The High-Resolution Industry-Based Trawl Survey: Methodology and Data Processing Procedures
The High-Resolution Industry-Based Trawl Survey: Methodology and Data Processing Procedures.

by

Rodney A. Rountree, Ross Kessler, Glenn Strout, David Martins, Darin Jones and Frank Bub

31 October 2004
ABSTRACT

The High-Resolution Trawl Survey project was established as a collaboration between the New Bedford commercial fishing industry and the School for Marine Science and Technology (SMAST) of the University of Massachusetts Dartmouth. One of the major incentives for the project was to recognize the desire of the commercial fishing community to participate more directly in the management of the resource. The primary focus of the project has been to create an industry-based program for the collection of trawl catch and environmental data suitable for use by researchers and fisheries managers. The SMAST High-Resolution Industry-Based Trawl Survey was a highly successful demonstration of a program to train fishermen to collect environmental and biological data. The data generated by this project is expected to provide significant information to regional managers by characterizing both the Georges Bank fishery trawling operation and its catch. The data collection goal of the project was to record catch, vessel, and environmental data for individual tows made during normal fishing operations by commercial trawlers operating on Georges Bank. The work has been conducted in collaboration with 20 vessels from the New Bedford trawl fleet, including vessels that fish primarily on Georges Bank. Trawl data were collected during two periods: November 2000 - October 2001, and August 2002-July 2003. In order to provide researchers with guidance on the strengths and weaknesses of data generated by this project, this report provides a detailed summary of data collection, keypunching, processing and data quality control and error checking methodologies used in the program. Summaries of the data generated by this project will be reviewed in subsequent data reports to be issued by SMAST. In addition information contained in this report provides significant insight on how to better design future industry-based trawl fishery surveys. The most difficult component of the fishery data to quantify with an industry-based survey is the catch discard data, because the collection of discard data is in direct conflict with the fishermen's need to efficiently process the kept catch, and to quickly return discard species to the sea in order to clear space for subsequent hauls and reduce mortality.
# TABLE OF CONTENTS

ABSTRACT ......................................................................... iii

LIST OF FIGURES ............................................................. vi

LIST OF TABLES .............................................................. xi

INTRODUCTION .................................................................. 1

MATERIALS AND METHODS ........................................ 1
  Fishing Boats .............................................................. 2
  Vessel Inspection and Training ...................................... 3
  Overview of Field Operations ....................................... 4
  Overview of Data Keypunching and Quality Control ........ 6
  Trawl Gear ..................................................................... 8
    Otter trawl components ............................................. 8
    Variations of the otter trawl ....................................... 10
    Gear performance ..................................................... 11
  Data Collection .......................................................... 12
    Data concept ............................................................. 12
    Logsheets ............................................................... 14
  Date Entry and Processing ........................................... 16
    Trawl master data files ............................................. 16
    Vessel log processing ............................................... 17
    Gear log processing .................................................. 18
    Trip log processing .................................................... 19
    Environmental log processing ................................... 19
    Water temperature data logger processing ................. 20
    Haul log processing .................................................. 23
    Length-frequency log processing ............................... 27
  Overview of Working Data Sets .................................... 28

CONCLUDING REMARKS AND RECOMMENDATIONS .......... 30

ACKNOWLEDGMENTS .................................................. 32

FIGURES ........................................................................ 33

TABLES .......................................................................... 60

APPENDICES .................................................................. 172
  Appendix A. Initial Sign-up and Contracting .................... 173
Appendix B. Original Data Processing Procedures. ................. 182
Appendix C. Summary of Data Processing Procedures. ............ 188
Appendix D. Keypunch Instructions to Trade Quote, Inc. ........ 201
Appendix E. Samples of All Haul, Environmental, and Length Frequency Log Variations Used in the Trawl Survey. ................ 206
Appendix F. Major Haul Log Changes ................................. 227
Appendix G. Look up Tables for Coded Fields Found in the Master Data Set Files. .................................................. 234
Appendix H. Look up Tables for Coded Fields Found in the Working Data Set Files. .............................................. 247
Appendix I. SAS Programs Used to Create the Trawl Survey Working Data Sets ...................................................... 264
LIST OF FIGURES

Figure 1. Flow chart illustrating the steps taken to train fishermen to conduct fisheries sampling during their normal fishing operations. .......... 34

Figure 2. Schematic diagram of an otter trawl illustrating the component parts, dimensions and swept area. H=trawl height, n=trawl width, which is also the width of the swept area, d=door-to-door distance, which is also the maximum width of the potential herding area. ......................... 35

Figure 3. Schematic illustration of the effect of trawl height (h) on the bottom swept and herding areas, and on fish catchability for different types of fish: a) cod, b) monkfish located inside of the swept area, c) haddock, d) monkfish located outside of the swept area. ....................... 36

Figure 4. Conceptual data structure of information gathered on the trawl survey. See text for explanation. ................................. 37

Figure 5. Actualized structure of data collected from field logsheets. Differences between the logsheet structure and the conceptual structure illustrated in Fig. 3 arise from logistical restraints and the need for field sheets that can be used efficiently by the fishermen. ......................... 38

Figure 6. Illustration of the vessel logsheet used to collect data on fishing vessel characteristics during the SMAST Trawl Survey. Note that it is almost an exact copy of the NMFS Observer Program Vessel and Trip Information Log. ......................................................... 39

Figure 7. Illustration of the Gear log used in the SMAST Trawl Survey. It is a modification of the Gear Characteristics Log used in the NMFS Observer Program. ............ 40

Figure 8. Reproduction of the SMAST Trawl Survey trip log. It is derived from the trip components of the NMFS Observer Program Vessel and Trip Information Log and the Vessel and Trip Information - Six Month Questionnaire. ......................................................... 41

Figure 9. Illustration of the environmental log used during the early stages of the SMAST Trawl Survey. It was soon replaced by a more efficient version. ......................................................... 42

Figure 10. Environmental log sheet used during most of the trips in the SMAST Trawl Survey. ......................................................... 43

Figure 11. Original haul log used during the early trips for the SMAST Trawl Survey. It is derived from the NMFS Observer Program Haul Log. ............ 44

Figure 12. Second generation of the SMAST Trawl Survey haul log. ............ 45

Figure 13. Third version of the SMAST Trawl Survey haul log. ............ 46

Figure 14. Fourth version of the SMAST Trawl Survey haul log. ............ 47

Figure 15. Fifth version of the SMAST Trawl Survey haul log. ............ 48

Figure 16. Sixth version of the SMAST Trawl Survey haul log. ............ 49

Figure 17. Seventh version of the SMAST trawl Survey haul log. It was used in year three (beginning August 2003). ......................... 50

Figure 18. Initial species length frequency log used to record fish lengths in the
Figure 19. Second generation length frequency log used in the SMAST trawl survey. .................................................. 52

Figure 20. Third generation length frequency log used in the SMAST Trawl Survey. ................................................... 53

Figure 21. Relationship between the master data sets created from the field data logsheets with the final working data sets used for analysis. Steps needed to create the working data sets from the master data sets are described in the text. ........................................... 54

Figure 22. The Tidbit temperature data logger and optic base station made by Onset Computer Corporation (Pocasset, MA). ........................................... 55

Figure 23. Examples of Tidbit water temperature data recorded on two different trips (left and right panels). Each panel depicts what is seen in the window while executing the matlab program Tidbitselect.m for each trip. In the first example, one Tidbit (B) was deployed on the gear being used during the haul, while the other Tidbit (A) was attached to another net stored on the deck. In the second example, Tidbits D and E were both deployed on the gear, one on each door. The program allows for up to three Tidbit records to be displayed for a given trip. In these cases, only two Tidbit recorders were deployed on each trip (A and B in the left panel and D and E in the right panel). ............................... 56

Figure 24. Two example Tidbit temperature profiles taken during a trawl tow. Both of the above panels depict typical haul profiles for boats fishing on Georges Bank. Note that the graphic on the left has one turn and the one the right has two. .................................................. 57

Figure 25. Measuring board and length frequency tally sheet used to record fish length frequency data during the SMAST Trawl Survey. Fishermen measure individuals of a given species directly over the insert on the measuring board. The tally sheet is labeled with species, trip, haul and other information. Totals for each length are then transcribed onto the length frequency data log (Figure 20). ................................. 58

Figure 26. Schematic of the relational working data set structure. Primary fields are the minimum fields necessary to code for a unique data record (in italics under data set name). Key linking fields are those fields needed to link (merge) files together (shown next to linking arrows). Arrows indicates primary links between files (others are possible). All are one-to-many except for that between station and the derived catch files, which is one-to-one. ................................. 59

Figure A.1. Vendor information form to set up contracts between SMAST and collaborating fishermen/vessels. ................................. 175

Figure A.2. Template for formal contract between SMAST and collaborating fishermen/vessels. Page 1. ................................. 176

Figure A.3. Template for formal contract between SMAST and collaborating fishermen/vessels. Page 2. ................................. 177
Figure A.4. Template for formal contract between SMAST and collaborating fishermen/vessels. Page 3. .................................. 178
Figure A.5. Template for formal contract between SMAST and collaborating fishermen/vessels. Page 4. .................................. 179
Figure A.6. Template for formal contract between SMAST and collaborating fishermen/vessels. Attachment A. .................................. 180
Figure A.7. Invoice template to compensate fishermen for fishing trip. .............................. 181
Figure B.1. Environmental log spreadsheet record. .................................. 184
Figure B.2. Fish haul log spreadsheet record. .................................. 185
Figure B.3. Species length frequency log. .................................. 185
Figure B.4. Vessel and trip information log. .................................. 185
Figure B.5. Gear information log record. .................................. 185
Figure B.6. The raw Tidbit temperature record. .................................. 186
Figure B.7. Trawl temperature data with position and time. .................................. 186
Figure B.8. Merged trawl position, temperature and fish data. .................................. 187
Figure B.9. Merged trawl position, temperature and fish data #2. .................................. 187
Figure C.1. Flow chart illustrating data processing procedures. Superscripts are explained in the text. .................................. 189
Figure C.2. Trip info log. .................................. 190
Figure C.3. Trip info logs.xcl. .................................. 190
Figure C.4. TRAWLXXXENVLOG4.xcl, Align. .................................. 191
Figure C.5. TRAWLXXXP_ENVLOG4.xcl, Environmental Log. .................................. 192
Figure C.6. An exploded view of the links and formulas used to link Align and Environmental Log on Line 11 on TRAWLXXXP_ENVLOG4.xcl. Note, the links and formulas are expressed on line 11. .................................. 193
Figure C.7. TRAWLXXXP_ENVLOG4.xcl, Trawl###P_, Time-Posit Data. .................................. 194
Figure C.8. An exploded view of the links and formulas on line 11 used to link ENVIRONMENTAL LOG and TIME-POSIT Data on TRAWLXXXP_ENVLOG4.xcl. .................................. 194
Figure C.9. TRAWLXXXP_ENVLOG4.xcl Posit Plot. The above pictured ship track is an example of how the POSITPLOTS sheet acts as a visual check of haul location validity. .................................. 195
Figure C.10. Trawl###P.txt. .................................. 196
Figure C.11. Haul data (TRIP##).xcl, Sheet 1. .................................. 196
Figure C.11. Raw haul data (TRIP##).xcl. .................................. 196
Figure C.12. Sheet 2 of Haul Data(TRIP##).xcl. .................................. 197
Figure C.13. Trawl###F.txt. .................................. 197
Figure C.14. Raw TRAWL_LENGTH_FREQ.xcl. .................................. 197
Figure C.15. Trawl_length_freq,xcl, for text. .................................. 198
Figure C.16. Trawl_length_freq.txt. .................................. 198
Figure C.17. Boxcar 4.0 software displaying the dtf file, Trawl_301_Isabel S_A. .................................. 198
Figure C.18. The Tidbit temperature monitor, optical reader, and some related
products. .......................................................... 199

Figure C.20. Trawl_###Boatname_A.txt (or B). ....................... 200
Figure C.21. Trawl###A.txt (or B). .................................. 200
Figure D.1. Environmental log. ................................... 202
Figure D.2. Typical haul log. ..................................... 203
Figure D.3. Typical length frequency log. ......................... 205
Figure E.1.1. Sample of environmental log version 1. .......... 216
Figure E.1.2. Sample of environmental log #2. .................. 217
Figure E.1.3. Sample of environmental log #3. .................. 218
Figure E.1.4. Sample of environmental log #4. .................. 219
Figure E.1.5. Example of environmental log #5. ................. 220
Figure E.1.6. Example of environmental log #6. ................. 221
Figure E.1.7. Example of environmental log #7. ................. 222
Figure E.1.8. Example of environmental log #8. ................. 223
Figure E.1.9. Example of environmental log #9. ................. 224
Figure E.1.10. Example of environmental log variation #10. .... 225
Figure E.1.11. Example of environmental log variation #11. ... 226
Figure E.2.1. Example of haul log variation #1. ................. 228
Figure E.2.2. Example of haul log variation #2. ................. 229
Figure E.2.3. Example of haul log variation #3. ................. 230
Figure E.2.4. Sample of haul log variation #4. ................. 231
Figure E.2.5. Example of haul log variation #5. ................. 232
Figure E.2.6. Example of haul log variation #6. ................. 233
Figure E.2.7. Example of haul log variation #7. ................. 234
Figure E.2.8. Example of haul log variation #8. ................. 235
Figure E.2.9. Example of haul log variation #9. ................. 236
Figure E.2.10. Example of haul log variation #10. .............. 237
Figure E.2.11. Example of haul log variation #11. .............. 238
Figure E.2.12. Example of haul log variation #12. .............. 239
Figure E.2.13. Example of haul log variation #13. .............. 240
Figure E.2.14. Example of haul log variation #14. .............. 241
Figure E.3.1. Example of length frequency log variation #1. ... 243
Figure E.3.2. Example of length frequency log variation #2. .. 244
Figure E.3.3. Example of length frequency log variation #2. .. 245
Figure E.3.4. Example of length frequency log variation #4. ... 246
Figure E.3.5. Example of length frequency log sheet variation #5. ... 247
Figure E.3.6. Example of length frequency log variation #6. ... 248
Figure E.3.7. Example of length frequency log variation #7. ... 249
Figure E.3.8. Example of length frequency log variation #8. ... 250
Figure E.3.9. Example of length frequency log variation #9. ... 251
Figure E.3.10. Example of length frequency log variation #10. ... 252
Figure E.3.11. Example of length frequency log variation #11. ... 253
Figure F.1. The first version of the haul log used by the SMAST Trawl Survey during year 1. .......... 229
Figure F.2. Second haul log used in the SMAST trawl survey. ............... 230
Figure F.3. Third variation of the haul log. .................................. 231
Figure F.4. Fourth haul log. .................................................. 232
Figure F.5. Fifth haul log used. ............................................. 233
LIST OF TABLES

Table 1. List of twenty vessels that collaborated in the High-Resolution Trawl Survey ................................................... 61
Table 2. Types of data logs collected during fishing operations .................. 63
Table 3. Fishing trips included in the SMAST Trawl Survey data. Trip is an arbitrary number assigned to each fishing trip. Vessel code is a randomly generated code given to each participating fishing vessel to insure confidentiality .............................................. 64
Table 4. Types of otter trawls used during the survey. Brand names or standard net type provided by the vessel captain. These generally fall into three major categories: flat nets, hi-rise nets, and roller/rock hopper nets. . . . 78
Table 5. Description of fields labeled on the vessel data field log sheet (Vessel log). The logsheet was derived by combining elements of two logs from the NMFS Observer Program (Vessel and Trip Log and the Vessel and Trip Log - Six Month Questions). Because this field log was derived from existing NMFS Observer logsheets, it retains fields that were not used during the SMAST Trawl Study. The “field identifier” is the label found on the log sheet. ......................................................... 80
Table 6. Description of fields labeled on the gear data field logsheet (gear log). The “field identifier” is the label found on the log sheet. ........... 83
Table 7. Description of fields labeled on the trip data field logsheet (trip log). This logsheet was derived the NMFS Observer Vessel and Trip Information Logsheet. The “field identifier” is the label found on the log sheet. .................................................... 88
Table 8. Description of fields labeled on the environmental data field logsheet (environmental log). * Local time = Eastern Standard Time +/- Daylight Savings Time. ............................................. 89
Table 9. Description of field labels on the haul data field logsheet (haul log). "Not used*" refers to log sheet fields which were carried over from the original NMFS log sheets used as templates or from previous versions of the SMAST log sheets, but were not used (i.e., no data recorded). The “field identifier” is the label found on the log sheet. ............... 91
Table 10. Summary of major changes made to the haul logs through to the end of the 2003 season. ......................................................... 95
Table 11. Description of fields labeled on the length frequency field data logsheet (trawl length frequency log). ............................................. 96
Table 12. Description of the trawl survey master data set tables. ............... 97
Table 13. Description of variable fields contained within the master vessel data file (trawl_vessel_log_05May04.xls). These data were keypunched from data recorded on the vessel data field logsheet (Table 5), but may include modifications and additions inserted during data processing. Fields used are modified from the NMFS Observer Vessel and Trip log and the Vessel
and Trip Log - Six Month Questions. Forms: V=field used in the NMFS Observer Program Vessel and Trip Information Log, VM = V form field after modification by SMAST, S= field unique to SMAST project, 6=field used in the NMFS Observer Program Vessel and Trip Information - Six month Questions, 6m= 6 form field after modification by SMAST.

Table 14. Description of variable fields contained within the master gear data file (trawl_gear_log_12Aug04.xls). This data was keypunched from data recorded on the gear info data field logsheet (Table 7), but may include fields added during data processing.

Table 15. Description of variable fields contained within the master trip data file (trawl_trip_log_17Aug04.xls). These data were keypunched from data recorded on the trip data field logsheet (Table 6), but may include changes and additions made during data processing. Fields are based on the NMFS Observer Vessel and Trip Information Logsheet.

Table 16. Description of variable fields contained within the master environmental data set (trawl_env_log_21jul04.xls). These data were keypunched from data recorded on the environmental data field logsheet (Table 8), but may include fields added during data processing.

Table 17. Description of variable fields contained within the master haul log data set (trawl_haul_log_05May04.xls). These data were keypunched from data recorded on the haul log data field logsheet (Table 9), but may include fields added during data processing.

Table 18. Description of variable fields contained within the master length frequency data file (trawl_length_frequency_log_05May04.xls). These data were keypunched from data recorded on the length frequency field logsheet (Table 20).

Table 19. Description of lookup tables for coded fields found in the trawl master data sets.

Table 20. Description of fields in the master data file Captain_code (trawl_captain_cod_10sep04.xls).

Table 21. Description of fields in the master data file Vessel_code (trawl_vessel_code_10sep04.xls).

Table 22. Description of fields in the master data file Target_species (trawl_target_species_10Sep04.xls).

Table 23. List of species or species groups targeted by commercial fishermen participating in the SMAST survey.

Table 24. Description of fields in the master data file Species_categories (trawl_species_categories_10sep04.xls, see Table 25).

Table 25. Species and species group categories that occurred in the SMAST trawl data during years 1 and 2. (Note, species market categories such as whole fish, gutted, tails, etc. were originally identified in the data with a combination of the dressed codes and species codes. The speccat code was created to simplify the coding of the market category/species codes. However, after quality control processing only monkfish have a speccat
code different from the species code.) .......................... 123
Table 26. Description of fields in the master data file for species (trawl_species_10sep04.xls). ................................. 127
Table 27. Standardized species name and code list used during the survey. The codes and species list are modified from the codes used by the NOAA Fisheries Observer Program. ................................. 128
Table 28. Field definitions for master data file species_category_conversions (trawl_species_category_conversions_10sep04.xls). ................................. 130
Table 29. Species codes and weight conversion factors for all species and species groups encountered in the survey. A conversion factor of 1 is applied to species weights recorded from whole fish, while species weights recorded after dressing must be converted to whole weight equivalents by multiplying by the appropriate conversion factor obtained from the NOAA Fisheries Observer Program. A dressed code of "302" indicates the sample was dressed prior to obtaining a weight estimate, while a code of "301" indicates that the weight estimate was obtained for whole fish. ........................................................ 131
Table 30. List of trips with unreliable discard data with technician comments. Discard data was set to missing for these trips. ................................. 142
Table 31. Species reported in the data that required conversion to a higher taxa level or to a mixed species group during data quality control processing due to the uncertainty of field identifications. ................................. 143
Table 32. Description of the trawl survey working data set tables. These data sets are derived from the trawl master data and are formatted for easier use by investigators and for construction of a relational data base. ................................. 144
Table 33. Description of lookup tables for coded fields found in the trawl working data sets. .................................................. 147
Table 34. Description of variable fields contained in the primary working data set vessel (trawl_vessel_26Aug04.xls). This data is derived almost directly from the master vessel data file (Table 11), but may contain changes inserted in the SAS Program 1 (Appendix I). ................................. 151
Table 35. Description of variable fields contained in the primary working data set gear (trawl_gear_26aug04.xls). This data is derived almost directly from the master gear data file (table 13), but may include modifications resulting from data processing in the SAS Program 1 (Appendix I). ................................. 153
Table 36. Description of variable fields contained in the primary working data set trip (trawl_trip_26aug04.xls). This data is derived almost directly from the master trip data set (table 12), but may include modifications resulting from data processing in the SAS Program 1 (Appendix I). ................................. 156
Table 37. Description of variable fields contained in the Primary working data set station (trawl_station_26aug04.xls). This data is derived from the master env log data file (table 14) and the master haul log data (table 29). The data has been modified to remove tow event data and insert aggregate catch data by the SAS Program 1 (Appendix I). ................................. 157
Table 38. Description of variable fields contained in the primary working data set haul_event (trawl_haul_event_26aug04.xls). This data is derived from the master env log data file (table 14). The data has been modified to insert a field to indicate the temporal sequence of haul events by SAS Program 1 (Appendix I). 163

Table 39. Description of variable fields contained in the primary working data set catch (trawl_catch_26aug04.xls). This data is derived from the master haul log data file (table 15). The discard_data_code is derived from the master trip info log data file (table 12). Total catch and total discard data are set to missing when discard_data_code=2. The data has been modified to remove discard by reason data, and to determine species totals for catch, kept and discard by summing over species categories and applying weight conversion factors, by the SAS Program 1 (Appendix I). 165

Table 40. Description of variable fields contained in the Primary Working Data Set Discard_by_Reason (Trawl_Discard_by_Reason_26Aug04.xls). This data is derived from the Master Haul Log Data File (Table 15). This provides discard by reason data for the subset of the catch data where discard reason was recorded, often when multiple reasons were recorded for a single species. Derived by the SAS Program 1 (Appendix I). 166

Table 41. Description of variable fields contained in the primary working data set length frequency (trawl_length_frequency_26aug04). This data is derived almost directly from the master length frequency data file (table 21). Note that the variable speccat has not been converted to species. Therefore, it is necessary to combine species categories to obtain length species for a given species (use the trawlspeciescategorytable23aug04.xls in the lookup table directory). 167

Table 42. Field definitions for the derived catch files working data set total catch (trawl_total_catch_23aug04.xls). This is an example of the "derived catch" tables created from the "primary working data set" catch (trawl_catch_26aug04.xls) by the SAS Program 1 (Appendix B). Note the prefix "T" is used in front of the species codes to form the catch variables for each species. The other derived tables are the same, except for the prefix: "T" - Total catch, "TC" - Total catch CPUE, "K" - Kept catch, "KC" - Kept catch CPUE, "D" - Discard Catch, "DC" - Discard catch CPUE, "PD" - Percent discard of a species catch, "DP" - Percent discard of total catch. 168
INTRODUCTION

The High-Resolution Trawl Survey project was established in November 2000 as a collaboration between the New Bedford commercial fishing industry and the School for Marine Science and Technology (SMAST) of the University of Massachusetts Dartmouth. One of the major incentives for the project was to respond to the desire of the commercial fishing community to participate more directly in the management of the resource. The primary focus of the project has been to create an industry-based program for the collection of trawl catch and environmental data suitable for use by researchers and fisheries managers. The design and implementation of the project relied on the ability of SMAST researchers to design a data collection system that made sense to fishermen and produced scientifically sound data. Goals of the project included:

- to obtain fisheries data with higher spatial and temporal resolution than available through industry-independent surveys;
- to involve industry in collection of fisheries data and management of resources;
- to develop methods for training commercial fishermen to record scientifically acceptable data during normal fishing operations;
- to demonstrate the feasibility of a cooperative project between the fishing fleet and scientists;
- to characterize the effects of environmental conditions and fishing operations on spatial and temporal patterns in catch statistics.

The purpose of this report is to document in detail the methods used to collect, process and archive the resulting data.

MATERIALS AND METHODS

The data collection goal of the project was to record fishing data for individual tows made during normal fishing operations for commercial trawlers operating on Georges Bank. Fishing data included catch, vessel, fishing operations, and environmental. The work has been in collaboration with 20 vessels from the New Bedford fleet of trawlers that fish primarily on Georges Bank (Table 1). Trawl data collections began in November 2000 and continued for a period of 12 months until temporarily stopping after October 2001. After a six-month gap (November 2001-July 2002), sampling resumed in August 2002 and continued through 2003. This report will cover the first two 12-month periods of data collection (November 2000 - October 2001 and August 2002-July
It is important to note that the participating vessels operate under a standard commercial fishing regime; they follow all normal regulations and receive no special research permits. The captains choose the methods, gear, location, and time of each trawl, based on their personal knowledge and commercial considerations. While vessels that participate in the project are paid $300/day to collect the required data, their primary role at sea is to catch fish. The vessel captains oversee data-logging activities by the crew, and are important participants in the data collection effort. SMAST researchers provide information and guidance on sampling techniques, log keeping, and observed results, and oversee the quality of the information returned. The returned log sheets and electronic sensors are processed by SMAST technicians using the methods described in this report. Since the fishing locations and catch quantity are implicitly sensitive information among economically competing commercial vendors, data availability is controlled during the processing cycle and stored on computers to which access is restricted. Except for a list of participating vessels (Table 1), vessel and captain identities are "blinded" in all data sets and reports produced by this project to preserve the confidentiality of vessel operations.

Fishing Boats

The participating New Bedford fishing vessels are groundfish trawlers that traditionally target cod and haddock during the spring and summer and flatfish during winter months (Table 1). The participating vessels’ primary fishing grounds extend from Nantucket Shoals through the Great South Channel across Georges Bank. In early October 2000, vessels associated with the New Bedford Trawler Survival Fund (TSF) began to work collaboratively with SMAST to collect data while fishing. A TSF representative to the project, a fisherman and vessel owner, was instrumental as a liaison between the fleet and SMAST. The TSF boats, which are in the 75-100 foot range, are required to have current Coast Guard safety inspection. These boats normally carry a crew of five. Roughly half of the participating vessels have crews of Portuguese-speaking fishermen that are immigrants or descendants of fishermen from traditional Portuguese fishing ports found in the region between Porto in the north to Figueira da Foz in the south central region of Portugal. A smaller number come from the Azorean archipelago, Madeira and the Cape Verde Islands. The vessel captains oversee the crew data-recording activities and participate in the data collection themselves. For this project, the vessels were paid $300/day to collect the required data once they had signed a contract with the University. Due to budget constraints, only the seven vessels that performed the most complete and reliable data recording were retained for subsequent years. For a vessel to become part of the SMAST High Resolution Trawl Survey, a contract and tax information was needed to define the business relationship between SMAST and the vessel’s owners and crew members. Information on vessel sign-up and contracting are
Vessel Inspection and Training

Steps taken to train fishermen to collect fisheries data during commercial fishing operations are outlined in Figure 1. Once a vessel captain schedules a fishing trip, SMAST technicians assemble a package that includes a set of five different log sheets (Table 2), an observer manual, a common fish identification handout, and two activated Tidbit temperature recorders. They are also provided with hand-held scales and measuring boards. Before a vessel can independently collect data, it must first undergo a training trip on proper data collection and recording methods. Training trips are otherwise normal fishing trips in which an SMAST technician participates and provides training to the captain and crew. All vessels receive one training trip prior to beginning in the program, and at least one additional training trip during each year of the program. Training consists of an iterative process whereby the SMAST technician observes the fishing activities, instructs the captain and crew in the proper recording methods, and reviews the results following each tow for the duration of a fishing trip. After the initial training trip, the captain and crew record data without the aid of a technician during fishing operations on subsequent trips. The crews are initially trained by SMAST technicians when they join the project, and the technicians provide additional training and feedback on the quality of data collected thereafter. SMAST researchers provide information and guidance on fish identification, sampling techniques, log keeping, and observed results. The training specifically involves going over each of the steps necessary to collect the data, as well as numerous practices in filling out all the log sheets. The time is recorded on the logsheets as Eastern Standard Time (EST) local time on a 24-hour clock. Information on the vessel and gear used are also recorded for each vessel. SMAST observers participated on approximately 18% of the 165 trips conducted during the two survey years (21 out of 115 in year 1, and 9 out of 50 in year 2; see Table 3).

Procedures and events for a typical training trip are outlined on the top portion of Figure 1 labeled “Setting Sail.” A meeting is arranged where introduction to the owner, captain, or both can be completed. The program is explained to the vessel’s representative. The vessel representative is required to give a tour of the vessel to an SMAST technician. During the tour the technician performs a general inspection to confirm that the vessel is seaworthy and looks for a current United States Coast Guard inspection sticker, life raft and other safety equipment. Once those criteria are established and the required paperwork (contract, tax information) is completed and recorded, a training trip is arranged. The vessel agrees to house, feed, and accept direction from an SMAST technician for the duration of a normal groundfish fishing trip in the Georges Bank fishery. The day the fishing trip is scheduled to depart, the SMAST technician meets the vessel at the dock. The technician is given a bunk and place to stow personal and
work-related gear. Other preparation includes placing Tidbit temperature
monitors on the fishing gear and describing what on-deck and wheelhouse duties
will be required from the captain and crew on the training and subsequent fishing
trips.

Overview of Field Operations

Data collection and processing is a complex task. Before describing data
collection and processing for each type of data, a brief overview of the entire
process will be presented below. The middle portion of Figure 1 provides a
schematic flow chart overview of data collection for the “setting out,” “haul-
back” and “post-haul” stages of trawl tows. Before the ship leaves the dock, the
captain is given a packet of logsheets, measuring boards, and length-frequency
template sheets, and Tidbit temperature data loggers are placed on either the trawl
doors or the nets. In the first year, the Tidbit recorders were generally placed on
the nets, while in subsequent years, they were typically placed on the trawl doors.
The captain generally knows what fishes he will be targeting for the trip before
setting sail, but target species can change from haul to haul depending on fishing
success, economic news and chance catches. The captain readies the logsheets
once on the fishing ground and a trawl tow is about to take place. The location
and time when the net is "set" is recorded on the environmental log. The net is set
when it reaches the bottom, the doors have spread out the desirable distance, the
brakes on the winch have been set stopping the free spooling of warp/wire and
the engine is put into gear initiating towing. Trawl tows were typically of three-
hours duration, but could be as long as 10 hours. Towing speeds were typically
around three knots, but varied by vessel, gear and target species. During a tow it
was often necessary for the captain to change course by turning about. During
turns the net is typically, but not always, raised off the bottom to avoid fouling
and tangling, but it is not taken on board ship. How high the net is raised depends
on the circumstances and is at the captain's discretion. Captains were instructed
to record all the data required on the environmental log (i.e., location, time, depth,
wire out, weather, etc.) for each haul event. Haul events were defined to include
the set, major turns, and the haul-back. The time and location for the start of the
tow were recorded at the time of the set, while the time and location for the haul-
back were recorded at the time the net was begun to be raised (i.e., when the
wrench brakes have been released and the wire/warp has begun to be reeled in).
Once the codend is brought on board, the catch is released into a checker box or
directly onto the deck for processing. Since this is a commercial fishing
operation, the crew has often developed detailed fish processing routines that
insure economic efficiency. Fishermen make minimal adjustments to their
routine in order to collect scientific data.

Fishes to be kept for sale are culled from the catch and placed either into
standard fish baskets or fish boxes. Fishes to be discarded are typically quickly
swept, shoveled or thrown overboard. It is in the fishermen's interest to do this as
quickly as possible to: 1) decrease mortality of discarded species to the extent possible, and 2) to minimize the loss of valuable time and energy better reserved for fishing and processing the kept fish. For these reasons, data on the discard species are poorly quantified. Fishermen estimate the total weight of each discarded species. They further partition the discard into weights for specific discard reasons. For example, it’s possible that a given fish might have both a minimum size limit and a maximum size limit. In that case some fish might be discarded that are too small or too large. It must be recognized that similar looking fishes of no economic value may not be distinguished by the fishermen (e.g., the various sculpin species). Likewise, similar looking species that are economically important, but that are sold under the same market name, are usually not distinguished by the fishermen. For example, although fishermen can be trained to distinguish red and white hake, it is not economically feasible for them to take the time and effort to do so during commercial fishing. Therefore, red and white hake are classified as a "red and white hake mix."

Many species of fishes kept and processed by the fishermen are dressed (i.e., butchered and cleaned) in various ways according to market demands. For this reason, the National Marine Fisheries Service has developed correction factors to convert the weight of dressed fish to an estimated whole fish weight. The SMAST trawl survey program uses the NMFS conversion data tables with a few minor changes (details can be found in a separate methods report). Because the SMAST High-Resolution Industry-Based Trawl Survey was initially modeled after the National Marine Fisheries Service (NMFS) Observer Program, we strongly recommend readers to review of the NMFS Observer Manual (current copies can be downloaded at http://www.nefsc.noaa.gov/femad/fsb/). Fishermen may estimate the weight of the catch for a given species either before or after dressing. The method used can change from haul to haul, and even between species within a given haul. Fishermen estimate weights based on visual inspection of the number of baskets or fish boxes that are filled or partially filled. It should be noted that fishermen have developed very specific methods of filling baskets and boxes because their livelihood depends on efficiency. Therefore, a packed basket or box of a particular species tends to be a standard quantity, and weights are relatively efficiently estimated. To date quantification of the weight estimation error made by fishermen during this survey has not been attempted. In the future, we hope to obtain species-specific and dressed-specific data on weight measurement variation.

Lastly, a sub-sample of one species of fish is collected during the sorting of the pile on roughly half of all hauls for use in length-frequency data. The fish are collected from the main pile in such a way as to ensure that the sample is as randomly distributed as possible. The project guidelines ask that thirty individual fish be collected per tow when a length-frequency is to be performed. In the first year of the project any species the fishermen chose was acceptable. Due to some very small sample sizes of individual species collected in year one, in year two the project asked that only six species (codfish, haddock, yellowtail
flounder, winter flounder, plaice flounder, and grey sole flounder) be considered for measuring. The actual measuring of fish was performed on a wooden measuring board and recorded in cm on a reusable template.

After all on-deck duties have been completed, the haul information is recorded on a haul log. The length-frequency data is transcribed from the measuring board onto a length-frequency data log. Usually a haul log is then prepared for the next haul. This set of processes is completed for every haul until the trip ends.

SMAST trawl project duties that must be completed by the fishermen at the end of a trip are outlined on the bottom section of Figure 1 labeled “ending the trip.” These duties include removing the Tidbit temperature data loggers from the fishing gear and stowing them safely with the rest of the SMAST trawl project data. All data sheets are brought to a central location. During the fishing trip some of the data sheets are used in the wheelhouse by the captain and mate, and some are generally kept in the galley for the deck crew to complete. A crew member places all data logs and Tidbit data loggers in a package to be collected or delivered to an SMAST technician upon landing in port.

Overview of Data Keypunching and Quality Control

Data keypunching from the field logs and subsequent quality control auditing of the trawl data has undergone a series of changes since the inception of the project. A summary of data processing procedures used in the first year of the study is provided in Appendix B. Data processing steps used through the end of year 2 are summarized in Appendix C. During the first year and a half, all data generated by the project were keypunched by SMAST technicians, students and summer interns. Starting with Trip #223 (January 2003), keypunching was outsourced to Trade Quotes, Inc. (Boston, MA). Trade Quotes provided a professional keypunching service and significantly improved the quality of data entry. A summary of keypunching instructions provided to Trade Quotes by SMAST is provided in Appendix D. As a quality control measure, the company was required to double keypunch all data and then to cross-reference the resulting files to identify keypunching errors. Double keypunching of data is an effective way to minimize keypunching errors. However, the company was also instructed to keypunch written fields exactly as they appeared on the log, even when the data was known to be in error. This was required as it was felt that the professional keypuncher lacked adequate knowledge of the fisheries and fisheries data to make interpretive decisions regarding data. Therefore, all “problem” records were subject to review and editing by SMAST technicians that were capable of interpreting the data. In either case, SMAST technicians provided some degree of preliminary quality control by reviewing logsheets for obvious errors prior to keypunching.

Once data was keypunched, it was then subjected to two different levels of quality control auditing. During the first year and a half of the project, data were
checked for errors at the trip level (see Appendix C). Basically data were
keypunched into separate files by trip number. These files were then examined in
an excel spreadsheet and with various Matlab programs to check for common
errors. For example, one of the checks was to plot the location of each haul to
verify that the location information was consistent among hauls within a trip.
However, due to temporary loss of funding and several changes in program
management, no extensive auditing of the data was conducted until after the 2nd
year of the survey had been completed. After this time, data from all trips were
combined into a single file for each data type (trip log, environmental log, haul
log, etc.). Then a series of steps were taken to examine the data in its entirety for
statistical outliers, unusual values, inconsistent records, and the integrity of data
links between related files. All numeric fields were examined for trends that
might indicate data errors. Usually the 1st and 99th percentile values were checked
against the actual log sheets. Frequency analysis was used to detect any unusual
modes that would then be checked. Character fields were checked for spelling
variations, and were also subjected to frequency analysis. Coded fields were
checked for valid code values, and any uncommonly occurring code values were
checked.

A great deal of time was spent verifying the linkage between the
environmental log data and the haul log data. Because the environmental log and
the haul log contained an internal hierarchy of data, it was necessary to check for
consistency of fields among levels in the hierarchy. For example, the target
cpecies value should be the same for all records in a given haul, and the
total_pounds_discarded field should equal the sum of the weights recorded for all
discard reasons for a given species and haul. Data auditing was first carried out on
each data table separately (e.g., trip, environmental data, haul data), and then
related tables were merged to check for the link integrity. For example, if there
were data for cod length frequencies in the length-frequency table, there should
be a corresponding entry for cod catch in the haul log. All audit checks were
repeated in an iterative fashion until no errors were encountered. Typically, audit
checks took the form of an SAS program that performed various manipulations
(merges, reductions, transpositions) of the data, identified statistical outliers or
high and low values, and checked for linkages between data tables. At other
times, similar checks were done using the index function in an excel spread sheet.
For example, if a particular field was supposed to contain one of three possible
values, the presence of invalid values could easily be detected with the index
function.

Previous auditing with MATLAB programs was superceded by auditing of
combined trip files conducted after year 2. For various reasons, it became
necessary to use the combined trip files (i.e., all data for all trips appended
together into a single file for each data type: environmental log, haul log, length
frequency, etc.) as the master data after auditing was completed. The original
trip-by-trip files have been archived, but no longer maintain the data integrity.
Basically because of all of the data editing, and the loss of some intermediate
files, it is not possible to revert to the original trip-by-trip files without losing most of the data edits made during the auditing process. To retain these edits, we would have to generate new trip-by-trip files from the master data files.

**Trawl Gear**

The New Bedford Massachusetts trawler fleet uses various types of otter trawls throughout the year (Table 2). Otter trawls are large funnel-shaped nets that are pulled through the water behind the vessel. What makes an otter trawl effective is its ability to open wider than the stern of the vessel as it is being towed. This is achieved by attaching the net to two "otter boards" or "doors" that spread the net apart due to hydraulic pressures as they are towed behind the boat. In addition, the top opening of the net is made to be positively buoyant while the bottom opening is made to be negatively buoyant. These factors cause the net to open vertically and horizontally.

**Otter trawl components**

The typical otter trawl net is cone shaped and has the following parts:

*Warp or wire* - steel wire, usually 1-1.5" in diameter, going from the vessel back to otter trawl doors. The length of warp used is an important factor in how the net performs. Warp length is determined by depth, and to some degree by vessel speed, weather and gear type.

*Otter boards/doors* - rectangular "doors" used to spread the net. In addition, on a bottom trawl (trawl that fishes on the sea floor) the doors create a lot of noise and a mud cloud on the sea floor that acts to herd fish in front of the net until they grow tired and fall back into the net. The size and weight of the doors used depend on the type of net deployed and the fishing power of the vessel. Using doors that are too big or too small for a given net will result in poor gear performance.

*Ground cable* (optional) - extends from the doors back to the bridle. Most modern ground cable is actually 1" or 1.5" cable with rubber cookies covering it. Ground cables may or may not be used depending on what the vessel is targeting. The ground cable affects the catch through its influence on fish behavior. The warp affects fish behavior in several ways. First it increases the mud cloud otherwise generated by the doors. Second, the wire itself can herd species that exhibit visual and/or acoustic fright responses to the wire. Thus the ground cable can serve to increase the effective swept area of a net to a larger area where fish are herded into the net (herein termed the "herding area"; see Fig. 2). The effectiveness of the ground cable in fish herding is dependent on the fish's behavioral response to the fright stimulus and its swimming speed. For example,
slow-moving fish like monkfish may be frightened by the doors and ground cable, but are less likely to be herded into the mouth of the net because they swim slowly relative to the speed of the oncoming net (Fig. 3). In contrast, faster swimming fishes such as cod can be herded from a greater distance. The effect of the ground cable on fish behavior has received relatively little attention, and would be an interesting area of investigation.

**Tickler (optional)** - a chain strung between ground cables and used to rile the bottom and stir any fish that may be dug into the bottom substrate so as to make them more vulnerable to the net (Fig. 2). The tickler chain can be placed anywhere along the ground cable from just in front of the footrope to just behind the doors (if the ground cable is relatively short). Unfortunately, data on the use of tickler chains have not been recorded during this project, as they are not recorded in the parent NOAA Observer Program log sheets. They were used most often with flat nets when monkfish or flounders were targeted. However, tickler chain use can change on a tow-by-tow basis. Tickler chains can have a strong influence on the catch, but little is known about how species-specific avoidance responses and swimming speed interact with towing speed and the distance between the tickler chain and the footrope to determine catchability.

**Bridle** - the place where the single wire splits forming a top and bottom sections leading to the net headrope and footrope.

**Wing** - A place where the split in the bridle has netting attached to it, thus forming the net opening. Some nets have an upper and lower wing.

**Headrope** - the top of the net extending between the wings. The headrope usually has plastic floats attached to it to add buoyancy and keep the headrope above the footrope. The headrope and footrope lengths together determine the size of the horizontal opening of the net. However, the horizontal opening is typically somewhat less than the headrope and footrope lengths because the net usually has a considerable bow or sag. Of course, the horizontal opening fluctuates as the towing speed changes, and as drag forces change during the tow. The amount and placement of floatation on the headrope has a strong influence on gear performance. Too little floatation will cause the headrope to ride lower in the water than optimal. However, too much floatation can have a similar effect by increasing net drag forces resulting in the collapse of net height.

**Footrope** - the bottom extension of the bridle is negatively buoyant so as to keep the bottom of the net on the sea floor. Its length, together with the headrope length, determines the horizontal width of the net's mouth opening. Although the footrope is usually set to drag along the bottom, sometimes it is raised off the bottom through the use of rollers or other means. The height of the footrope off the bottom is an important consideration, as some fishes might
escape the net by passing under the footrope.

Sweep - the entire area of the bottom of the bridle, footrope, and any rollers that are added to the bottom of the net, though the sweep may be taken to mean any rollers and cable specifically inside of the wings. The construction of the sweep has importance consequences for net performance and the catchability of various species. Various hardware may be added to the sweep, including rollers and chains of many types and configurations.

Fishing circle - the widest opening in the net, usually measured in number of meshes.

Belly - panels of mesh sewed together between the net mouth (bridle, sweep, etc.) and the codend.

Codend - the end of the sock shaped net where the fish are collected during fishing operations and subsequently dumped from after the haul has been brought on deck. The back of the codend is drawn closed like a purse with a rope and knotted or bound with a "pucker." The end of the codend is called the "puckered end."

Variations of the otter trawl

There are a number of configurations that are commonly used in the otter trawl fishery. Nets come in different sizes (lengths, widths, and heights) for different size vessels and fishing types. In addition, nets are built with different components for varying fish and bottom types. There are a number of net types that are marketed by net manufacturing companies. Most captains customize their nets after they are purchased.

At least fourteen otter trawl variations were used during the first two years of the survey (Table 4). These can all be grouped into one of three general trawl types described below. It should be noted that most if not all of the nets used in the SMAST trawl project have been modified from these general configurations in some way.

Flat nets - are a type of Yankee trawl used primarily for flounder and monkfish and in smooth bottom (sand, mud, etc.) areas. The Yankee is the "classic" flat net that the others are derived from. Some identifying characteristics of a Yankee net include a footrope that is 20' longer than the headrope and a 9' tall mouth when fishing. In most cases, flat nets are fished with a cookie sweep or small rollers. A cookie sweep is made by attaching 3" diameter rubber discs all along the sweep (bridal, footrope, etc). When fishing for flounders, fishermen usually choose a square meshed codend. The flat net is built to stay on the bottom and sweep the entire towed area. Some specific flat
nets used in this project were the Levin Marine, flat 5" sweep, 76x96, and 76x97 (Table 4).

**Roller and rock hopper nets** - are made for fishing in the hard bottom (rocky terrain). The large rollers and rock-hopper-like sweeps allow the nets to ride, bounce, or spring over large obstructions. These nets do not stay on bottom as well as a flat net. The ability to get over obstructions without being damaged is a compromise that fishermen choose when building and designing their nets. These nets are often named after the size of the rollers used. Center roller 12", roller 18", and Yankee rock hopper nets were used in this project.

**Hi-rise nets** - These nets have a taller mouth opening than flat nets (Fig. 3) in order to maximize their ability to catch roundfish such as cod, haddock, pollock, and other demersal fishes that might avoid a flat net by swimming over the top of the headrope. A diamond meshed codend is also used in most cases on these nets. Types used in this project include the Shuman, Gurock and Balloon trawls (Table 4).

**Gear performance**

Many details of the net construction can influence gear performance and fish catchability. Some of these effects are subtle, while others are more obvious. There are at least three major factors that affect catchability: 1) encounter rate, 2) avoidance, and 3) retention. Encounter rate is the likelihood of a fish encountering the mouth of the net and is a function of interactions among the physical size of the net (net width and height), the spatial distribution of a given fish species/size class, and the speed at which the net moves through the water (and other factors related to the fish's behavior). The swept area (Fig. 2) and swept volume represent the physical aspects of the encounter rate. The swept area is the area of sea floor over which the net physically passes. It is influenced by gear performance such as fluctuations in how well the net hugs the bottom during the course of the tow, and how much the net's length and height wax and wane as the net is hauled through the water. All things considered, any fish that is present within the swept area and does not avoid the net will encounter the mouth of the net and be subject to possible capture. The swept area can be artificially enhanced through the use of ground cables and associated hardware that act to herd fish from a larger area into the mouth opening. The swept area is determined by the distance between the doors, which is a function of the ground cable length and the net width (Fig. 3).

However, the actual herding area experienced for any given fish is also dependent on the fish's reaction distance, reaction time and swimming speed. For example, the effective herding area of a slow-swimming fish like the monkfish would be expected to be much smaller than that of a relatively fast-swimming species like the cod. The monkfish would be more likely to be overtaken and
passed over by the net/ground cable before it had time to be herded towards the net mouth, while the faster-swimming cod can easily move fast enough to enter the swept area and be captured (compare "a" and "d" in Figure 3). For a given fish, its likelihood of encountering the net decreases the farther to the side of the center of the trawl track it is located (compare "b" and "d" in Figure 3). A fisherman can compensate for this somewhat by towing at a slower speed. Although these relationships are intuitively obvious, few studies have been conducted to quantify the relationships between capture rate and swimming speed, reaction rate, reaction type, distribution relative to the trawl track, swept area, herding area, and tow speed. This would be another interesting area of future research.

However, mere encounter of a fish with the net does not insure its capture. Fishes can avoid the net in many ways. Some fish may be able to avoid the net by actively swimming under, around or over the net. Others may avoid capture by reacting to the doors and ground cable in different ways, such as moving laterally away from the trawl track, or by allowing the ground cable to pass over or under them instead of being herded into the net mouth. Both visual and acoustic stimuli may be important in net avoidance.

Once in the wings and belly of the net, capture may yet be avoided due to retention effects. The most obvious of these is the mesh size. If a fish is small enough to pass through the mesh, it may escape by doing so, though many fish are captured that could have escaped because of strong net avoidance behavior, or because they are caught up in the general mass of fishes in the net and fail to find a way out. Several characteristics of the net influence the ability of a fish to escape through the netting. The type of net mesh interacts with fish body shape to influence fish retention. For example, a larger flatfish can escape from a 6-inch-stretch mesh than from a 6-inch-square mesh, while the opposite is true for a typical roundfish such as a cod. The type of twine used to construct the net also has an influence on fish retention as it can make it more difficult or easier for a fish to slip through the mesh. Twine type also has a strong effect on the gear performance by influencing net drag. Thus two nets of the same configuration, but constructed of different twine types, might perform differently and result in different catch characteristics.

**Data Collection**

**Data concept**

Data generated from any trawl survey have many commonalities. Typically trawl data have a characteristic conceptual structure, as that depicted in Figure 4. The data structure reflects a natural hierarchy in its design. Most have some version of a "trip," "station," "species" and "specimen" component. The trip (often called "cruise," as in the NMFS Bottom Trawl Surveys) is defined as all the data collected from the time a ship leaves port to the time it returns to port. However, the trip data set itself only contains the fields that describe the trip, such
as date of departure, days at sea, etc. (The content of the individual data sets will be described in more detail below.) Other data collected during the trip are entered into the appropriate data sets, all of which are nested under the trip data.

The next level is the "station" data. A station is typically defined by a location and time where and when sampling is initiated. In our case, observations are made on environmental conditions and a trawl tow is initiated. In other programs, many types of sampling may take place at a station. For example, on NMFS bottom trawl surveys, a CTD profile and a plankton net tow are made at the station, followed by trawl deployment. The location where the trawl is "set" determines the station location. Station data typically include information describing the station, such as location, time, depth, weather, etc. In the case of trawl surveys, station data also usually includes data describing how the trawl was deployed, such as the amount of wire set out, the speed and direction of the ship's travel, etc. The station data table is nested under the trip data table (actually under the appropriate trip record within the trip table). Nested under the station data table is the species data which consists of information collected for species in the catch (hence the data is often referred to as the catch table). Species data usually consist of some measure of quantity (total weight of the catch for a species or number of individuals). Finally, the "specimen" level includes any data collected on individual specimens, such as weight, length, age, stomach contents, etc.

There are two other important components of the data that are treated in many different ways depending on the organization of the trawl survey. A data set containing vessel information is often necessary. However, in some surveys where only one or two vessels participate, vessel data is usually not formalized in a data set (e.g., the NMFS Bottom Trawl Survey). In other cases, where there is a different vessel for each trip, and few vessels are used more than once, the vessel data is usually incorporated into the trip data as it logically is an attribute of the trip (e.g., the NMFS Fisheries Observer Program). In the case of the SMAST trawl survey, where many vessels are used repeatedly, the vessel data table logically constitutes of a separate data set under which the trip and all other tables are nested.

The SMAST trawl survey is somewhat unusual in that the primary sampling gear, the otter trawl, is not standardized. Data on the gear attributes (size, construction, etc.) are gathered in a separate table once a year and updated as necessary. Because the type of trawl used can change within a trip, the gear table is most properly considered a component of the station table. Finally, the SMAST trawl survey differs from many programs in the way that water temperature is collected. A Tidbit temperature data logger is mounted on the gear before the trip starts and is removed at the trip’s conclusion. The raw data, therefore, is collected on a trip basis and should be a component of the trip. However, because the logger records data continuously, whether the gear is in the water or not, much of the record is not useful. The data corresponding to each trawl tow in the trip must be extracted. The resulting temperature data sets are a
component of the station table.

To summarize in Figure 4, in concept we have a nested data structure consisting of vessel, trip, station, species and specimen levels where one record in a higher level data set corresponds to multiple records in the lower data set (one-to-many). Gear is functionally an attribute of station (one-to-one). Because the functional temperature data consists of temperatures recorded at 1- to 2-minute intervals continuously during each trawl tow, they constitute a nested data set under station (one-to-many).

Logsheets

The actual structure of the trawl data as recorded in the field diverges somewhat from the conceptual model described above, because of logistical constraints. To record all the data described in the conceptual model (Fig. 4), six types of data logsheets, and an electronic temperature data logger, were used during the survey (Table 2; Fig 5). The data structure and logsheets used in the SMAST trawl survey were modified copies of data and data logsheets used by the National Marine Fisheries Service Observer Program. Over time, some of these logsheets were modified from the original NMFS format to streamline recording in order to provide simplicity and efficiency for the fishermen. However, the logsheet, the resulting data collected, and the user data set structure have been designed for maximum compatibility with the NMFS Observer Program data.

The field logsheets can be grouped into three categories: trip, station, and species. Station and species data are nested within a trip. The trip logsheet itself simply records data describing the trip, such as date of departure, crew size, consumption of fuel and ice, etc. The vessel log can be thought of either as a higher level data group with all data from all trips nested under it, or simply as an attribute of the trip log that defines the vessel and its characteristics (length, age, tonnage, etc.) for a given trip. The trip logsheet and Tidbit temperature logger provide data collected on a per-trip basis. The gear information logsheet provides information on the type of trawl being used at each station. The gear sheet is usually recorded once for each vessel and modified as the vessel discards old gear and/or changes gear. In the future, this practice will change to allow recording of gear information that can change on a tow-by-tow basis.

**Vessel log** - The vessel logsheet is completed by an SMAST technician once during each year of the survey, but can be updated as necessary when vessel data changes (for example, addition of new electronic equipment). A near duplicate of the NMFS Observer Vessel and Trip Log is used for the SMAST vessel log (Fig. 6). Data fields on the logsheet are defined in Table 5. Because the NMFS logsheet is intended to record both trip and vessel data, many of the fields on the data sheet are not applicable to the SMAST Vessel log.

**Gear log** - The gear logsheet is also completed by an SMAST technician once during each year of the survey (Fig. 7; Table 6). Updates are made as necessary.
Trip log- The trip log sheet is also based on the NMFS Observer Vessel and Trip logsheet, but includes the trip related data not recorded on the SMAST Vessel log sheet (Fig. 8). Explanations of the data fields on the logsheet are provided in Table 7.

Station (environmental log) - The "station" is the basic sampling unit in the survey. It consists of all the data associated with a trawl tow. The terms "tow," "haul," "trawl tow" and "station" are considered synonymous for the purposes of this report. The term “haul-back” will be used to indicate the retrieval of the trawl in place of “haul” to avoid confusion. At each station, environmental data are collected and recorded on the environmental log (Figs. 9 & 10; Table 8), and the fishermen begin a trawl tow to harvest fish. The environmental log has an additional level of complexity in that separate data are recorded for each major event during the trawl tow. Data are always recorded when the net is set and again when it is hauled-back as is typical of many research surveys, including the NMFS bottom trawl survey. However, data are also recorded whenever the ship makes a turn or other major change in its course direction. Although these data are recorded on the environmental log, they actually belong to a lower level data hierarchy termed “Events” in Figure 5, and will be separated out during data processing (see below). The environmental log underwent a series of changes during the study (Appendix E.1). The greatest change occurred when an early version was restructured to improve data collection (Figs. 9 & 10).

Species (haul log)- Once the net is hauled, biological data on each species are recorded in the haul log and the length frequency log. The haul log primarily records the weight kept and discarded for each species, but includes two additional levels of complexity. First, data are recorded separately for each dressed category of a given species (see explanation below). Second, data are recorded separately by discard reason, if multiple reasons for discard occur in a haul. The haul log has undergone extensive revisions since the start of the SMAST trawl survey (Figs. 11-17; Appendices E & F). The complete series of changes are shown in Appendix E.2. Major changes are described in detail in Appendix F. Briefly, the first version was modeled after the NMFS Observer Haul Log Sheet (Figure 11). An explanation of the fields found on the haul log is provided in Table 9. Again, most fields were taken directly from the NMFS Observer Haul Log Sheet. However, it soon became apparent that the original haul log required changes to make it easier for the fishermen to understand and to eliminate data redundant with other logs. A second version of the logsheet was used throughout most of 2002-July 2003 (Fig. 12). One major change was the separation of species that are primarily kept and those that are usually discarded into separate sections of the log. This streamlined the recording for the fishermen. In addition, in recognition that the estimates of the weights of discarded species were subjective, weight bins were added to the log. Finally, additional changes have been made for year 3 to reduce confusion caused by multiple dressed type and discard reason possibilities for a given species (Fig.
A summary of major changes made to the haul log from the beginning of the project through the end of year 2 are provided in Table 10. Samples of all haul log variations, many consisting of minor cosmetic changes, are provided in Appendix E.2.

**Length frequency log** - Length frequency data were collected from species of special interest to SMAST by recording lengths on the length frequency log (Figs. 18-20; Table 11). Minor modifications to streamline the length frequency log were made over the course of the study (Appendix E.3).

**Date Entry and Processing**

**Trawl master data files**

The data recorded on the field logsheets were keypunched by SMAST technicians, student interns and student volunteers prior to January 2003, and by Trade Quotes afterwards. After extensive quality control auditing and data processing (see below), a master data set was produced (Table 12). These data included data tables corresponding to each of the logsheets, including the Vessel_log (Table 13), Gear_log (Table 14), Trip_log (Table 15), Environmental_log (Table 16), Haul_log (Table 17) and Length frequency_log (Table 18). Tables 13-18 provide a description of data fields contained in each of these master data sets. In addition, “lookup” tables were created for each of the coded fields found in the Master data. A description of these lookup tables can be found in Table 19, and the lookup tables themselves are provided in Appendix G. Several other data sets were added to the master data files for auditing and data processing purposes. Captain_code (Table 20) and Vessel_code (Table 21) data sets were added to allow the creation of random labels for captain and vessel names during the creation of the working data sets. This was necessary to safeguard the confidentiality of fishermen that collaborated with SMAST during the project. In addition, data sets containing a list of target species (Tables 22, 23), species categories (Tables 24, 25), and species collected (Tables 26, 27) were created for use in data analysis and reporting. The species_category_conversions table (Tables 28, 29) was created to allow conversion of dressed catch weights to whole body catch weights, based on data provided by the NMFS Observer Program.

The master data tables were subjected to rigorous quality control and editing checks. The contents of every field was examined for: 1) values consistent with the data field definitions (e.g., invalid species code values), 2) range extremes, 3) statistical outliers, 4) clusters of “unusual” values. These auditing procedures were conducted manually within excel, and also through SAS programming. Fields and records with incorrect, unusual, or outlier values were then manually checked against values on the paper logsheet by a SMAST technician. After any necessary corrections were made, the auditing checks were repeated. An iterative process of identifying errors and potential errors, checking them, and correcting them, followed by repeating the auditing check, was
continued until no more errors were found. The next step was to test for the integrity of the data hierarchy. For example, for a given trip, there should be only one correct value of vessel. If more than one value was found, all records were checked against the paper log to determine the source of the error. Again, an iterative process of identifying potential errors, correcting them, and then rechecking for the same errors was performed until all errors were identified and corrected. Next, the integrity of data linkages was checked. For example, vessel data from the trip data might be compared to the vessel data in the vessel data set. Obviously, vessel names unique to one data set suggest an error in one of the data sets. The various data sets were all merged with data sets above and below in the data hierarchy, and the integrity of linking fields and redundant fields was examined. It is beyond the scope of this report to describe all these types of audit checks; however, some of the more important examples will be described below.

A brief summary of the details of the data processing is provided in Appendix C (also see Appendix B). The summary is intended to serve as a stand-alone document for those persons interested in a quick review of the major details of the processing up to the point of production of the raw data files that were subsequently converted into the master data files after extensive quality control auditing of the aggregate data sets (2 years combined). Additional details of the initial processing and quality control auditing are presented below for each major component of the master data files.

**Vessel log processing**

The vessel log sheet documents the characteristics of the vessel and its equipment. It is based on both the "Vessel and Trip Information Log" and the "Vessel and Trip Information Log - Six Month Questions" forms from the NMFS Observer Program. Because the SMAST trawl survey project repeatedly sampled from the same vessels in a relatively small pool of participants (20 vessels in year 1 and 7 in year 2), it was decided to remove much of the largely redundant information from the NMFS Observer forms by splitting the vessel and trip forms into separate vessel and trip field logsheets. Further, some data from the NMFS six month question form was incorporated into the SMAST vessel log. However, many of the trip-related fields remained on the logsheet, despite not being used. This could potentially result in confusion if the fishermen were required to complete this form themselves, but since SMAST technicians were responsible for completing the vessel log, this was not an issue. In all, the SMAST forms were much reduced in complexity compared to the NMFS forms, but collected less economics data. A sample of the SMAST vessel logsheet is provided in Fig. 6. Table 5 provides a brief description of data fields collected. Since there are numerous electronic systems and many manufacturers for each system, these components are documented to provide the distinct combination of components aboard each vessel. The vessel characteristics include information about the construction of the vessel hull and engine characteristics.
The vessel log sheet is generally filled out once per year, as the characteristics of the vessel do not change during the course of a single fishing trip. Often this information was recorded during the initial walk-through inspection of the vessel or during the initial training cruise. The definitions and descriptions of the fields are shown in Table 5. If changes are made to the vessel, updates are tracked by filling out and dating additional sheets as necessary. Much of the information comes from documentation (licenses, including NMFS, U.S. Coast Guard, etc.). The accuracy of the information provided by the captains to SMAST is verified by comparing with information available from NMFS. Most of the pertinent information about the vessels in the project can be referenced at: http://www.st.nmfs.gov/st1/commercial/landings/cg_vessels.html.

Most of the fields in the vessel log come directly from the NMFS Observer Vessel and Trip Information Log and Vessel and Trip Information - Six Month Question Log. However, in several cases, SMAST enters character values instead of using the NMFS numeric codes (e.g., second_engine, trip_type, etc.). The “NMFS”, “typical_crew_size” and “gear_number fields” found on the SMAST vessel logsHEET are unique to the SMAST trawl survey. The “NMFS” field contains the National Marine Fisheries Service vessel documentation number and is added by SMAST technicians during data keypunching. The typical number of persons crewing the vessel is provided in the typical_crew_size field. Each net used by the vessel is given an identification number (gear_number: e.g., 1, 2, 3, etc.) to allow cross references of the environmental and haul logs with the gear log. As depicted in Figure 21, the master vessel log data set is created directly from the field log during keypunching. However, some field names are changed from the more descriptive logsheet labels to shorter data set variable names (compare Tables 5 and 13).

**Gear log processing**

Though regulations exist which specify many of the legal requirements for the fishing equipment used by the trawlers in the pursuit and capture of ground fish, many variables affect the fishing efficiency of each vessel. The gear log records the properties of the trawl fishing setup used aboard each participating vessel. The gear log is filled out by SMAST technicians in consultation with the vessel captain once each season. If gear is modified or new gear is added to a vessel it is included in the gear log under the “gear number” field and is keypunched as net_number. Net_name encodes a specific net brand name or type designation provided by the vessel captain. The gear log, as part of a series of log sheets designed to separate the recorded information into distinct sections, is linked to the other log sheet information by association based on the trip ID, vessel name and vessel number. An example of the gear log sheet is shown in Figure 7. The definitions associated with each of the fields on the gear log are listed in Table 6. The keypunched master gear log data file is a nearly direct transfer of the gear log. Data fields are shown in Table 14, and lookup tables are
described in Table 19 and printed in Appendix G.

**Trip log processing**

Trip information is recorded on a Trip Information Log for every trip (Table 7). The purpose of the Trip Information Log is to document vessel name, dates, fuel use, ice use, number of crew, and comments outlining any vessel or gear alterations from previous trips. As mentioned above, the trip log is derived from the trip components of the NMFS Vessel and Trip Logsheet, with the redundant vessel information excluded. This was done to reduce the fishermen's paperwork. While SMAST technicians are responsible for the vessel log data, the fishing vessel captain is responsible for the trip log data. An SMAST technician assigns a unique 3-digit number to every trip, prior to sailing. Trip number codes are consecutive within year. Trip numbers ranged from 1-119 (trip 101 excluded for lack of data) in survey year 1 (November 2000-October 2001), and 200-249 in year 2 (August 2002 - July 2003). A total of 167 trips were initiated, but catch data was not obtained from 3 trips. The trip logsheet is keypunched almost directly into the master trip log data file (Table 15). Port sailed was not keypunched because it was always New Bedford Harbor (except for two trips out of Hyannisport). The fields "Captain", "DAS" and "Comments" were added to the keypunched record (Table 15) by SMAST technicians. Days at sea (DAS) was calculated automatically by the Excel spreadsheet as the difference between the date landed and date sailed.

**Environmental log processing**

The environmental log was designed to record data on the ship’s operations during each haul, together with various environmental variables. An example of the environmental logsheet used at the start of the project is shown in Figure 9. It was intended that position and other data would be collected every 15 minutes during a trawl tow. This was immediately found to be impractical. Data was being recorded only at the start and end of the haul (i.e., the set and haul-back). As a compromise, it was decided that environmental data would be recorded at the set, haul-back, and at any point during the haul when the vessel made a turn. A new environmental logsheet was implemented to record this information efficiently (Figure 10; Table 8) after trip number 5. At the same time, additional fields were added to record weather information and bottom type.

During standard fishing practices, a typical trawl tow would include several events such as the set, a possible turn or turns, and haul-back of the net. A new line on the environmental log is filled out for every set, turn, or haul during a tow. Certain trawls may only have two lines, a beginning and an end that correspond to a set (first line) and a haul (second line), in column 1. Other hauls may have one or more turns. is completed. When duplicate information appears
from line to line, such as date, a quotation mark (") or an arrow pointing down
was deemed sufficient to indicate repeating information. At each of these steps,
date, time, location, boat direction, water depth, wave height, wind speed, and
direction are recorded. The bottom type data was later removed from the
keypunched data during quality control auditing as it was found to be unreliable
due to a poor choice of habitat types listed in the key on the logsheet and
inconsistent and widely varying interpretations of the meaning of these habitat
type labels by the fishermen.

Fishing operations are conducted in various weather and sea conditions
that can affect the ability of a vessel to catch fish. The purpose of the
environmental log is to record relevant information about the atmospheric and
sea-state conditions during fishing operations. The environmental log sheet, one
of a series of relational logs filled out by either a fisherman or an observer (if
present during operations), is linked with the other logs by the vessel name and
cruise dates. Environmental data information is usually recorded in the
wheelhouse of a vessel. The information contained in the environmental log
includes data pertaining to temporal and spatial atmospheric and oceanic
conditions, and the detail of fishing operations taking place within those
conditions. The data sheet contains sixteen columns. The headings encompass
when a haul begins and ends as well as if the boat turns during fishing operations
(see Figure 10). Data related to the vessel’s position includes: time, date,
position, haul number, course, and speed. Net number is a randomly assigned
number given to every net used by a vessel. Characteristics of each net are
recorded separately in the gear log. Examples of the environmental log sheet are
shown in Figures 9 & 10. Units, such as fathoms (1 fathom = 6 feet or 1.83
meters), on the data sheet are in common units used within the community of
fishing vessels. A comment section is included at the bottom of the
environmental log to document unusual occurrences or circumstance. This
comment was not keypunched and should not be confused with the technician
comment field in the keypunched data set that was added to the master data
during quality control auditing procedures (Table 16).

Water temperature data logger processing

To record water temperature, TidBit sensors (Onset, Inc., Pocasset, MA)
are used. A Tidbit temperature monitor is a programmable water resistant
instrument the diameter of a quarter, which records temperature at user-preset
time intervals. Since sampling began, the relatively inexpensive ($90) Tidbit
time-temperature data loggers (Figure 22) have been attached to the fishing gear
at the start of each trawler trip. Tidbit temperature loggers sample with an
uncertainty of about 0.2° C, for a period that is limited by the amount of memory
in the sensor, which is about 22 days of temperatures recorded at 30-second
intervals. The Tidbit recording interval is set by the SMAST technicians prior to
being delivered to the cooperating fishing vessel. The Tidbits provide bottom
temperatures while the nets are fishing, as well as vertical profiles when the gear
is set and hauled, and air temperature when the gear is on deck. These data,
along with the inventory of fish catch during the same haul, provide simultaneous
sampling of environmental temperature and fish catch. Since the fishing trips
usually range from 7-10 days, there is a 2- to 10-day delay between the actual
measurement and receipt of the observations.

During the first year of the project the Tidbits were secured to the nets, but
this was later changed to mounting a Tidbit on each trawl door. Attaching data
loggers to the nets involved many problems, ranging from loss and breakage to
the fishermen’s forgetting to attach the monitors to a net that is not used until the
middle or end of the trip. The fishermen’s habit of frequently changing nets also
complicated the problem of extracting temperature data corresponding to specific
trawl tows (see below and Appendix C). Since an analysis of the response time of
the Tidbit logger indicated that 30 seconds was shorter than the response time of
the logger, the time interval was changed at the beginning of the second year to
1-minute intervals.

At least two Tidbit temperature monitors are deployed on a vessel’s nets
or otter trawl doors during normal fishing operations when data is being collected
for SMAST for the purpose of creating a temperature time series of bottom
temperatures. The information on the temperature monitors is then offloaded as a
.dtft file using Boxcar 4.0 software (Onset, Inc., Pocasset, MA) and converted to a
text file. For trips where Tidbits were mounted on individual nets, it was
necessary to determine which Tidbits were used for each tow. SMAST
technicians developed a Matlab routine, tidbitselect.m, which combines the Tidbit
data with the environmental data recorded aboard the vessel. In addition to
finding the proper Tidbit, an accurate assessment of bottom temperature at the
start and end time of an individual haul must be ascertained. For this another
Matlab procedure, trip.tempselect.m, was utilized. Both Matlab procedures
require a technician to make decisions in determining the proper monitor, and the
start and end times of a given haul. Extracting the correct temperature data
corresponding to trawl tows is a difficult process because the trawl tow
deployment and retrieval times are not accurate enough, and are not synchronized
with the internal clock on the Tidbit. Often, temperature data alone is not
sufficient to determine when the trawl is in the water. The following paragraphs
and referenced figures briefly describe the process for temperature data
extraction.

The first Matlab program, tidbitselect, utilizes two windows to (1)
graphically display the temperature time series on the one, two, or three monitors
and (2) act as a mechanism to select only the monitor which was recording
bottom temperature. Currently, tidbitselect is run on a UNIX box through Matlab.
In order to run tidbitselect, the user needs to go through the following procedures:

Log on to the SGI UNIX system (any SGI terminal)
Enter “xhost pacific”
Enter “ssh pacific” or “telnet pacific” (the first is more secure)
Enter “cd /hosts/indian/mnt/disk2/home/user/flbub/MATLABX”
Enter “setenv DISPLAY your_terminal:0.0” where your_terminal is the
SGI computer at which you sit
Enter “matlab” and the proper “startup.m” will run and create your
required paths.

The “startup.m” program is named tidbitselect.m. After entering
“tidbitselect,” a second viewing window will automatically pop up. The program
then prompts the user to select a trip. All trips are numerically assigned 1-119 for
year 1, 200-249 for year 2, and 300-330 so far for year 3. Once a trip is entered,
the temperature profile for the Tidbit temperature monitors will pop up in the
view window on a haul-by-haul increment. There can be as few as one
temperature profile and as many as three for any given trip. If none of the profiles
accurately reflect what a bottom temperature should be (usually somewhere
between 3 and 13 deg C, depending on depth and time of year) on an individual
haul then a fourth choice, “no data” is chosen. At the end of a trip, the Matlab
prompt pops up again in the data window and the process starts again. The output
of cumulative chosen temperatures is stored in:
/hosts/indian/mnt/disk2/home/user/flbub/MATLABX as "Atidbit###.txt."

The determination of which Tidbit is the “correct” one for a given tow is
guided by several indicators. In most cases there will be a spike at the beginning
and the end of a haul. In addition, climatology suggests what the temperature
should be at a given depth and time of year relative air temperature. Some cases
are generally easy to determine; water temperatures at depths of about 85 fathoms
and greater are relatively stable and easy to determine. On the other hand, in
shallow water (less than 50 fathoms) in the fall and late spring, air temperatures
are similar to water temperatures, and the water column is well mixed making it
difficult to extract trawl temperatures from the record (Figure 23).

The file Atidbit###.txt contains a temperature time series based on the trip
hauls as recorded in the corresponding environmental log. The Tidbit record and
environmental log differ somewhat in almost every case. The reasons for this
difference include the fact that different clocks are used to set the Tidbit logger
(SMAST technician’s PC clock) and to record the trawl times (fishermen’s
watches), the time the gear reaches the bottom may differ from the time the
captain records for the trawl set, the Tidbits take a few minutes to equalize to
temperature at depth, among many other possibilities.

In order to determine what the actual bottom temperature is for the
beginning and end of a haul, the Matlab routine triptempselect.m is employed
(Fig. 24). In this Matlab program, the first prompt asks for a randomly assigned
number on the x axis where the haul begins fishing on the sea floor. The program
then asks for another of these numbers on the x axis that corresponds to the point
where the Tidbit left the sea floor, signifying the end of a given haul. Some hauls
contain peaks which represent turns made during fishing operations. In many
cases when a turn is made, the otter trawl doors and net are lifted from the sea floor to reduce the chances of damage to the gear. The program then prompts the user to verify if they want their requests processed. If the answer is yes, the program moves to the next haul. If the answer is no, the user is prompted to repeat that haul. At the end of the trip file the file Btidbit.txt is created.

A summary of the steps involved in Tidbit temperature processing follows:
Step 1. Collect 1-3 Tidbit records per trip.
Step 2. Select the best Tidbit record for each haul in each trip, output trip temperature data file.
Step 3. Determine the start and end time for each haul in the temperature data. Mark beginning and end of any observed tows. Output as haul_temperature data (step a, Fig. 21).
Step 4. Determine the haul bottom temperature profiles by excluding data during ascent, descent and any turns. Also remove any obvious data outliers. Output as haul_bottom_temperature profile data (step b, Fig. 21).
Step 5. Determine temperature statistics for each haul including mean temperature. Input into the environmental (station) data record (step c, Fig. 21).

Haul log processing

The haul log documents the catch for each tow during a trip and is the most critical component of the data collected. The haul log includes a space for listing the vessel name and trip number (Figs. 10-16). The date field refers to the date when the haul commences. The hauls are numbered consecutively beginning with one within each trip and recorded in the haul number field. The target species field is used to record the species of fish that is sought for each tow. If more than one species are being sought in a specific tow then the target species is designated "M," for multispecies.

In this first log, fishermen were asked to write in the names of fish species captured in the haul. They also were to record the weight of each species that was discarded and kept.

Because of the relatively high complexity of the haul log compared to the other data logs, and the need to continually re-evaluate the effectiveness of the log when used by fishermen, the haul log has undergone numerous revisions during the course of the SMAST trawl survey. Examples of all variations of the haul log used during this study are provided in Appendix E.2., and a detailed summary of the major changes in the haul log is provided in Appendix F. The initial haul log was based on the NMFS Observer Haul Log (Fig. 11). The first version of the haul log contained many fields that were redundant with the environmental log or that were not applicable to the SMAST survey. This initial log sheet format proved less than ideal as it required more writing by the fishermen and was overly
complicated. Four variations of this logsheet were implemented in rapid succession (Figures 12-15). Most of the revisions involved the layout of the form, removal of unused fields derived from the NMFS Observer Haul Log, and the inclusion of various codes and keys to assist the recorder (e.g., discard reason codes, explanation of the d/r and a/e fields, etc.).

Two problems with the haul log format became apparent during the first year of the study. First, writing out the names was time-consuming and resulted in much confusion due to spelling and common name variations. Second, the format did not take into account the way that fishermen process the samples. Kept fish and discarded fish were processed completely separately, and often by different crew members. Discard was subject to neglect, or was often recorded on a separate line from the kept data. The sixth and seventh versions of the haul log (Fig. 16 and 17) sought to address these problems. First, it included the most common species of kept fishes and discarded fishes as pre-printed lines on the form. Second, it separated kept species from those that were typically discarded. For example, since there is no market for things like longhorn sculpin and sea raven, it was much more efficient for the fishermen to record discard data in a separate section. The new log included common names for twenty-five of the most common species with market value and nineteen of the most common bycatch species. The common names that fishermen use are the ones listed on the data sheet. These common names were chosen because fishermen, and not scientists, need to be comfortable with the data sheet they are filling in. In addition, the species were grouped into species groupings that were familiar to the fishermen, such as flounders.

Another major change to the logsheet was that discard weights were to be recorded in pre-printed weight bin categories in an effort to improve the accuracy of discard estimates. Because discarded species are typically not sorted and stacked into baskets, discard estimates are based on eyed estimates. In contrast, kept fish are meticulously stacked into fish baskets or fish boxes, where weight estimates can be made more efficiently based on the fishermen’s experience in estimating weights for his livelihood. The weight bins were selected to take advantage of the influence of weight scale on weight estimation. For example, although it was felt that most fishermen could accurately distinguish fish weights less than 5 lbs from those between 5 and 25 lbs, the same fishermen might not be able to accurately distinguish between 105 lbs and 125 lbs. Fishermen were instructed to write in the weight of discards estimated to be over 100 lbs. The maximum range in a given bin was keypunched (i.e., in the 1-5 lb bin, 5 was keypunched).

The seventh and current version of the log sheet, which was implemented in August 2003 (year 3 of the project), included changes to clarify special situations that arise for a few species that often have more than one dressed type in a given haul and/or frequently have multiple discard reasons within a given haul (Fig. 17). For example, weights for monkfish can be estimated from whole fish, from fish dressed so that the head is intact but the viscera are removed, or
dressed so that only the tail is retained. In the early data, it was sometimes ambiguous as to which situation applied to the data in a given haul.

On deck, the catch content is logged by species and estimated weight distribution. Estimating the catch is the most subjective aspect of the project. The bycatch is by far the most difficult to accurately assess. This aspect of training is the most difficult in almost all cases and for two main reasons. A lot of fishermen view bycatch as simply stuff that takes time to dispose of and sometimes damages equipment. In many cases certain fish that have been seen thousands of time by fishermen remain nameless. Some fishermen do not understand the importance of monitoring bycatch. They don't fully appreciate the connection that our goals are scientific and not economic. They only see value in what can produce income. Getting fishermen to learn the names of fish that are disposed of has been a challenge. Another challenge is assessing exactly how much bycatch comes aboard in a given tow. The bycatch is not sorted into species-unique containers as are kept species. It is swept overboard. For of these reasons, the discard data is the most poorly quantified component of the catch data. In addition, it became apparent during the first year that some vessels were not providing reliable discard data (see Table 30). In these extreme cases all discard data, and hence all total catch data, were set to missing values for the entire trips during creation of the trawl working data sets (see below). This was done in a programing step. The erroneous discard data is still contained within the master data files. It is important to point out that it was not possible to a posteriori determine when discard data was unreliable on a haul-by-haul and species-by-species basis. Therefore, we recommend that all discard and total catch data from this study be used with caution.

When a haul is completed, the net is brought aboard and the fish are dumped on the deck. A normal haul will fit into a large partitioned section of deck known as a checker. The resulting "pile" is then sorted into smaller checkers for washing and handling before being transferred into the fish hold below deck for storage until the end of the trip. When the pile is dumped on the deck, a quick visual inspection of it will give the observer a good idea of most species in that haul. While sorting the pile, the main focus of attention should be on discarded species. This is the only chance to ascertain how much and what kinds of fish are discarded. Most fish are discarded for reasons ranging from market desirability to regulatory restrictions. Once the pile is sorted, the kept species will be present in their respective checkers. Some flatfish may be mixed, as may be some gadiformes and similarly shaped fishes. The fish in all likelihood will then be sorted into baskets that will be lowered into the fish hold. When the catch is put into baskets, most all of the fish will be separated. This is the best time to ascertain kept weights. At the end of the sorting process, the observer or a crew member documents what was in the haul.

The review of the haul log data and the quality control auditing of the data were the most difficult and time consuming component of the data processing. In many cases, technicians had to manually compare species identification, and more
specifically species dressed type information, between what was recorded on the logsheet and the data files. Because of confusion over the various dressed types and multiple species codes used for monkfish, most of the monkfish data from the first year of the survey had to be checked by inspecting the logs. Catch data partitioned by dressed type, kept or discard status, and discard reason had to be cross-referenced and checked to see that they could be combined to arrive at total catches. For example, it was possible to have whole monkfish, both kept and discarded, and monkfish tails in the same haul. In many situations a record coded for monkfish tails included both discard data and kept data. This resulted from fishermen recording the weight of whole fish discarded together on the same record as the weight of monk tails kept. In some cases the species could be coded as monk tails (species code 120) or whole monk (species code 124). In other cases, data for whole monkfish and monkfish tails would be recorded on separate lines of the logsheet. The resulting multiple keypunched records had to be combined appropriately by summing them after the dressed weight conversion factors to create the final working data sets.

The species names themselves had to be carefully reviewed and checked for authenticity. Thirteen species or species groups were accepted as valid target species (Table 23). Fifty species categories (Table 25) resolved into 49 species or species groups recorded in the catch (Table 27). A larger number of species were reported in the data, but some of these were edited during quality control auditing to a higher taxa (Table 31). These changes were made because it was felt that fishermen were not accurately identifying some fishes to species. For example, although both red and white hake appeared in the data, these were consolidated into a red and white hake mix because fishermen do not make a distinction between these two species. Although they could be trained to do so, it is not economically feasible for the fishermen to separate these species in the catch because it would be a time-consuming effort, and the market does not distinguish between the species.

As part of the quality control processing, it was necessary to develop our own table providing dressed type descriptions and the appropriate conversion factors needed to determine the whole body weight (Table 29). To do this we used the Conversion factors developed by the NMFS Observer Program together with our own dressed type descriptions. This data is compatible with the NMFS observer data in all but a few cases unique to the SMAST study. During the course of this process, it became apparent that the conversion factors used by the NMFS Observer Program should be studied in some detail. We were not able to find any supporting data to validate the conversion factors, and no information on estimate error is available. This would seem to be a minor point at first look, but on reflection it seems that the accuracy of the conversion factors could have a significant impact on stock assessment. A small error in the conversion factor would translate into a relatively large error in biomass estimates. In addition, a correction factor appears to be applied to some species regardless of their dressed state. For example, all cusk are given a 1.13 conversion factor whether dressed or
whole. The conversion factor is important to the SMAST data as it is used to determine the whole body weight of the kept catch for each species. If changes are made to conversion factors in the future, the SMAST data will have to be regenerated from the master raw data using the new values. Unfortunately, because dressed weights must be corrected, and then added to undressed weights within each trawl tow and species to obtain the final species kept weight, it is not possible to work backwards from the final “catch” files in the primary user data set to apply different conversion factors.

Finally, during data processing it was necessary to insert a field into the data (total_pound_discard) that represents the sum of the weights discarded for different reasons (Table 17). That made it possible to break out the discard-by-reason data from the main catch table (see “Overview of Working Data Sets” below).

**Length-frequency log processing**

The length-frequency log records the lengths of fish measured from a given haul. When a length frequency is taken, one of five predetermined species is chosen by the recorder. The five species of interest are: cod, haddock, yellowtail flounder, greysole flounder, and black backs (winter flounder). Thirty individuals of the chosen species are randomly removed from the pile of fish on the deck to be measured. The randomly chosen fish are measured on a measuring board by placing the nose of the fish on the beginning of the board and locating the total length (TL) of the fish. The length is recorded in centimeter bins in pencil with hash marks on inserts provided with the measuring boards (Figure 25). Once all measurements are made, the recorder writes the species and haul number directly on the insert. By recording the haul and species information, the recorder can finish the processing at a later time if necessary.

The template from the measuring board is removable and is brought into the boat to be transcribed onto the length frequency data sheet. Three variations of the length frequency logsheet have been used to date (Figs. 18, 19, & 20). The first version (Fig. 18) was based on the NMFS observer form and included a number of fields not used in the SMAST program. The second version (Fig. 19) greatly simplified the log, and removed the unutilized fields. The last version (Fig. 20) included a minor change in the log layout.

The length frequency log data includes the date, trip, and haul number. The species that was measured is circled from the selection of species. Any comments relating to the measuring can be entered in the comments section. Lastly, the measurements from the template are transcribed into their appropriate boxes. The log has slots for fish lengths ranging from ten to one hundred and twenty-nine centimeters. If a fish exceeds 129 cm its length is manually entered in the comments section. The aim of the project was to measure the length-frequencies on roughly half of all hauls during a trip. Unfortunately, because of the time involved in collecting length frequency data, this was perhaps
the most difficult data to get fishermen to collect. Continued encouragement by
SMAST technicians resulted in a greater commitment to record length frequency
data by the fishermen during the second year of the project.

Overview of Working Data Sets

Once the keypunched data have been fully audited and have passed quality
control testing, they are converted into a set of working data files in a relational
format (Figure 26) for use by SMAST scientists and for distribution to the public.
Working data sets are provided in both Microsoft Excel and Access formats and
contain extensive metadata documentation of all fields. An overview of the steps
to construct these files was presented above and is illustrated in Figure 21. The
working data files are organized into four folders: 1) primary data sets, 2) derived
catch data, 3) field descriptions and 4) lookup tables (see Tables 32 & 33). The
primary files consist of the Vessel (Table 34), Gear (Table 35), Trip (Table 36),
Station (Table 37), Haul_event (Table 38), Catch (Table 39), Discard_by_reason
(Table 40) and Length_frequency (Table 41) data tables. Secondary files, called
“derived catch files,” are also created for the convenience of researchers, but are
not essential as they simply contain information derived from the catch (Table
42). The derived catch files contain a single record for every unique trip and haul
(i.e., for each trawl tow) with separate fields for every species collected during the
entire study. File structures and field definitions for all the working data files are
provided in Tables 34-42. Lookup tables for all coded fields contained within
these data sets are described in Table 33 and printed in Appendix H.

Currently, the working data files are created from the master data files
using a set of three Statistical Analysis System (SAS) programs (Appendix I). The
SAS programs output the working data files in excel spreadsheets that can then be
loaded into Access. In addition, a great deal of metadata is provided to document
the files, including tables of all file structures and field definitions. Details of the
steps needed to accomplish this are documented in the SAS programs.
Researchers interested in recreating working data sets from the master data are
encouraged to carefully review documentation provided in these SAS programs to
avoid misuse of the data. The most important SAS programs is the first one titled
“trawl_master_SAS_data_set_creation_10sep04.sas”. Note that the SAS program
strips out extensive auditing control fields that have been added to the master data
files during data processing (e.g., old_fields, edit_fields, and comment fields).

The relationships between the field data logsheets and the final derived
data tables are shown in Figure 21, while Figure 26 summarizes the relationships
among the working data sets. Master data sets from the vessel, trip, gear and
length frequency logs are converted directly into vessel, trip, gear, and length
frequency working data tables with little structural change, although net name and
net_category are added to the gear table for convenience. However, master data
from Tidbit temperature records, environmental log and haul log data are each
split into multiple working data tables. The continuous temperature recordings
stored on 1 to 3 Tidbit recorders for each trip are consolidated to remove periods when the gear was not deployed (i.e., when the net is sitting on the deck or is in storage) to produce a data set with continuous data for each haul (Fig. 21, step a). The haul_temperature data table thus provides a record of the temperature experienced by the gear during the trawl tow from approximately the time of set to haul-back. However, this includes periods when the net is lifted off the bottom during vessel turns, and thus includes data other than the bottom temperature data. In step b (Fig. 21), intervals when the trawl is off the bottom during turns are removed so that the resulting haul_bottom_temperature data table includes a time series of bottom temperatures observed during the trawl tow. In step c (Fig. 21), mean bottom temperature and other statistics (variance, range, etc.) are calculated and inserted into the station data table. Unfortunately, at this writing the creation of the master and working temperature data has not been completed. Earlier versions of the data were discovered to have been significantly corrupted, necessitating the time-consuming re-extraction of the temperature data from the original Tidbit data records. The temperature data will be incorporated into the trawl data base as described above as soon as processing is completed.

The environmental log is converted into the haul_event table with minor changes and provides data on the environmental conditions, and ship performance during each set, turn and haul-back event during the trawl tow. In step d (Fig. 21), the environmental data are aggregated over haul events within the haul_event table to provide statistics on the tow and to create the station data table. These include conditions at the set and haul-back, as well as the mean, minimum, maximum, and variance of conditions across all tow events.

The haul log is spit into multiple data sets after extensive manipulations in the SAS programs. First, data partitioned by discard reason within a species are separated out into a separate discard_by_reason table. This includes the weight of a given species that is discarded for each different reason. The total amount discarded for a species is summed up over all discard reasons and inserted into the catch data table. The catch data table includes data on the total weight of the catch for a given species (total discard plus total kept), the total discard (summed over all discard reasons from the Discard_by_reason table), and the total weight kept. The total weight kept must be computed by calculating the total weight of each dressed type and the summing over all dressed types. In step e (Figure 21), catch data are aggregated over all species and inserted into the station table to provide data on the total catch (total weight of all species, total weight discard, number of species, etc.). Finally, several derived catch files are created from the catch table. Each derived catch file includes only one catch estimator statistic determined for all species encountered during the entire project for each trawl tow. For example, the total_catch table contains the total catch of all 50 species (49 species plus valid “water tows”) for all trawl tows; hence many species have zero weights for any given tow, but all species have at least one positive weight for at least one tow.
CONCLUDING REMARKS AND RECOMMENDATIONS

The SMAST High-Resolution Industry-Based Trawl Survey was a highly successful demonstration of a program to train fishermen to collect environmental and biological data. The resulting data generated by this project is expected to provide significant information to regional managers by characterizing the Georges Bank fishery trawling operations and catch. Summaries of the data generated by this project will be reviewed in subsequent reports. However, information contained in this report provides significant insight on how to better design future industry-based trawl fishery surveys. Probably the most difficult component of the fishery data to quantify with an industry based survey is the catch discard data. For many logistical reasons described above, fishermen have only limited opportunity and incentive to collect discard data during normal fishing operations. The collection of discard data is in direct conflict with the fishermen’s need to efficiently process the kept catch and to quickly return discard species to the sea in order to clear space for subsequent hauls and reduce mortality. Nevertheless, we believe it is still possible to obtain useful discard data from industry-based surveys.

One problematic aspect of the discard data encountered during this study was the uncertainty as to whether the lack of discard data represents a true lack of discard or a lack of reporting on a haul-by-haul and species-by-species basis. This problem was partially resolved in year 2 by modifying the haul log sheet to improve discard reporting (pre-printing kept and discard species names on the form, and including discard weight estimation bins to encourage reporting). Thus discard data reporting significantly improved in year 2. However, it was still not possible to be certain that all occurrences of zero discard were valid. Sometimes fishermen may skip recording of discard data for a particular haul because of conditions at the time (fatigue, weather, lack of time, etc.). Inclusion of a field on the log sheet to allow fishermen to indicate that discard data was not reported, together with training to educate the fishermen on the importance of being able to distinguish true zero discard from non-reported discard, should help reduce this problem. Beyond that, there is a need for researchers to develop efficient and accurate methodologies that allow fishermen to quickly estimate, or measure catch discard without undue disruption of their processing of the catch.

Our industry-based study would also have benefitted from subsampling by scientists to estimate the error of the fishermen’s estimate of the discard and kept catch weights by species. The chief difficulty carrying out such sampling is the difficulty of placing scientists on board vessels during normal fishing operations, and the greater difficulty of persuading the fishermen to allow the scientist to disrupt their catch processing in order to collect the needed data. It must be recognized that having a scientist on board collecting the appropriate data for error estimation (letting fishermen record kept and discarded catch estimates, and then coming behind them to measure the actual weights of the kept and discarded catch) does represent a significant burden on the fishermen and would
significantly reduce their efficiency. Obviously, the problem is greater for
discard data compared to catch data because of the necessity of returning the
discard to the water as soon as possible.

Data describing the reason(s) for discard were not reported as often as we
would have liked, partly because the NMFS reason categories are extensive and
somewhat confusing, especially when decisions have to be made quickly. In
addition, it is cumbersome and confusing to record such information on the log
sheets as it adds another layer onto the data structure. Changes to the haul log
that were implemented at the end of year 2 (and throughout year 3) improved
reporting of discard reasons, but allow for multiple discard reasons to be recorded
only for lobster. Recording multiple discard reasons for other fishes basically
requires hand-writing in additional lines of data by filling in one or more of the
blank lines included on the log (but usually that means data from the same species
get recorded in two different places).

Another improvement to the study would be to collect gear information on
a tow-by-tow basis rather than on a trip basis. This is because fishermen
frequently modify the nets and can change things like ground cable length and
ground gear between hauls to take advantage of changing conditions and catches.
Further, as described in the gear and gear performance sections of this report,
more research on the effect of these modifications on catchability by species and
size class is needed. In particular, the effect of changing the ground cable length
and ground gear composition on fish behavior and consequently on catchability is
needed.

During the course of this study we came to believe in the need for better
data on the influence of various methods of dressing and preparing fish on the
conversion codes used by NMFS to calculate whole body fish weights. We were
unable to obtain information on how the conversion factors were determined by
NMFS. No information on the sample sizes or variance of the conversion factor
estimates appears to be available. We suspect that a wider variety of dressed
types occurs, with potentially different conversion factors. Further, NMFS has
assigned conversion factors greater than 1 to the whole body weights of several of
the species collected during this study, including cusk, Atlantic halibut, Atlantic
wolffish and scallop (Table 29). Since by law, scallops must be shucked at sea, it
is reasonable to assume that a conversion factor should always be applied if data
are always collected in port (but then, why have a whole body code category?).
At least in the case of our study, weights could be estimated either from whole
fish prior to dressing or from dressed fish immediately after processing the catch
at sea. Therefore, no conversion should be applied when whole body codes are
used.

Finally, the recording of bottom temperatures using the Tidbit data loggers
was much more problematic than anticipated. Normally, on a scientific survey,
the data loggers would be placed on the nets or trawl doors at the time of set and
removed for data download immediately after the haul-back. In addition, clocks
used by the recorder would be better synchronized with the internal clock in the
temperature recorder. In contract, in the industry-based survey, the Tidbits are supplied to the fishermen already recording prior to the trip departure and are not turned off and downloaded until well after the trip returns to port. On top of that, a fishermen’s concept of accurate time keeping is not the same as that of a scientist, and times recorded on the log sheets can be 30 minutes or more off that recorded by the temperature data logger. These factors make it surprisingly difficult to extract the bottom temperature data associated with each trawl tow as described above. This problem can be virtually eliminated by using data loggers that record both temperature and depth, though these are more expensive.

ACKNOWLEDGMENTS

The School for Marine Science and Technology at the University of Massachusetts Dartmouth would like to acknowledge the hard work and dedication of many of the commercial fishermen who worked so assiduously to gather fisheries data for this project under the challenging conditions that is life at sea. The success of this study would not have been possible without their efforts. Our thanks and appreciation is also extended to members of the Trawlers Survival Fund, New Bedford’s local fishermen’s association, and especially to the TSF’s Director Bob Lane. The advice and perspectives he provided on all aspects of the program were invaluable. We would also like to thank Dr. Paul Diodati and other scientists at the Massachusetts Division of Marine Fisheries for their support. Frank Smith provided editorial assistance. This project was funded by the National Aeronautics and Space Administration under grant number NAG 5-9752, NAG 13-02042 and NAG 13-03021. Additional funding was provided by a contract to the Center for Marine Science and Technology (now SMAST), University of Massachusetts - Dartmouth, from the Northeast Region, National Marine Fisheries Service, NOAA, DOC, under the Cooperative Research Partners Initiative (Contract No. 50-EANF-0-00062).