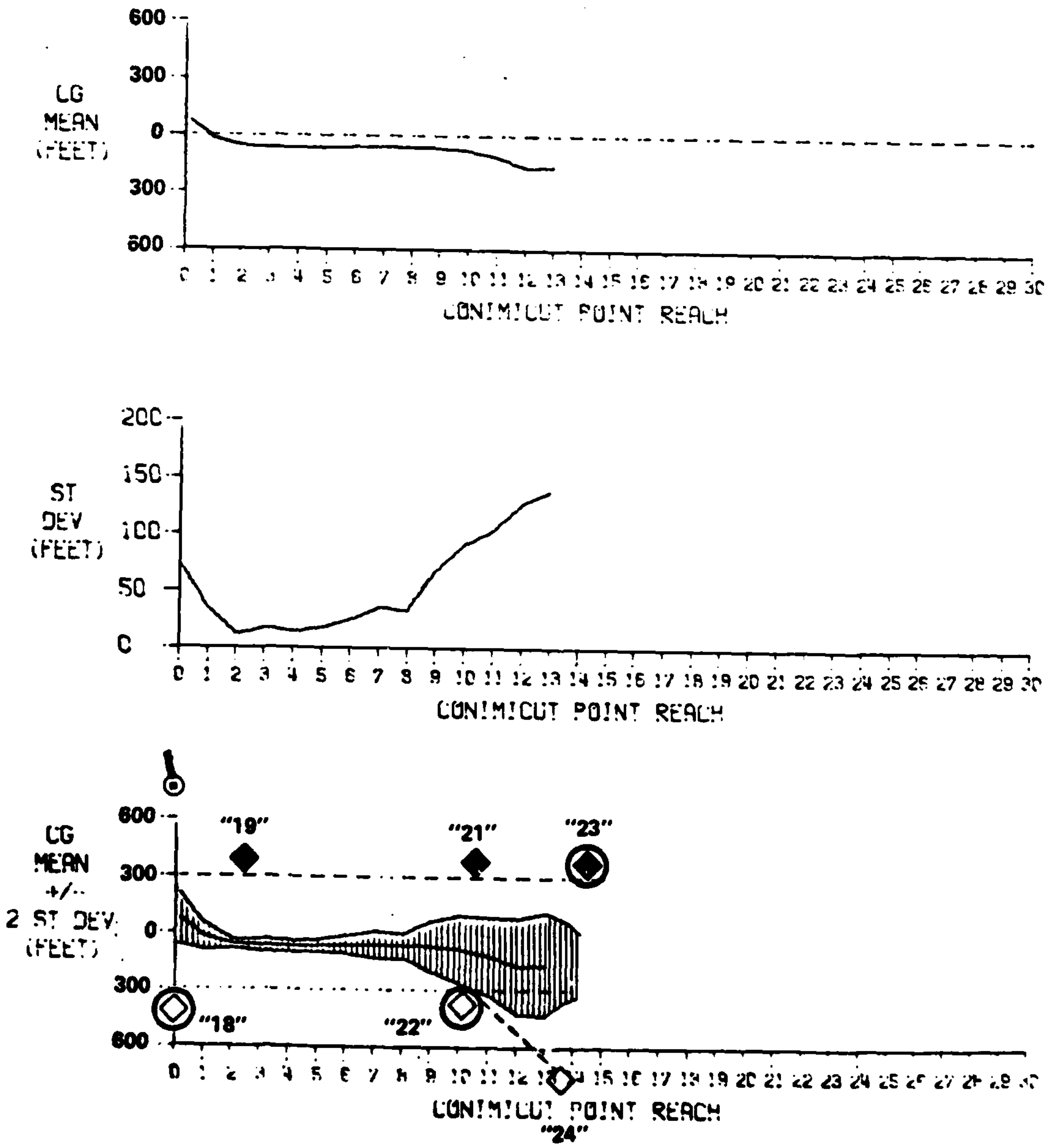


Figure E-12

# AT–SEA
# DAYTIME TRANSIT
# NARRAGANSETT BAY
# BULLOCK POINT REACH

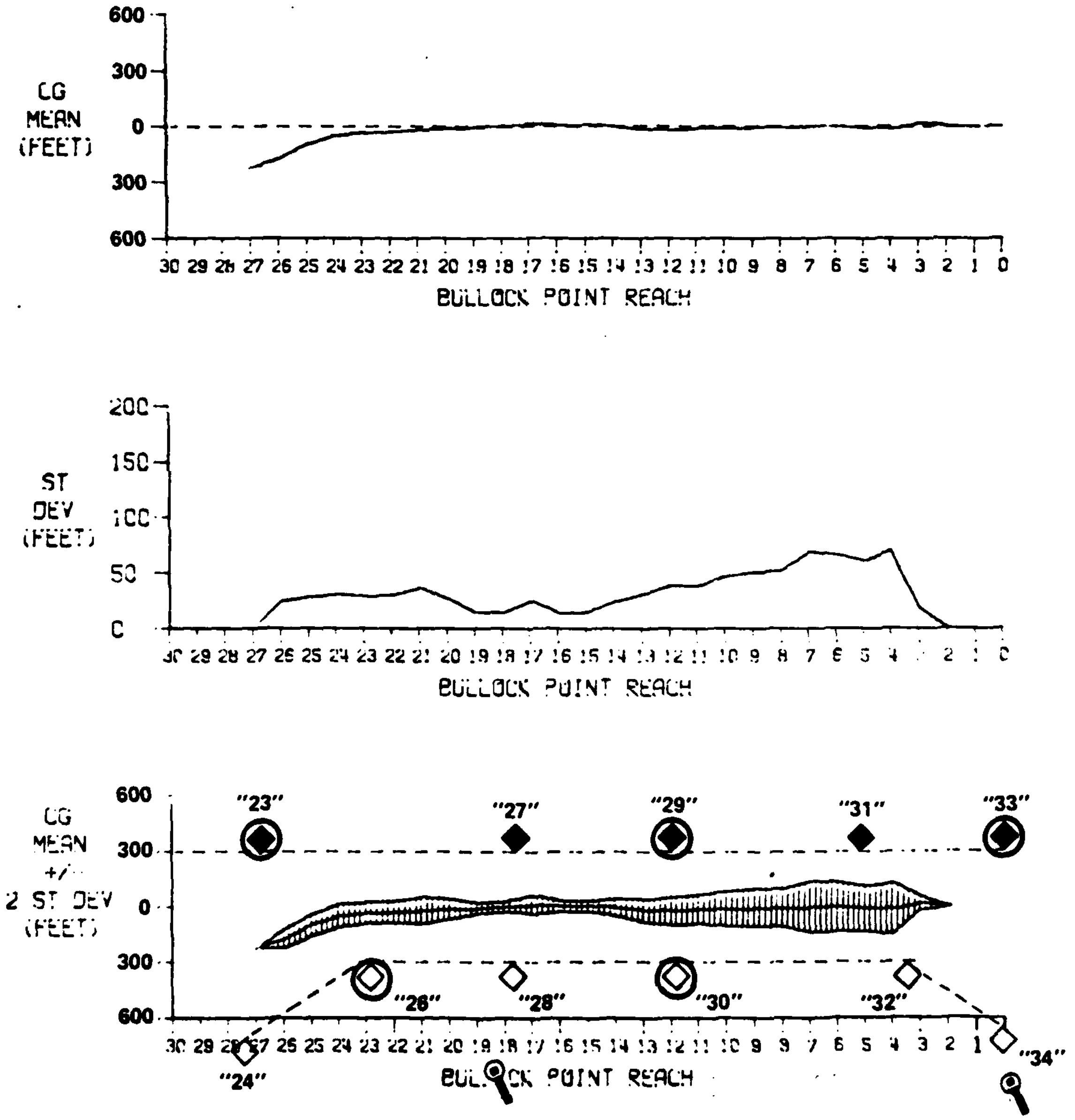


Figure E-13

# AT-SEA
# NIGHTTIME TRANSIT
# NARRAGANSETT BAY
# BULLOCK POINT REACH

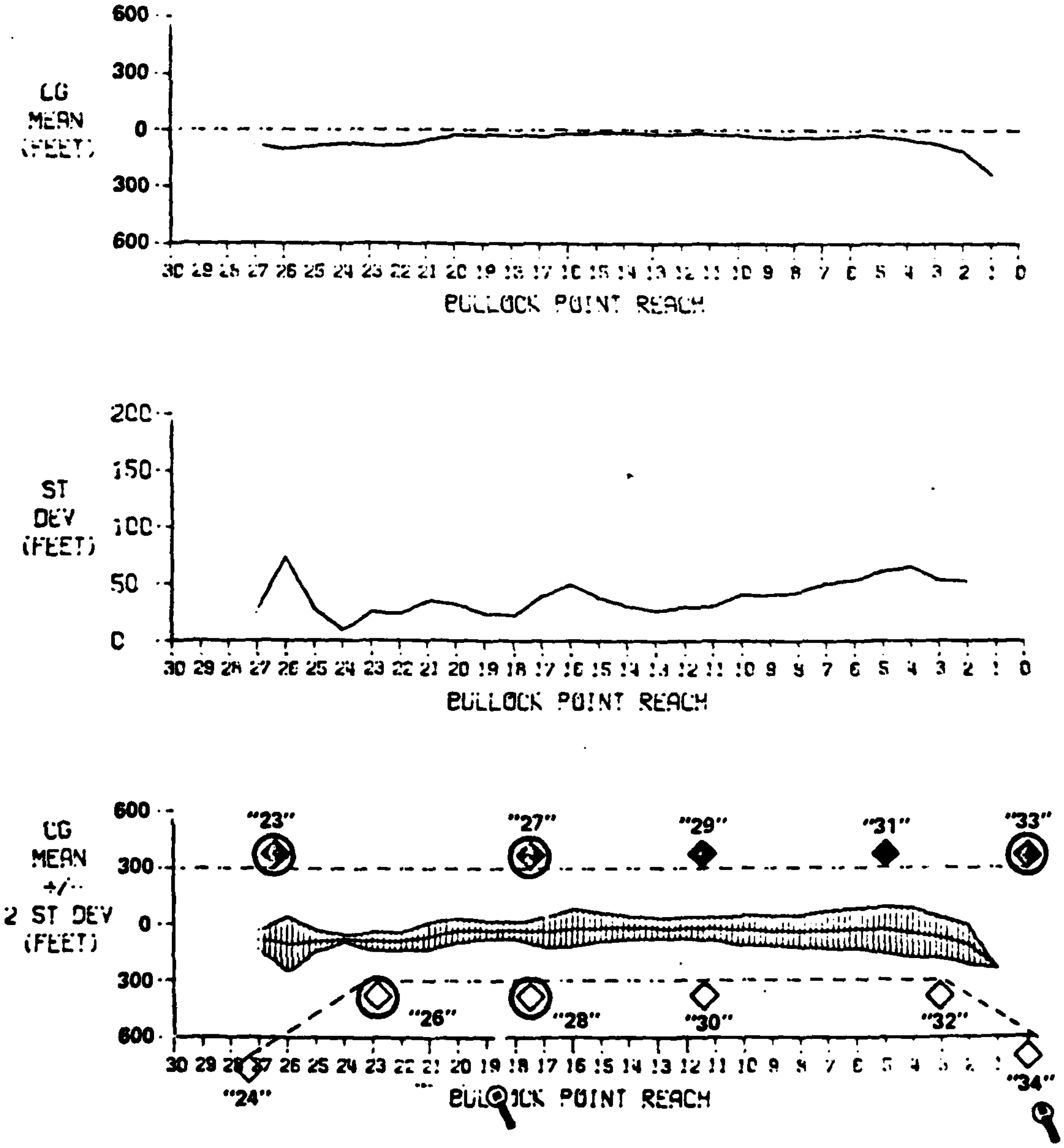


Figure E-14

AT-SEA
NARRAGANSETT BAY
DAY/NIGHT COMPARISON

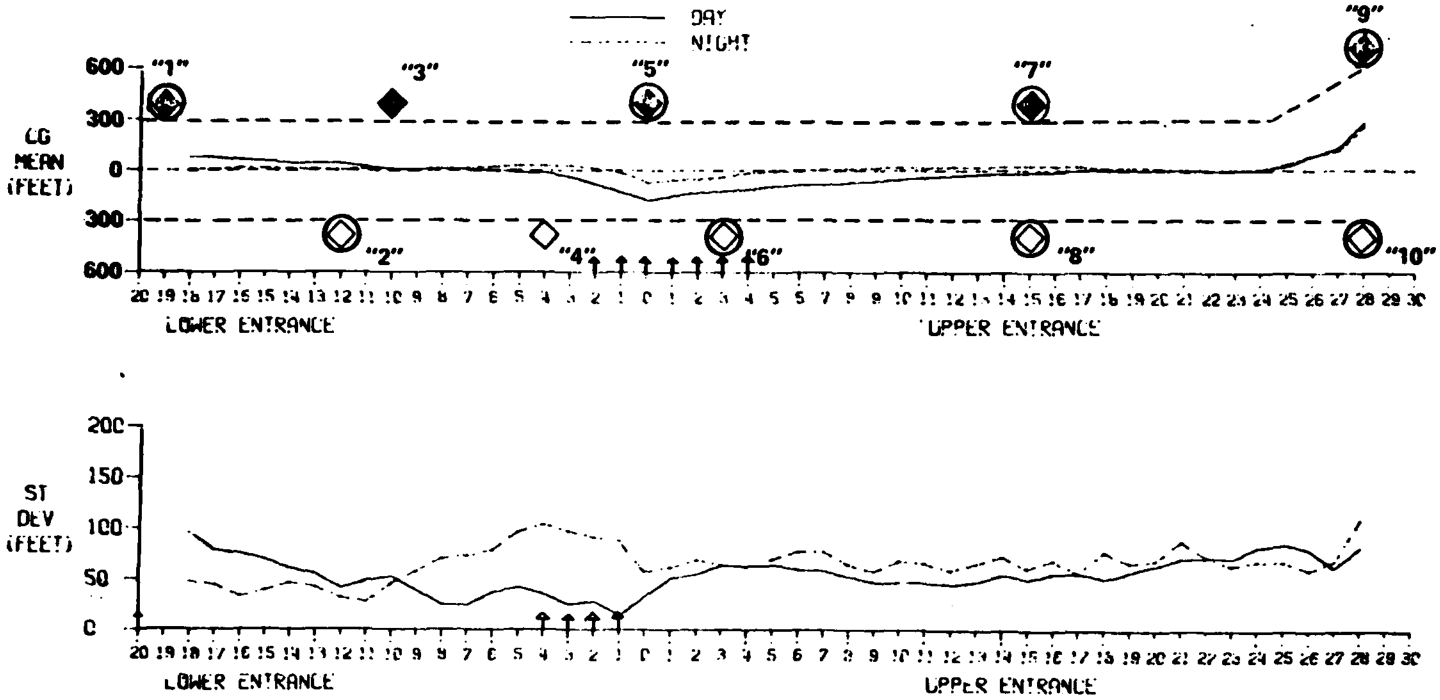


Figure E-15

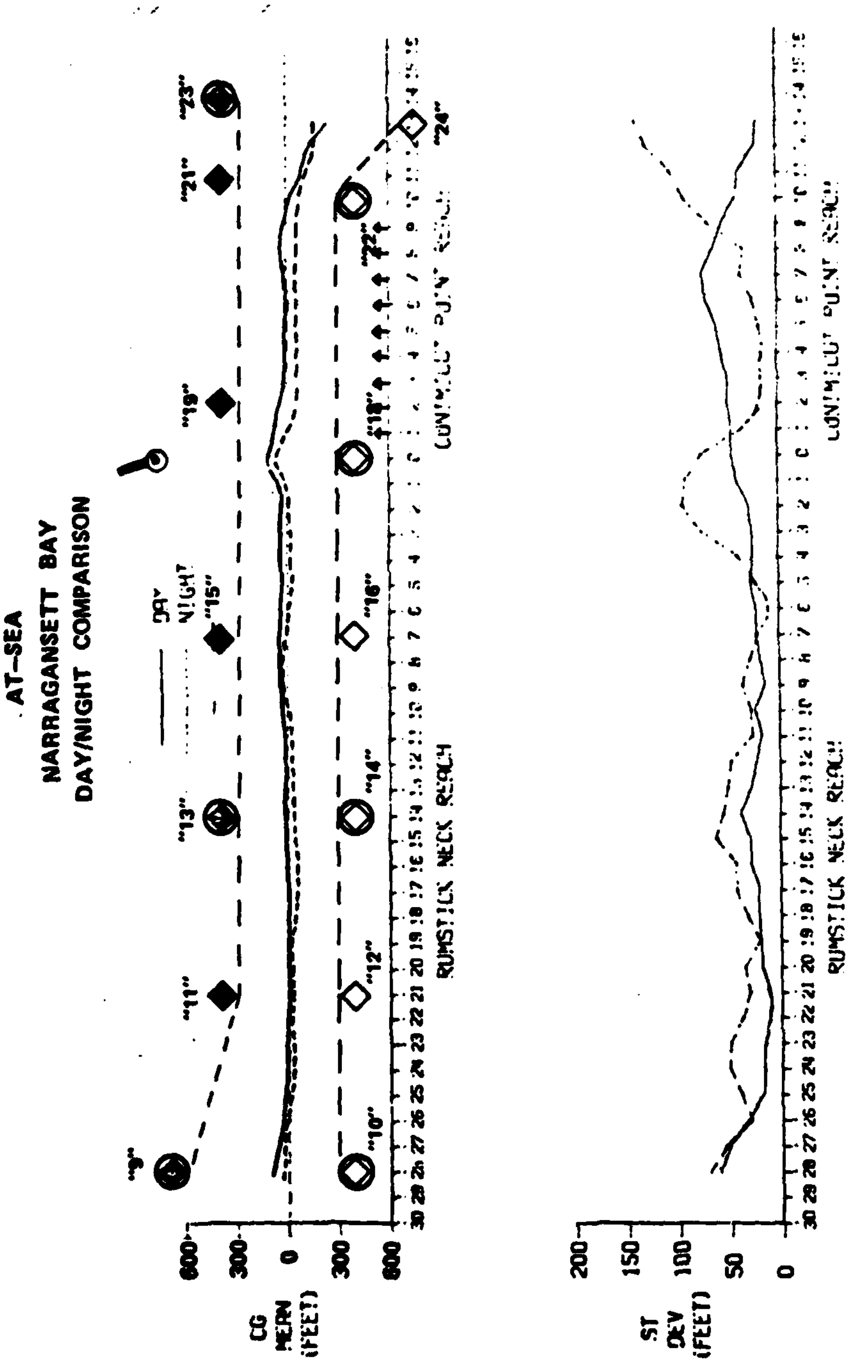

AT-SEA
NARRAGANSETT BAY
DAY/NIGHT COMPARISON
DAY
NIGHT
CG MEAN (FEET)
600
300
0
300
600
"9"
"11"
"13"
"15"
"19"
"21"
"23"
"10"
"12"
"14"
"16"
"18"
"22"
"24"
RUMSTICK NECK REACH
CONIMICUT POINT REACH
ST DEV (FEET)
200
150
100
50
0
30 29 28 27 26 25 24 23 22 21 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Figure E-16

## AT–SEA
## NARRAGANSETT BAY
## DAY/NIGHT COMPARISON

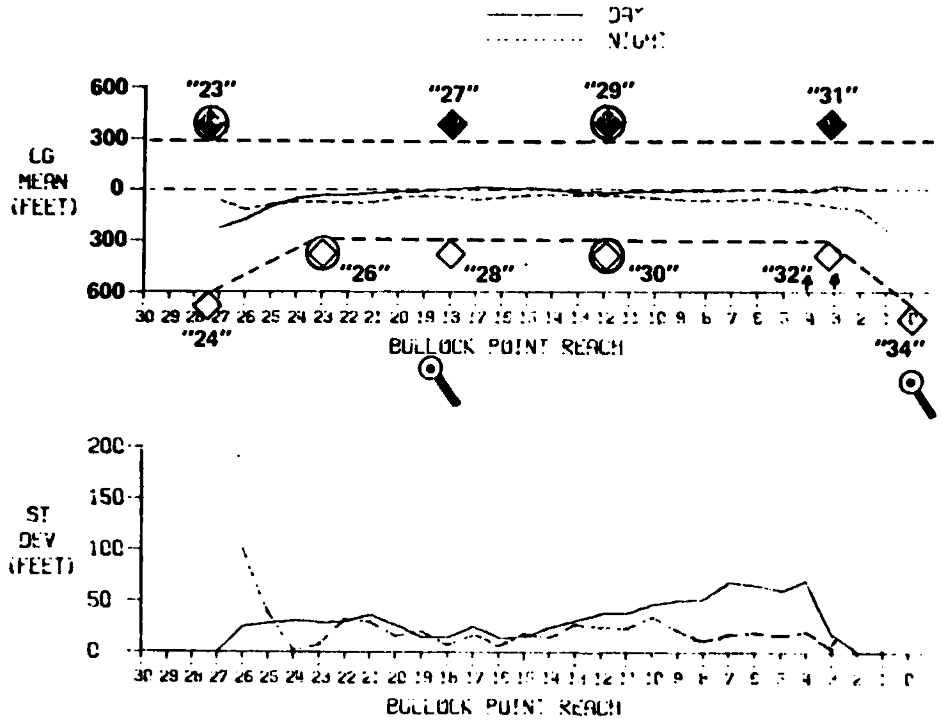


Figure E-17

THIS PAGE INTENTIONALLY LEFT BLANK

# SIMULATOR
# DAYTIME TRANSIT
## NARRAGANSETT BAY
## LOWER AND UPPER ENTRANCE CHANNELS

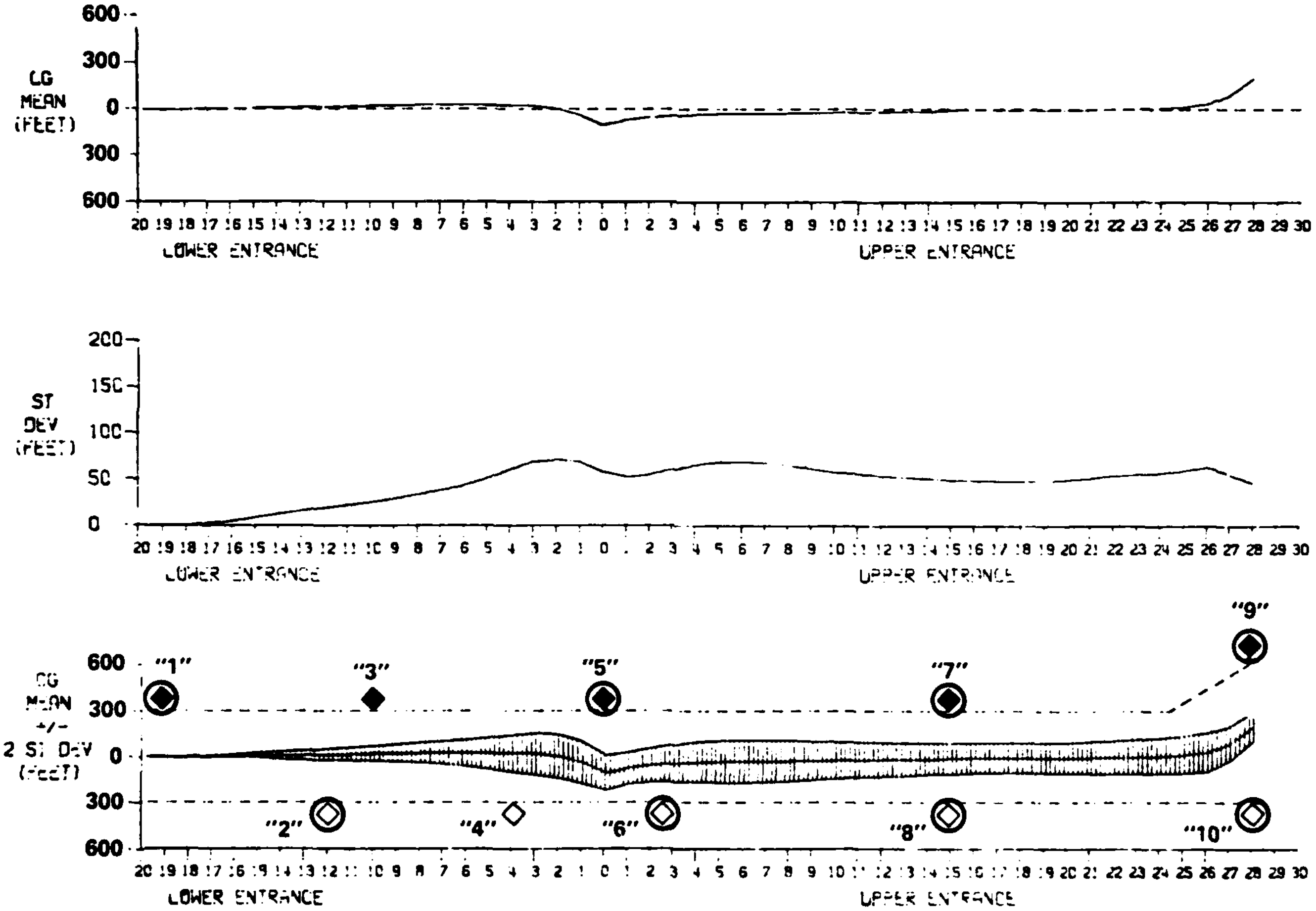


Figure E-18

# SIMULATOR
# NIGHTTIME TRANSIT
## NARRAGANSETT BAY
## LOWER AND UPPER ENTRANCE CHANNELS

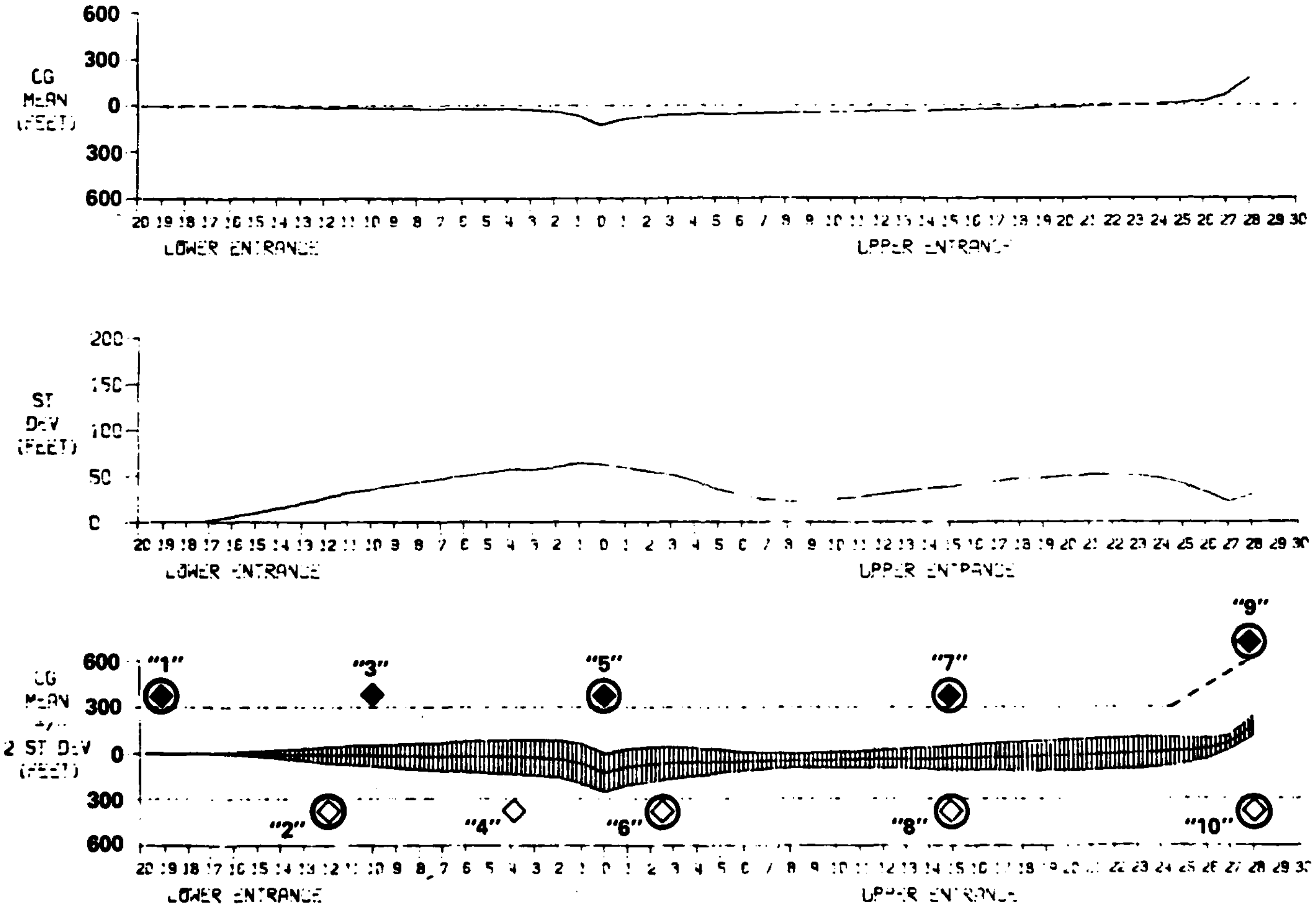


Figure E-13

## SIMULATOR
## DAYTIME TRANSIT
### NARRAGANSETT BAY
### RUMSTICK NECK AND CONIMICUT POINT REACHES

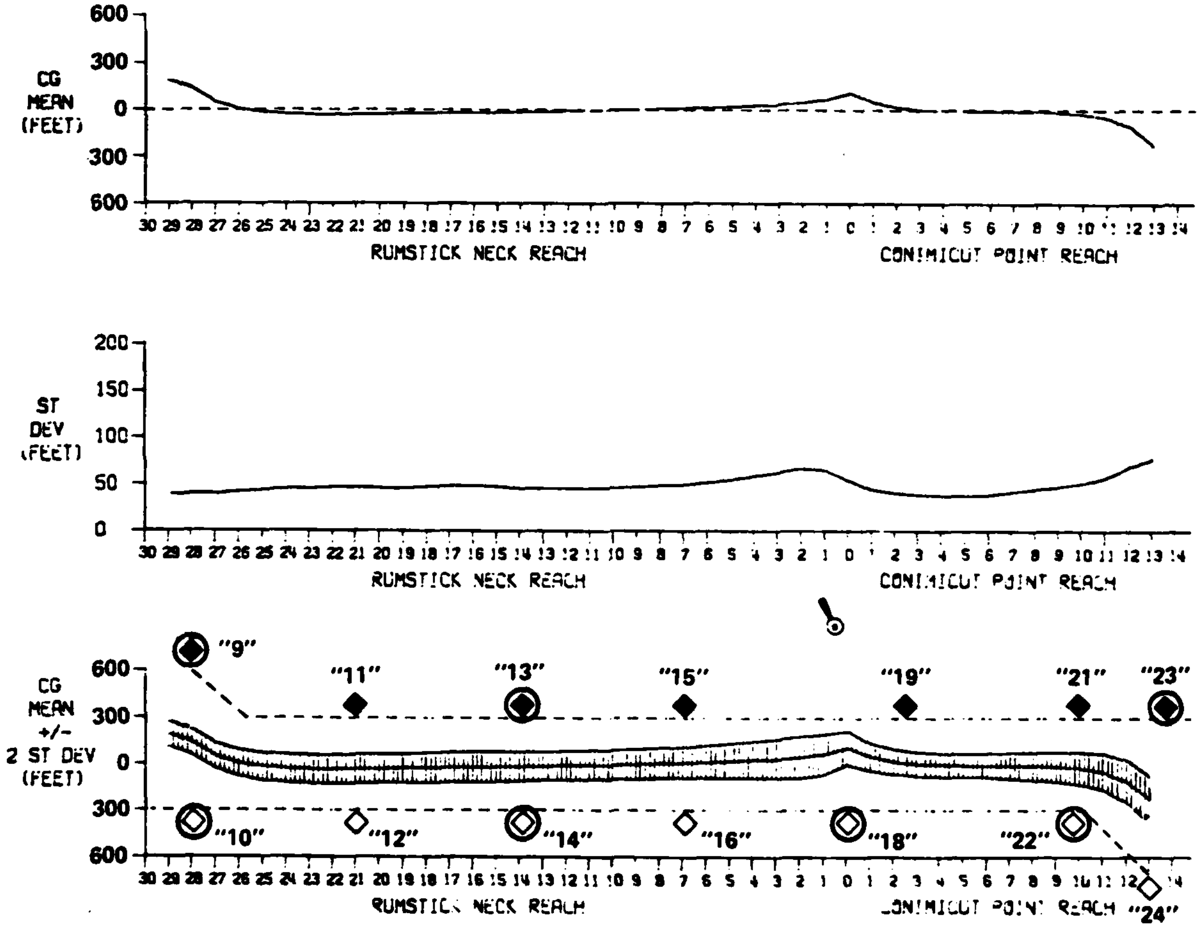


Figure E-20

# SIMULATOR
# NIGHTTIME TRANSIT
## NARRAGANSETT BAY
## RUMSTICK NECK AND CONIMICUT POINT REACHES

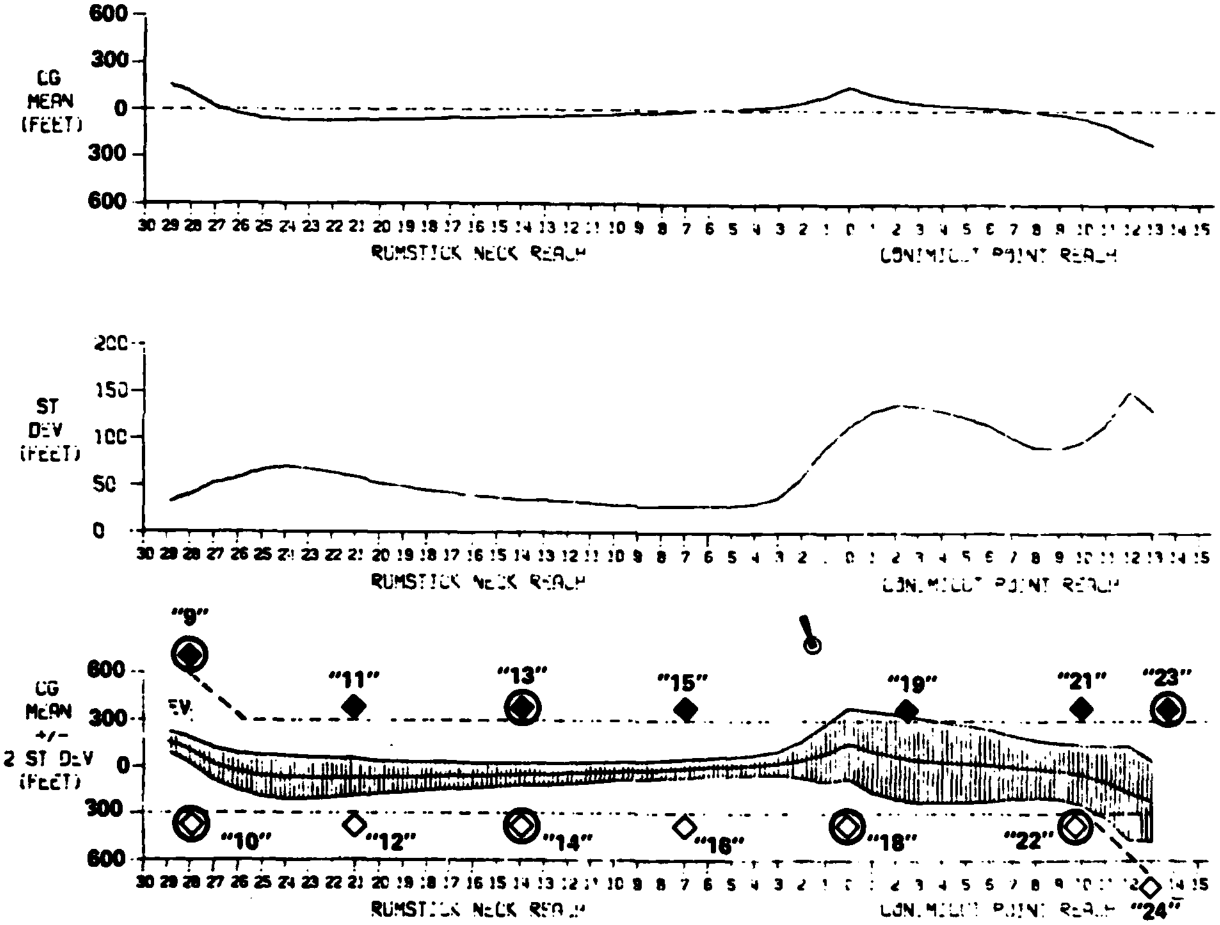


Figure E-21

# SIMULATOR
# NIGHTTIME TRANSIT
## NARRAGANSETT BAY
## BULLOCK POINT AND SABIN POINT REACHES

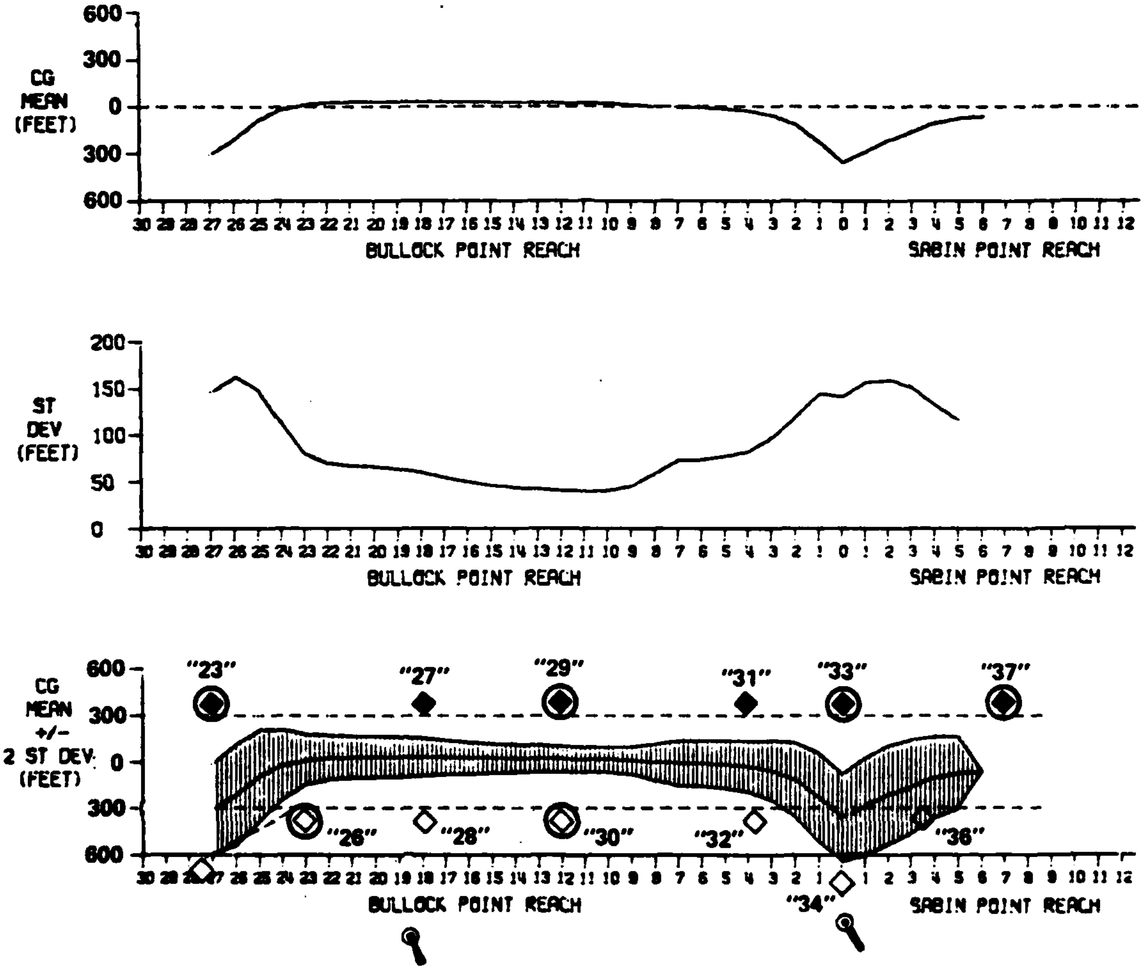


Figure E-22

# SIMULATOR
# DAYTIME TRANSIT
## NARRAGANSETT BAY
## BULLOCK POINT AND SABIN POINT REACHES

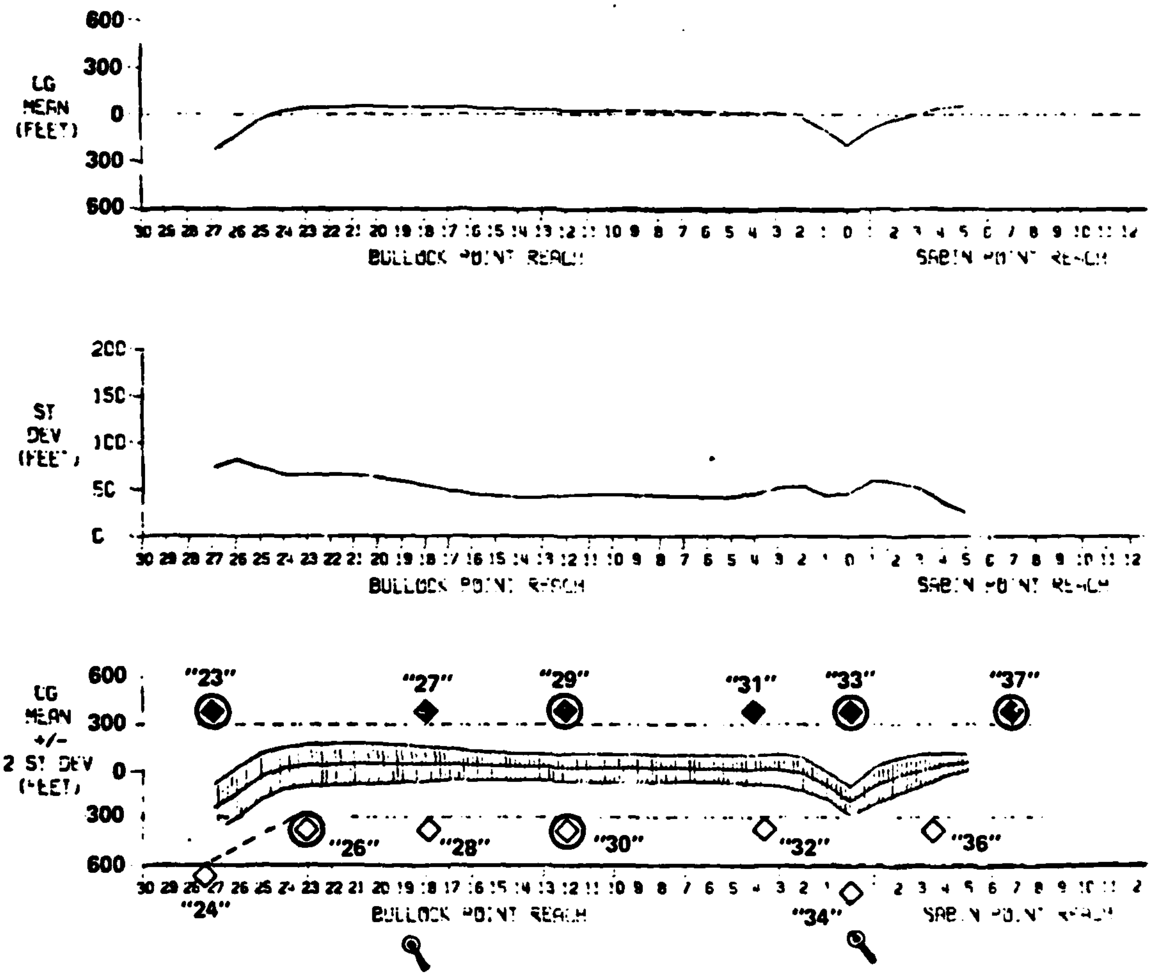


Figure E-23

# SIMULATOR

## NARRAGANSETT BAY

## DAY/NIGHT COMPARISON

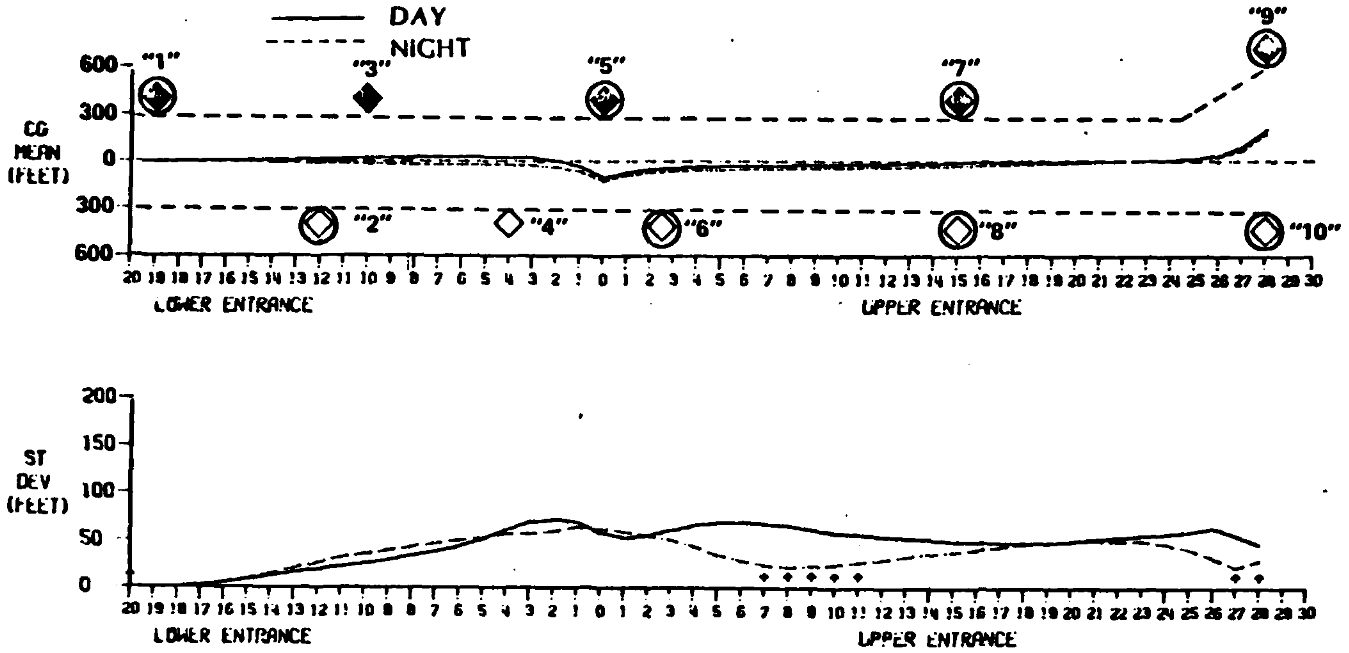


Figure E-24

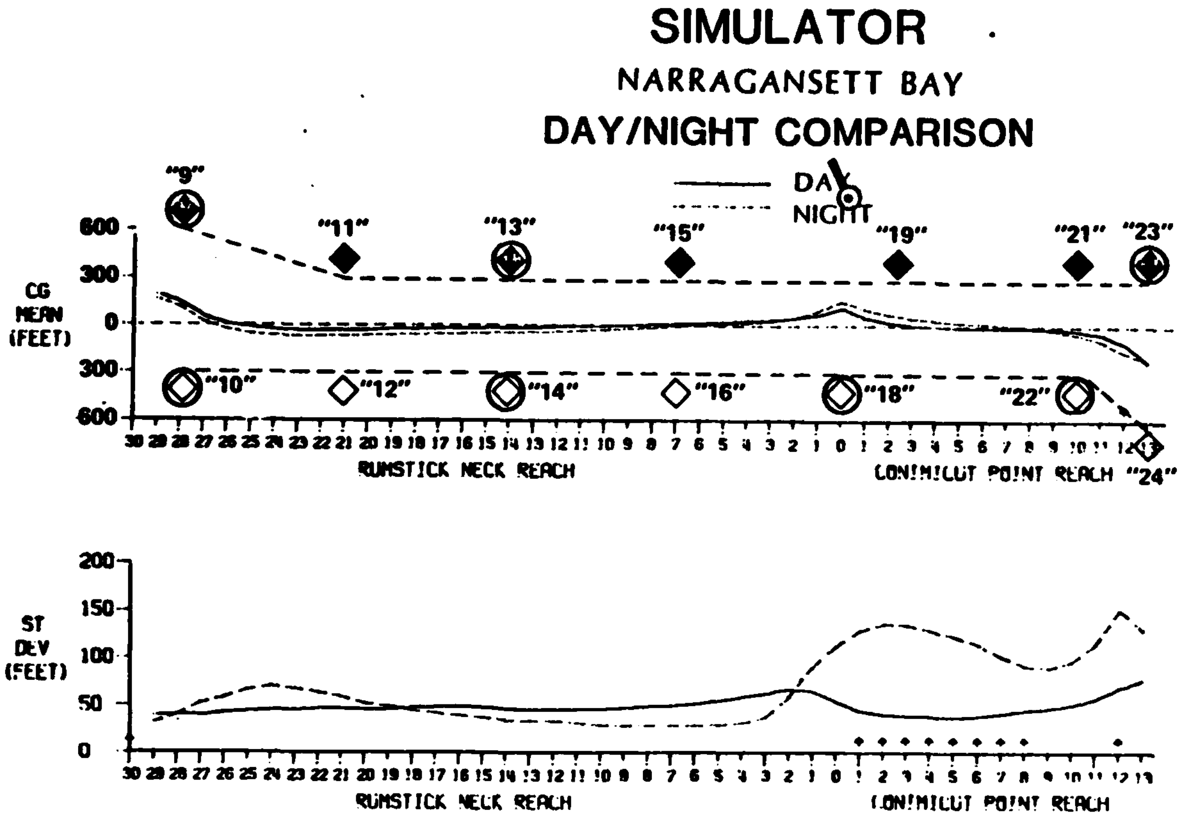

SIMULATOR
NARRAGANSETT BAY
DAY/NIGHT COMPARISON
DAY
NIGHT
CG MEAN (FEET)
600
300
0
300
600
"9"
"11"
"13"
"15"
"19"
"21"
"23"
"10"
"12"
"14"
"16"
"18"
"22"
"24"
30 29 28 27 26 25 24 23 22 21 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 10 11 12 13
RUMSTICK NECK REACH
CONIMICUT POINT REACH
ST DEV (FEET)
200
150
100
50
0
RUMSTICK NECK REACH
CONIMICUT POINT REACH

Figure E-25

# SIMULATOR

## NARRAGANSETT BAY

## DAY/NIGHT COMPARISON

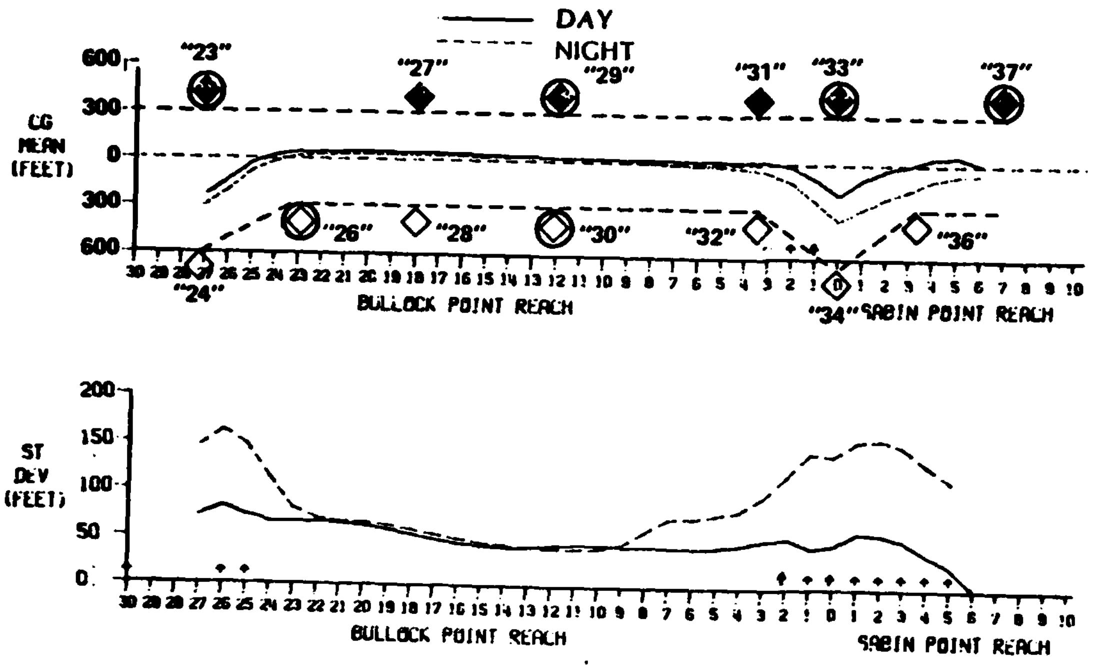


Figure E-26

# APPENDIX F

## HYPOTHESIS TESTING FOR VALIDATION

The experiments completed prior to the Validation experiments were done to determine the differences in performance that existed under a variety of conditions. Performance was compared between different levels of the independent variable or variables in question. To assess these differences statistically, a hypothesis testing procedure is used in which two mutually exclusive hypotheses are evaluated, the null hypothesis ($H_0$) and alternative hypothesis ($H_1$). The null hypothesis states that there are no differences between treatment conditions. The alternative hypothesis states that differences between treatment conditions do exist. Statistical tests are used to evaluate which of these hypotheses is true. In the case of the Aids to Navigation project the t statistic and F statistic were used to establish whether the observed differences between means and between standard deviations, respectively, were due to random fluctuations in performance or to the different experimental conditions being manipulated.

The appropriate use of these statistics depends in part upon the hypothesis testing rules used to determine whether two conditions are the same or different. Figure F-1 show two statistic distributions and the regions corresponding to the probability of making a Type I, or $\alpha$ (alpha), error and Type II or $\beta$ (beta) error. The alpha error represents the probability that the two conditions represented by a test statistic, will be considered different when in fact they are not different. The experimenter must decide at what level to risk making an alpha error. Thus, when the calculated statistic is the tabled value at $\alpha$, $\alpha$ is also the probability that the null hypothesis will be rejected when it is true. However, it is not the case that if the calculated statistic is less than the tabled value at $\alpha$, there is a $1 - \alpha$ probability the two means represent the same population. This is due to the beta error which represents the probability of not rejecting the null hypothesis when it is false, or of rejecting the alternative hypothesis that there is a difference when that hypothesis is true. For the AN project, statistical tests have generally used an $\alpha = 0.10$ as a significance level with that probability distributed 0.05 in each tail. For such a case, it is not true that a lack of significance means a 0.90 probability that the means are the same. The beta error has not been considered.

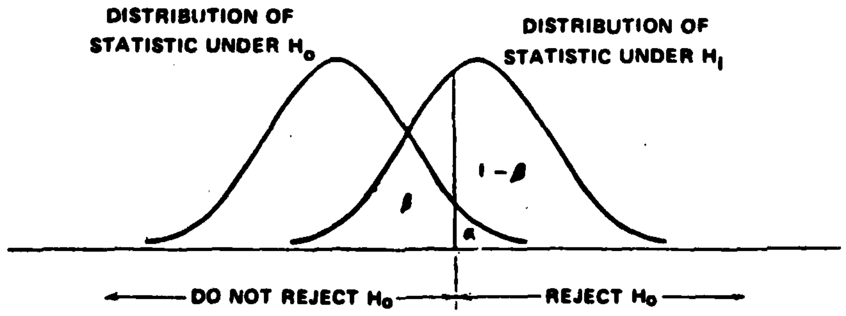


Figure F-1. Two Statistic Distributions and the Regions Corresponding to the Probability of Making a Type I Error ($\alpha$) and Type II Error ($\beta$)

The purpose of the Validation experiments was different from the previous experiments conducted for this project. The goal of this experiment was to evaluate the degree of similarity between conditions found at sea and equivalent conditions tested on the simulator. In this situation, where one is interested in measuring the degree of similarity between two conditions, the same hypothesis testing procedure used for the other experiment is not appropriate. The procedures are designed to measure difference between two conditions and are not sufficiently conservative in their probability estimate of sameness. There is, however, a procedure that is more conservative, to evaluate the degree of sameness between the means. No such procedure exists for an evaluation of the standard deviations.

To find the degree to which two means represent the same population, the experimenter must take into account beta error as well as alpha error. The beta analysis was performed to calculate the probability of making a beta error. Subtracting the probability of making both a alpha error and beta error from the total probability distribution for the null hypothesis ($H_0$) provides a more conservative estimate of the probability of not rejecting the null hypothesis when it is true $[1 - (\alpha + \beta)]$ than subtracting only the probability of making an alpha error from the total probability distribution for the null hypothesis.

The following is a simple example meant to illustrate the procedure used in estimating the beta error. Table F-1 shows sample means and standard deviations for two conditions and their respective t statistic and F statistic. The lack of statistical significance between the means suggests that the null hypothesis is not false. The beta error can be calculated only in those situations where the means are not statistically different.

TABLE F-1. SAMPLE VALUES FOR USE IN CALCULATION OF BETA ERROR

| | Mean (feet) | Standard Deviation (feet) | Test Statistic and Degrees of Freedom | Significant at $\alpha = 0.10$ |
|---|---|---|---|---|
| Condition A | 30R | 25 | t(14) = 1.08 | No |
| Condition B | 60R | 30 | F(7,7) = 1.44 | No |

The following formula was used to calculate $\phi^2$ (phi$^2$):

$$\Phi^2 = \frac{n\Sigma(\mu_1 - \mu_2)^2/a}{\sigma_e^2}$$

Where:

$\Phi^2$ = the ratio of the treatment mean square variance divided by the population error variance

$n$ = the number of subjects in each condition

$\mu_1 - \mu_2$ = the difference between the means

$a$ = the number of treatment groups

$\sigma_e^2$ = the population error variance

For this sample problem:

$$n_1 = n_2 = 8 \qquad a = 2 \qquad \sigma_e^2 = 762.5$$

Using the formula given above $\Phi$ (phi) is calculated for the desired difference between the means $[\Sigma(\mu_1 - \mu_2)]$. For a 20-foot difference the following phi would be calculated:

$$\Phi = \sqrt{\frac{8(20)^2/2}{762.5}} = 2.09$$

The next step is to find the power $(1-\beta)$ associated with the calculated phi using statistical tables showing the power function for analysis of variance. The power is found by identifying the calculated phi on the abscissa for the alpha error rate selected and a-1, a(n-1) degrees of freedom, and reading the power off the ordinate. These tables can be found in Myers (1979) and Kirk (1968). In this example the power associated with the calculated phi of 2.09 and 1 and 14 degrees of freedom equals 0.78 at $\alpha$ 0.05. These tables present the probability for one tail. This value is subtracted from 1 (representing the total probability distribution) to obtain the beta error. Here, the beta error is estimated at 0.22 for one tail.

Once the beta error is known it is a straightforward procedure to calculate the probability that the two means reflect the same population. The probability estimate differs from the calculation of beta error in that it is based upon both tails of the distribution. This is consistent with the analysis of the experimental conditions in general, in which the data was interpreted for a two-tailed distribution. The probability value is calculated by multiplying the alpha and beta error by a factor of two to reflect both tails of the distribution, adding the two error estimates together and subtracting this sum from one (representing the total probability distribution) as shown here.

$$\begin{array}{lrr} \text{alpha error} & 0.05 \times 2 = & 0.10 \\ \text{beta error} & 0.22 \times 2 = & \underline{+0.44} \\ & & 0.54 \end{array}$$

$$p = 1 - 0.54 = 0.46$$

The outcomes of these calculations are shown in Table F-2. Table F-2 shows the beta errors associated with different levels of differences between the means ($\mu_1 - \mu_2$) and the probability (p) that those differences represent the same population (i.e., the null hypothesis is true). For example, for a 20-foot difference between the means the beta error equals 0.22 and there is a 0.46 probability that the two groups represent the same population. For a 30-foot difference between the means the beta error equals 0.18 and the probability is 0.54 that the two means reflect the same population.

TABLE F-2. TYPE II ERROR AND PROBABILITY OF TWO MEANS REPRESENTING THE SAME POPULATION FOR SEVERAL LEVELS OF DIFFERENCES BETWEEN MEANS

| $M_1$-$M_2$* | Beta Error** | P |
|---|---|---|
| 20 | 0.22 | 0.46 |
| 30 | 0.18 | 0.54 |
| 40 | 0.35 | 0.83 |
| 50 | 0.01 | 0.88 |

*Units are in feet
**Beta error calculated for $\alpha = 0.05$, one tail $df_1 = 1$, $df_2 = 14$

The level of the alpha error, in this case $\alpha = 0.10$ for a two tailed distribution, sets an upper limit on the probability value that can be achieved. Assuming there was no beta error, the maximum probability would be .90. Due to the procedure by which the beta error values are tabled, the minimum beta error will be 0.02 for a two-tailed distribution. For these reasons, the maximum probability that two means will be accepted as reflecting the same population is 0.88.

While calculation of the beta error provides a better estimate of the degree to which two groups represent the same population than the t statistic, there is an important limitation that must be kept in mind. Their calculation uses the crosstrack standard deviation to estimate population error. For any given difference between the means ($\mu_1 - \mu_2$) an increased population error has the effect of lowering the probability that two means reflect the same population. However, a large population error does not necessarily indicate that the two conditions are not the same. A large standard deviation in both conditions may be just as indicative of similar performance as two small standard deviations. The standard deviation provides meaningful information about the similarity between conditions just as the crosstrack mean does. The formula for beta error does not take this meaningful information into account. Instead this measure is treated as random error.

This fact has two implications for the interpretation of the data. First, in the turn regions where the standard deviation tends to be lower than in the trackkeeping region the degree of similarity between the simulator and at-sea conditions will be underestimated. Second, in the trackkeeping regions where the standard deviation tends to be higher compared to the turn, the degree of similarity between simulator and at-sea conditions will be overestimated. A more thorough account of the reasoning behind the calculation of beta is in texts by Myers, 1979, Cohen, 1977, Kirk, 1968 or Pfaffenberger and Patterson, 1977.

# BIBLIOGRAPHY

Bertsche, W.R., D.A. Atkins and M.W. Smith. "Aids to Navigation Principal Findings Report on the Ship Variables Experiment: The Effect of Ship Characteristics and Related Variables on Piloting Performance." CG-D-55-81, U.S. Coast Guard, Washington, D.C., October 1981. NTIS AD-A108771.

Bertsche, W.R., M.W. Smith, K.L. Marino and R.B. Cooper. "Draft SRA/RA Systems Manual for Restricted Waterways." CG-D-77-81, U.S. Coast Guard, Washington, D.C., February 1982. NTIS AD-A113236.

Cohen, Jacob. Statistical Power Analysis for the Behavior Sciences. New York: Academic Press, 1977.

Cook, R.C. "At-Sea Data Collection Program for the Validation of Piloting Simulation and Evaluation of Radar Aided Piloting Techniques." U.S. Coast Guard, Washington, D.C., March 1981.

Cooper, R.B., R.C. Cook and K.L. Marino. "At-Sea Data Collection for the Validation of Piloting Simulation." CG-D-60-81. U.S. Coast Guard, Washington, D.C., December 1981. NTIS AD-A111978.

Eclectech Associates. "Aids to Navigation Preliminary Data Analysis for the At-sea Data Collection Program." Technical Memorandum. U.S. Coast Guard, Washington, D.C., August 1981.

Grant, G.E., J.D. Moynehan and M.W. Smith. "Aids to Navigation Validation Draft Principal Findings Report: Simulation I: Chesapeake Bay." U.S. Coast Guard, Washington, D.C., February 1983.

Gynther, J.W. and R.B. Cooper. "At-Sea Data Collection Plan for Prototype Implementation of Aid System Design Guidelines." U.S. Coast Guard, Washington, D.C., April 1982.

Kirk, Roger E. Experimental Design: Procedures for the Behavioral Sciences. Belmont, California: Brooks/Cole Publishing Co., 1968.

Marino, K.L. M.W. Smith and J.D. Moynehan. "Aids to Navigation SRA Supplemental Experiment Principal Findings: Performance of Short-Range Aids Under Varied Shiphandling Conditions." CG-D-03-84, U.S. Coast Guard, Washington, D.C., July 1984.

Moynehan, J.D., M.W. Smith and J.W. Gynther. "Aids to Navigation Presimulation Report: Aids in the Meeting Traffic Situation." U.S. Coast Guard, Washington, D.C. February 1984.

Moynehan, J.D. "Aids to Navigation Implementation At-Sea Preimplementation Draft Principal Findings." U.S. Coast Guard, Washington, D.C., September 1983.

Moynehan, J.D. "Aids to Navigation Implementation At-Sea Preimplementation Draft Principal Findings." U.S. Coast Guard, September 1983.

Multer, J. and M.W. Smith. "Aids to Navigation Radar I Principal Findings: Performance in Limited Visibility of Short-Range Aids with Passive Reflectors." CG-D-79-83, U.S. Coast Guard, Washington, D.C., December 1983.

Multer, J. and M.W. Smith. "Aids to Navigation Turn Lights Principal Findings: Effect of Turn Lighting Characteristics, Buoy Arrangements, and Ship Size on Nighttime Piloting." CG-D-49-82, U.S. Coast Guard, Washington, D.C., February 1983. NTIS AD-A-126080.

Myers, Jerome, L. Fundamentals of Experimental Design. Boston: Allyn and Bacon, Inc., 1979.

Pfaffenberger, Roger C. and James H. Patterson. Statistical Methods for Business and Economics. Homewood, Illinois: Richard D. Irwin, Inc., 1977.

Smith, M.W. and W.R. Bertsche. "Aids to Navigation Principal Findings on the CAORF Experiments: The Performance of Visual Aids to Navigation as Evaluated by Simulation." CG-D-51-81, U.S. Coast Guard, Washington, D.C., February 1981. NTIS AD-A107045.

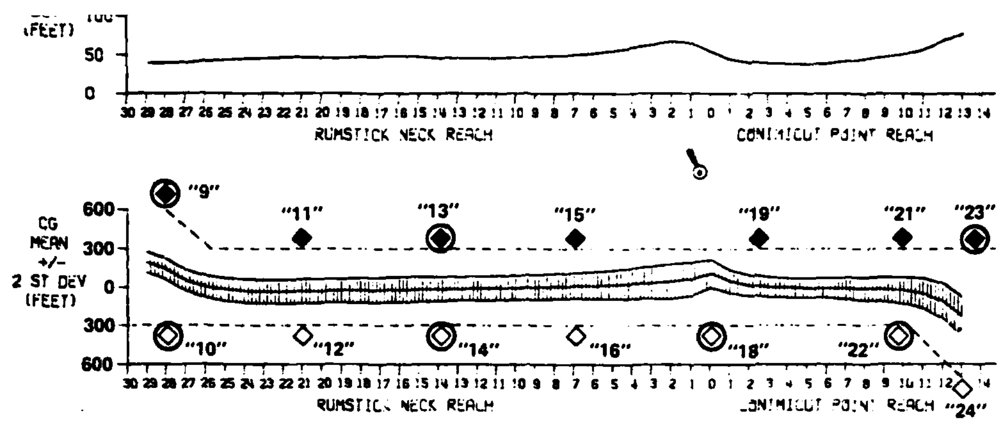


Figure E-20

END

DATE

FILMED