Environmental Monitors on Lobster Traps

Phase III Database Management

(Subcontract 03-659)

Final Report to Northeast Consortium

Abstract

A program such as eMOLT with over a hundred individuals contributing to the database requires efficient and reliable management strategies. The process has evolved from the first year of eMOLT. Routine standards and program conventions have been modified and are now stabilizing. The most difficult aspect of the project is not the collection of physical data but documenting data (i.e. mooring logs, metadata) associated with each deployment. From the beginning, we have implemented a number of options to record this information including the use of the Thistle Marine Electronic Logbooks. The difficulty in documenting data has, however, been reduced over the years as fishermen come to understand the notion of "fixed sites". In other words, after the position and depth of a particular site are accurately recorded once, fishermen simply redeploy the instrumentation at those sites.

Much of the burden in documenting deployments now falls on the lobster "association representatives". While we attempted to disperse some of this task to a set of "industry representatives" during phase III, we found that the additional layer of administration only added confusion. In phase IV, it is now the association representative's task to collect and transfer the handwritten logs to electronic spreadsheet files in specific formats. Much of the eMOLT funding goes to their time on this effort as well as the equipment needed to do so (laptops, printers, etc.). The difficulty of converting between units used by most New England fishermen (farenheight, fathoms, Loran TDs, local date/time) to those expected by the scientific community (celsius, meters, lat/lon, Greenwich Mean Time/yeardays) and then back is not insignificant.

While the scientific results of the eMOLT project in general can be obtained in various documents at http://emolt.org, the details of our data management experiences are presented here. Much of the report is contained in the "Methods" section as it describes both the procedure and the nuances associated with each step of database management.

Deleted:

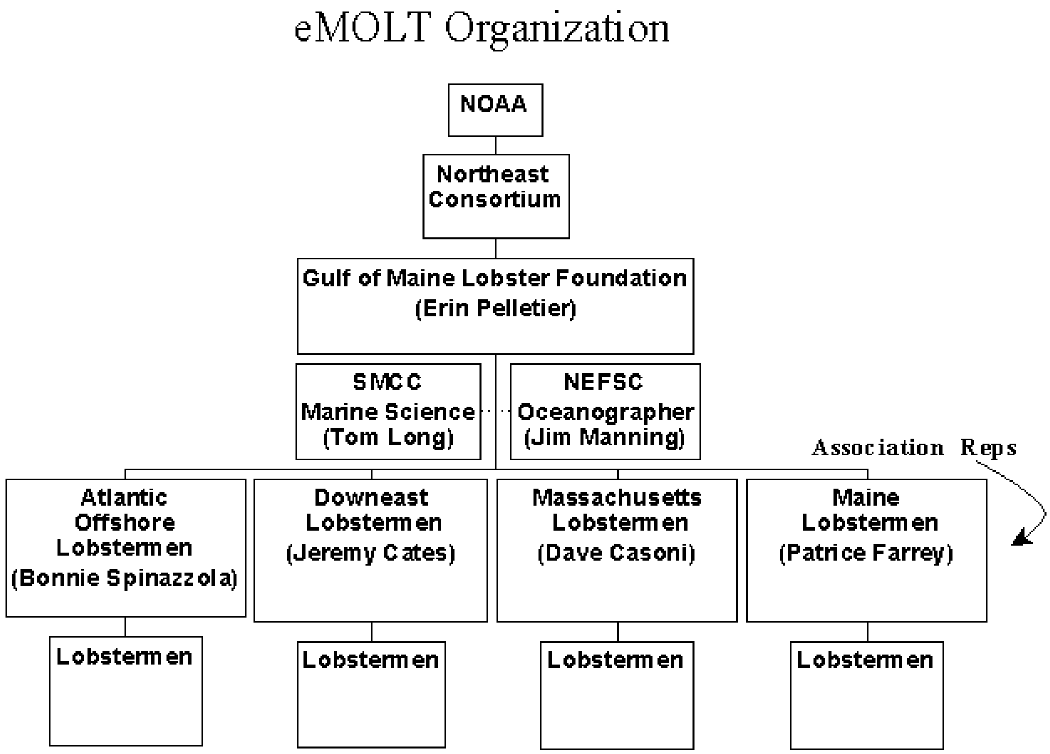
Deleted:

Background

In today's world of electronic devices capable of easily acquiring megabytes of data, it is important that the proper system be setup to process, archive, and serve relevant information in a an efficient and reliable manner. In all disciplines of science, the proper dissemination of well-documented data now requires more effort and expense than the collection of raw data. With the success of the temperature probe project in the earliest phase of eMOLT, it became immediately obvious that a serious effort to document data was necessary. Phase III was specifically designed to address this issue. It was necessary to build a standard operating procedure for every aspect of the project and to make it consistent for all participants and all associations. If the eMOLT concept is to be continued for years to come, it is essential that decades from now the pertinent information on data collection now be accessible to all researchers. For example, when depth is recorded by lobstermen, is it feet, fathoms, or meters and is that estimate of water column depth at mean low water or whatever the fathometer displays at the time? Are latitude and longitude in degrees, minutes, seconds or degrees, minutes, decimal minutes? These are just some of the subtle but important aspects of documenting data from moored instrumentation.

The need to standardize oceanographic data collected on moorings has recently caught national attention by the OCEAN.US data committee. A workshop entitled "Aggregation Data Model Structures for Time Series from Ocean Moorings" was held in January 2004 at NOAA's Coastal Service Center in Charleston, SC where both eMOLT and NEC were represented by Jim Manning and Bob Groman, respectively. The importance of metadata, data integrity, and QA/QC flags was discussed. The proceedings will help guide future protocols for integrated ocean observing system development.

The eMOLT project has worked due primarily to the organization of fishermen. Nearly all of the eMOLT participants are active members of "associations". These organizations and the individuals who participate care about their industry and genuinely interested in cooperative research with scientist. Most understand that it is unlikely that they will personally benefit but are willing to contribute to the long-term understanding of the ocean environment. As denoted in Figure 1, there are four major associations, each one with an eMOLT representative.



Methods

One of the first concepts communicated to all participants was "site codes". Since the eMOLT project is predicated on returning to "fixed sites", it was necessary to designate each of these standard locations/depths with a site code. In the beginning, since we were working with some of the larger offshore operations, we were naming sites according to the fishing vessel. The sites occupied by Colbert brother's various captains, for example, were "MJ01", "MJ02", etc. for the F/V MISS JULIE. After several sites were named in this way and more inshore participants joined the program, it was decided to name sites according to the individual's initials. The first site occupied by "John Chipman", for example, is "JC01". As the number of participants grew, it sometimes became necessary to switch the order of initials (i.e. Jon Carter's site is called "CJ01"). If switching initials does not work, the initials of the younger generation or a spouse's initials are then used. Jeremy Cates' site is called "CC01", for his wife, Charlene. In order to designate a new site, an administrator can view the current list of 160+ sites under the emolt.org "Data Access" link. By clicking on a particular site, the individual associated with that site and the approximate depth is listed.

These fixed sites are documented by lobstermen. It is critical to ensure that probes are deployed at the location and depth specified. While a site is valid with as much as a half mile radius of the nominal location, the probe must fall within 5% of the nominal depth to be considered the same site. It is not always easy to comply with this convention especially for fishermen who fish over steep topography and energetic tides. However, we are working to help them understand that, in order to properly interpret variations in temperature from one haul to the next, it is necessary to eliminate the depth-dependent temperature effect. The error associated with probes moving through multiple depth zones will need to be assessed on a site-by-site basis (see Discussion below).

The exact protocol for documenting sites has been adjusted as the program evolved. Four basic options have been available for all participants from the beginning.

The first option is the Thistle Marine Electronic Logbook. These units were successful to some extent for the offshore fishermen but not enough to be worth the fishermen's effort. For most of the inshore Massachusetts lobstermen, these units are too complicated to install, difficult to understand, and a challenge to use regularly while fishing. Nevertheless, the eMOLT program coordinated with Thistle Marine such that access to trawl data associated with the temperature probe is available. After downloading the raw data, a series of MATLAB routines are conducted to extract, plot, and archive the necessary information. Tracks of the individual's activity, assuming they implemented the "trawl track option", can be plotted, for example, to determine their GPS position for the entire time the string is hauled. These plots verify the approximate position of the probe from week to week and the variability of the fix with respect to the nominal position. Being sensitive information, none of these track plots are made public but are shared with the individual responsible deploying the probe. Since catch data is also available from the data stream, plots of temperature vs. "relative catch" can be generated. At the time of this writing, there are only a few inshore participants still using the "Thistle Box" to document the location of a temperature probe. The individual who sold the units has reduced his support significantly and, in fact, only maintains the business on the side. The use of these units has declined not only within our project but for many studies around New England. The states of Maine, Rhode Island, and Connecticut also have dozens of these units in storage.

The second option is to enter the information directly into the database through an on-line web page at: http://www.wh.whoi.edu/~jmanning/lob/prompt_info_set.cgi. While this is probably the most efficient method of data entry since the results would go directly into the master database, very few lobstermen, if any, have implemented the option.

Table 1. Example Spreadsheet for Participants Using Option 3

SN	Deployment	Site	Latitude	Longitude	date-haul	time-haul	Depth	#of pots/haul	#lbs kept	#lbs. eggers	#lbs shorts	trap type
566	01	TA01	4059.8	6733.2	4/10/2001	00:01am	100	50	100	20	30	50"wire
566	02	TA01	4059.8	6733.4	4/25/2001	12:30am	102	50	50	40	20	
344	01	TA02	4100.1	6822.0	4/12/2001	10:23am	80	50	123	15	40	

The third option is to maintain a spreadsheet record with each haul (see Table 1 above). As described in the emolt.org "Getting Started Manual for Participants", example template spreadsheets are available for download with the specific columns labeled and documented. The order and format of data entry is described. The units of latitude/longitude, for example, are requested in degrees, minutes, decimal minutes, depths are requested in fathoms, and columns of date & time are formatted accordingly.

The fourth and final option for fishermen is, of course, to fill out a handwritten log. This is the most commonly used option. As noted above, it is the association reps responsibility to then collect these waterproof paper logs and transcribe them into electronic files. The blank handwritten logs have gone through a number of iterations as each association attempted to simplify the requirements of their respective fishermen but the standard form is posted prominently as the first click under "Getting Started". Both one-page logsheets/instructions as well as a multi-page detailed manual about all of the above options are posted. Waterproof hardcopies of the former are distributed to participants either through the mail or at meetings by the association representatives.

In options three and four above, there is often the difficulty in specifying position in lorans vs lat/lon. We have accepted lorans in these first few years of eMOLT but hope to gradually migrate in the next decade to defining sites by their lat/lon. For converting the lorans to lat/lon, we have setup a web site to be used especially by association representatives while generating merged electronic spreadsheets at: http://www.nefsc.noaa.gov/~jmanning/lob/loranConv1.cgi . Since nearly all lobstermen have GPS positioning, it is only a matter of setting their navigational panel in the correct mode. In late 2003, we distributed a set of wallet-sized laminated "site code cards" to all inshore participants with their lat/lons listed on one side and loran on the other. These are provided so they are familiar with their standard eMOLT sites.

In option three and four above, there is also the difficulty in specifying the correct format of lat/lon. While the eMOLT convention is to use degree, minutes, and decimal minutes (DDMM.M) some lobstermen set their navigational display to degrees, minutes, and seconds (DDMMSS). Confusing the two formats can often lead to a half-mile or so error in position. A more serious problem is when position is recorded in decimal degrees (DD.DD) especially if the decimal points are out of place.

Additional problems arise when the depth is recorded without units. Is it feet, fathoms, or meters? Fortunately, part of the routine processing in defining a new eMOLT site is a MATLAB depth-checking routine which, given a specified lat/lon, estimates the depth based on interpolation of a 15 second-resolution bathymetry database. If the discrepancy between fishermen-reported depth and the empirically-estimated value is more than a typical tidal range, a flag is raised. As of this writing, we have made no attempt to specify what tide the nominal depths should be specified. In an attempt to keep requirements as simple, we have not required this detail. In the near future, we hope to make model estimates of tidal range for each site in order to define the envelope of variation expected at each location and gradually begin to quantify the exact "depth" specified at each site.

Problems arise with date and time. Fishermen deal in month, day, and local time. Scientist deal in yeardays and Greenwich Mean Time. The temperature probe records have a time stamp on each observation that is a function of the PC clock at the time the probe was initialized. Since most of the probes are initialized in early winter, we have chosen Eastern Standard Time as the eMOLT convention. An important part of the administrators protocol, therefore, is to ensure that the PC clock is set to EST when initializing a probe. Data users should be aware, researchers especially, of the EST convention. The error of a few hours may not mean much to a fishermen or a lobster but to a physical oceanographer looking to resolve tidal variations and coherences between sites, it is a nightmare.

In all of the above options, there is space to enter haul information. While we specifically requested this information at the start of eMOLT, we have since backed off this requirement. We are now focused on getting good position and depth information. Lobstermen may fill in the haul information if they desire, and plots of catch overlaid on temperature can then be produced. The catch is plotted in the form of "relative catch" such that the range of catch values span the range of temperature. Plots can be shared among fishermen without revealing sensitive catch information. Less than a quarter of the lobstermen are currently providing haul counts.

Laptops were distributed to all the major associations. Association reps can carry these units to various meetings and annual forums in order to a) enter participant documentation, b) download data, c) train participants to download their own data in the future, and c) display results. Given the LCD projectors purchased with phase IV funding, presentations can now be made to all attendees at various meetings. Aside from occasional news publications, this will become our primary outreach mechanism. It is assumed that all association reps have become proficient in software packages ONSET BoxCar, ExCel, and PowerPoint. In the near future, as we migrate to VEMCO probes, association reps will need to become familiar with another BoxCar-like package called "Minilog".

The temperature and salinity data is passed through a series of processing routines, plotted on the web, archived in web-accessible ORACLE tables, and served via OPEN Data Access Protocol. Since the salinity process is outlined in the final report of Phase II at http://www.nefsc.noaa.gov/~jmanning/lob/saltfinal.html, the process associated with managing temperature is outlined here. After a “calibration check”, association reps download data and the science party processes the data. While the association reps do not always conduct the download, they have done so at times.

- *Calibration checks:*
    - Whenever a set of probes (~6 or more) are collected for a download session, they are first subjected to an icebath calibration check. The exact protocol for this routine has been adjusted and refined with each attempt according to the manufacturers suggestions. The ONSET corporation has posted their recommended method for customer testing the accuracy of the probes. The current method involves submerging the probes in "crushed ice" in an insulated container within a refrigerator overnight. We have also subjected the probes to a "melting-ice test" where an iced-down set are subjected to room temperature overnight so that they tend to record the full range of temperatures they are normally exposed to in the field. It is necessary to test the full range of

temperature. The details of this method (as well as the findings from all four attempts) can be found linked from the eMOLT Administrators manual under "Probe Comparison/Calibration" link. A summary of the results is presented in the "Discussion" section below.

- *Association Reps download raw data:*
    - The first step in processing data involves extracting records from the probe via the manufacturers software. The serial port connection downloads a years worth of data in a few minutes time. The raw binary file is then saved on disk with filename following a particular eMOLT convention. For the first few years of eMOLT, the convention was to name the raw file ABBBCCDD.dtf where "A" was a "t" for TidBit or "m" for Minilog, "BBB" was the last 3-digits of the serial number, "CC" was the consecutive time the probe was deployed at the site, and "DD" was the two digit site code. After some discussion at a June 2002 administrators meeting in Woods Hole, it was decided to change the convention to be ABBCCDD.dtf where "A" was again a "t" or "m", "BBCC" is the site code (which typically includes the participants initials), and "DD" is the consecutive time a probe was deployed at that site. It should be noted that the operator can assign this filename at the time the probe is initialize but it is often revised at download time to more accurately represent the probes recent history. For this reason, it is advantageous to have all the probe documentation organized and available at the time of download so that the raw data file can be properly labeled. While some of the pertinent information such as the probes serial number is saved within the header of the raw data, it is best to save a raw data file with names that pertain to the deployment because having to rename them at a later date can lead to mistakes.
    - The probe is reset for the subsequent deployment after checking the PC clock setting.
- *Science party data processing:*
    - The next step in processing is to "export" to a raw ascii file by choosing the default "custom" option at in the export menu. The output name is the same with a ".txt" extension.
    - At this point, a set of MATLAB routines are executed to parse, check, crop, and reformat the data. Called from a main driver program called "emolt.m" are the following subroutines:
        - tidbit2mat.m to parse the record and calculate yearday
        - emolt_rawplot.m to make the detail plot of the deployment allowing the operator to manually:
            - crop the data to "in-water" records
            - specify whether the objective editing of "on-deck" records was adequate based on a default range and delta check and, if not, specify a different range and delta check based on the record at hand (note: the same range and delta check does not apply to all cases). The purpose of this editing is to objectively filter out records obtained during the haul operations.
            - convert to farenheight time series
        - subsequent statements to:

        - prompt user for necessary info such as serial number, site code, and water column depth
        - generates a reformatted file
    - After specifying this new filename in an "emolt_sensor.ctl" file an SQL loader program is executed to commit data to the emolt_sensor table
    - At this point, after the data is stored in ORACLE, the emolt.m file is run again. In this phase it performs a Perl shell routine to extract the data from the user specified site. The MATLAB code then conducts a user-specified running average (typically 7-days) and plots it on a calendar axis for comparing color-coded years.
    - The participants individual web site is then modified to include links to new plots.
    - Both the raw detail plot and the multi-year summary plot are sent to the participant in the US Mail. Note: For the first few years of eMOLT, an email was sent to the participant along with a copy to their association rep, but we found that these emails were not always read by participants and that the mailing was more effective in reaching the individuals.

Part of the eMOLT effort is in collection of historical data. As described in a draft document at http://www.nefsc.noaa.gov/~jmanning/whwt/newt.html, there is a significant set of both dockside and deeper bottom-water time series of temperature. We have been coordinating our efforts, for example, with the Mass Division of Marine Fisheries who now have nearly two decades of records from several sites in Mass Bay. Most of this data along with that of several other labs have been merged with the eMOLT database and, in most cases, is the sole internet source of the data. The century-long set of near-daily temperatures from both Woods Hole Harbor and Boothbay Harbor are part of this archive. This historical archive now applies not only to temperature and salinity records but to drifter paths as well. A recent effort to collect drifter data from various programs like ECOHAB and GLOBEC should allow users to view & download Gulf of Maine drifter trajectories from selectable years and months along with the SMCCeMOLT drifter tracks.

One of the most exciting partnership associated with eMOLT is GoMOOS. The Gulf of Maine Ocean Observing System has been very helpful in providing a professional web interface to eMOLT data along the coast of Maine. Linked from their main page is an eMOLT web mapping utility powered by the open source University of Minnesota MapServer routine that allows users to browse eMOLT data. It includes zoomable maps and time series plots (see .http://www.gomoos.org/emolt/emolt_map.shtml). We hope to expand this partnership with GoMOOS. In order to allow users to view eMOLT drifter tracks we have followed GoMOOS example by building a web mapping utility of our own currently posted at http://fish.nefsc.noaa.gov/circmods/first_init.html. The objective of this site is to be able to overlay various point objects on model-generated flow fields and observed drifter tracks in the Gulf of Maine. We appreciate the help of GoMOOS in leading the way.

eMOLT will be represented at a 11 March 2004 workshop on "Building a Spatial Data Framework in the Gulf of Maine to Support Benthic Mapping for Resource Mapping" at the Bedford Institute of Oceanography. Funded by the Federal Geographic Data Committee, this workshop will bring together various data servers around the gulf to encourage standardization in GIS layers of information. Partners are asked to comply with the protocol of "Web Mapping Service" and FGDC metadata standards.

Findings

Having described in some detail the various "methods" associated with our database management effort, what can we say of our "findings"? It may be best to list the most important items in brief bullets at follows:

- cooperative research project "methods" should be adjustable... the individual responsible for sampling design, whether it is a scientist or a fishermen, can not expect to devise a strict set of protocols at the start of the project until they have become better informed of one another's practice
- minimize the documentation requirements for fishermen with simplified logsheets and ensure they are supplied with them regularly
- metadata (i.e. data about data) is more difficult to collect than the data
- frequent contact with participants (every few months) is needed to remind them of the program and the protocol
- more than one project representative per fishing "association" does not necessarily help and can, in fact, cause confusion
- units of measurement are much different in the fishing industry (lorans, fathoms, farenheight, local time vs lat/lon, meters, Celsius, GMT)
- email and website communication is not the best means of getting information to the fishing industry, mailing information is better
- attending association meetings and annual forums is the best means of contacting participants
- if electronic loggers are to be used in the future, units need to be simple, adjustable by project personnel, and not dependent on commercial enterprise
- source of error need to be identified and quantified (see "Discussion" below) early in the project in order to minimize their effect on the quality of data collected

Table 2. eMOLT database statistics as of February 2004

# fixed sites documented	169
# temperature records	1,086,492
# salinity records	72,752
# drifter sightings reported	55
# drifter automated fixes	15,407
# people returning data	104

Discussion

Other than these general "findings" (ie statements on cooperative research in general), there are points to be made in particular on the subject of eMOLT's database management and documenting the degree of uncertainty. There are four sources of error which need to be addressed separately:

- *thermal sensor uncertainty*

According to the manufacturers, the ONSET temperature probes are accurate to within 0.1-0.2 degC, depending on the model and range of temperature it is engineered to monitor. In order to obtain records with less than 0.1 degC accuracy, we had the ONSET corporation engineer a probe specifically for our needs in the Gulf of Maine before we began the project. Given a $500 engineering fee, they provided a more accurate probe for the temperature range of 0-20 degC. In any case, accuracy of less than 0.2 degC has been sufficient to capture the signals of interest which are typically 10 to 100 times this value. The interannual signals of interest have been more than 2 degC and some of the tidal signals have recorded nearly 20 degC variations.

In order to check on the accuracy of each probe, we have implemented a protocol to ice bath a collection of the probes on at least an annual basis. We have found only 1 of the 55 tested thus far to register unacceptable values. This probe will be tested in future baths. As we have become more proficient in conducting the proper controlled experiments, we will be more apt to detect malfunctioning probes. For example, our protocol now includes both the "crushed ice test" and the "melting-ice test" in order to check for both absolute 0 degC as well as 0-20 degC range.

- *conductivity sensor uncertainty*

The error associated with fine grain sediment affecting the conductivity reading of the Microcat probe is discussed in detail in the Phase II Final Report. This, the primary source of salinity error, resulted in episodic variations of nearly 1.0 PSU in the records as the instrument was subjected to variable concentrations of mud.

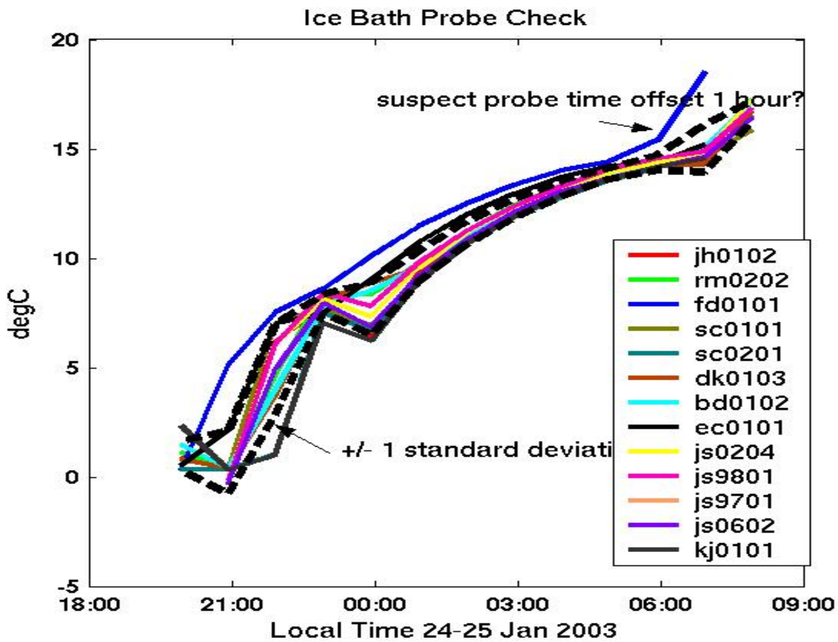


Figure 1. Example calibration check of several probes during the Jan 2003 download. The large differences (1-2 degC) between probes in these earlier test were due a combination of time offsets and the "cubed ice test" not being a controlled environment. Subsequent "crushed ice test" resulted in differences of less than 0.4 degC.

- *clock uncertainty*

The importance of time stamps associated with each record needs to be highlighted. In the figure above, for example, where more than a dozen probes were tested together in an ice-cubed bath overnight on 24 January 2003, one probe, in particular, appears to be offset by a few degrees. It was actually a clock problem however where that probe was initialized an hour different from the others. The errors associated with time stamps can therefore lead to several degrees of error in a point-by-point comparison of two probes. Because of this time stamp error in many of the early years of eMOLT datasets, detailed analysis of hour by hour events will be difficult. Daily and weekly averages should be unaffected by such problems. Probe clocks should be checked with each download. The ice-bath procedure that is typically conducted at the end of each deployment provides an opportunity to test for clock offsets. The EST convention needs to be strictly enforced.

- *depth-dependent temperature effect uncertainty*

As mentioned in the methods section above, the probes are deployed at "fixed" sites in particular water depths. It is expected that the probe depth may vary, however, by as much as 5% of the water column depth as it is hauled from day to day or, in typical offshore lobstermen's cases, week to week. How does this slight variation in depth affect the temperature record? In most of the study area, the variation in temperature with water depth is insignificant especially in the deepest portion of the water column. Since most of the probes are located well below the seasonal thermocline, vertical translation of their position from time to time does not affect the record. However, there are a number of probes that are situated either in steep bathymetry, shallow water, or in the vicinity of oceanographic fronts such that occasional vertical translation of the probe is significant. The error associated with this variation in depth can be greater than variations due to water movement and mixing process. This makes interpretation of the time series difficult. Since dropping a trap in different depths may result in daily to weekly (depending on how frequently the trap is hauled) variations in the record, the analysis of the record is restricted to longer signals. In some sites, only the interannual variations can be considered . In the future, an estimate of temperature gradients with respect to depth need to be quantified at each location in order to determine the associated error in trap depth placements.

In phase IV, a specific investigation of this issue will be conducted with a set of pressure sensors on a series of probes deployed on the steeply sloping shelf edge. Funding was allowed to provide some of the Atlantic Offshore Lobstermen with a set of temperature and pressure probes to deploy along a cross-isobath string. The results of this process study will help quantify the depth-dependent temperature variations.

Budget Report