Appendix IV

Marine fish monitoring at FORCE: Updated report on processing and analysis

Marine Fish Monitoring at FORCE:

Updated Report on Processing and Analysis

Historical Dataset: August 2011 thru May 2012
and
Contemporary Dataset: May 2016 thru August 2017

Prepared by: Louise P. McGarry, Ph.D. and Gayle B. Zydlewski, Ph.D.
University of Maine, School of Marine Sciences

May 2, 2019

Update: June 12, 2019 (Appendix C only)

Executive Summary

This is a report of the Environmental Effects Monitoring Plan (EEMP) for sixteen hydroacoustic fish surveys conducted in Minas Passage between August 2011 through August 2017. Six surveys were led by Dr. Gary Melvin between August 2011 and May 2012 (herein "historical"). The nine FORCE-funded surveys were conducted between May 2016 and August 2017 (herein "contemporary").

The four main objectives of this report are to:

1. provide FORCE personnel information relevant to understanding the historical and contemporary datasets, including technical guidance.

2. convey to FORCE a set of scripts that can be used to automate preparation of hydroacoustic data for analyses.

3. provide examples of data visualizations, including a case study example of drilling down into the data to gain insight into the summarized data.

4. provide a statistically rigorous analytical *approach* to quantifying the relationship between observed volume backscattering strength (S_v: proxy for fish density) and predictor variables (e.g., site, season, tide phase). *This approach was designed and approved by a University of Maine statistician*.

An overall *approach* to understand the information contained within the hydroacoustic datasets, including data visualization and statistical analyses are detailed in the report. In addition, the scripts with coding for the analytical approach and data visualizations are provided such that deeper explorations of the data may be taken to investigate questions specific to the needs of FORCE. Selected results are presented below.

Because of the large percentage of "zero" observations (59%) contained in the dataset, the analytical steps were separated into two steps: (1) implement a statistical modeling approach (GLM) to examine fish presence:absence in relation to the predictor variables, and (2) implement a statistical test (ANOVA) to examine relative fish density (backscattering strength, S_v, as proxy) in relation to the predictor variables. The predictor (or explanatory) variables available in the dataset were: temporal (historical vs. contemporary, or by survey), spatial (CLA vs. reference study area, or by transect), and environmental (tide phase, diel state, or with and against predicted tidal flow). Metrics of interest, such as minimum, maximum, mean, and median S_v as well as the estimated marginal means used in the relative fish density analyses are included within the report.

Highlights from the analytical approaches for understanding how fish presence:absence and relative density compare for the following data aggregation levels:

1. Research Program, the data collected included historical (2011-2012) and contemporary (2016-2017) data, when compared:

- highest maximum relative fish density was observed during the contemporary program
- the probability of observing fish was higher during the contemporary program (42.8%) and statistically differed from the probability of observing fish during the historical program (31.5%)
- relative fish density during the contemporary program was statistically different from the relative fish density during the historical program

Implications: Based on both metrics (presence:absence and S_v), the historical and contemporary datasets were statistically different. The differences may be artifact due to differences in survey design and execution, and therefore the datasets are simply not comparable, or the differences may signal that fish use of the site has changed. Additional analyses would need to be conducted to examine whether or not future comparisons with the historical dataset would be constructive. Given the statistical differences between the historical and contemporary research programs, the remaining analyses were conducted using the ***contemporary dataset only***.

2. Study Area, data were collected in two distinct locations, CLA and reference, when compared:

- highest maximum relative fish density was observed in the CLA study area
- the probability of observing fish was higher in the CLA study area (44.9%) and statistically differed from the probability of observing fish in the reference study area (38.4%)
- relative fish density within the CLA study area statistically differed from within the reference study area

Implication: The statistically significant differences between the CLA and reference site may indicate that the reference site is not sufficiently representative to serve as reference for the CLA.

3. Tide Phase, data were collected during the following stages: high-slack, ebb, low-slack, flood, when compared:

- highest maximum relative fish density was observed during low-slack
- the probability of observing fish was highest during the ebbing tide flow (49.3%) and statistically differed from the probability of observing fish during any of the remaining three tide phases
- relative fish density observed during the ebbing tide flow statistically differed from all other tide phases

Implication: The ebb tide is an important tidal phase to focus on for an understanding of fish in this site.

4. Diel State, data were collected during the following time periods: dawn, day, dusk, night, when compared:

- highest maximum relative fish density was observed during the day
- the probability of observing fish was highest during the night (52.5%) and statistically differed from the probability of observing fish during any of the remaining three diel states
- relative fish density observed during night statistically differed from observations during all other diel states

Implication: Data collection at night is important for understanding fish presence in this location.

5. Tide Phase and Diel State

Tide Phase and Diel State were used as an example scenario of combining explanatory variables where the effects of the tide phase and diel state add to one another (additive) and where the effects of the tide phase interact with the diel state (interactive):

- variance in the probability of observing fish was better explained (statistically significant) using the complexity of the interaction of the two explanatory variables

Implications: The influence of the variety of predictor variables and their additive versus interactive impact should be further explored.

6. Survey, data were collected during nine contemporary surveys, May 2016 through August 2017, when compared:

- the probability of observing fish and observations of relative fish density varied among surveys
- no seasonal trends were noted

Implications: Given the absence of a seasonal pattern and the preponderance of statistical differences between surveys, it may be advisable to increase sampling frequency within each month, sampling on consecutive days in order to get a finer scale understanding of the patterns and variability of fish presence and density in Minas Passage. May, with its particularly high and wide-range of observed backscattering values and apparently distinctive spatial pattern, appears to be an important month for surveying to continue to gather time-series data to help with interpretation.

7. Transect, data were collected along nine transects, six in the CLA study area (N0, N1, N2, N3, N4, N5) and three in the reference study area (S1, S2, S3), when compared:

- highest maximum relative fish density was observed on transects N0, N1, N2
- probability of observing fish on transect N0 statistically differed from all other transects
- relative fish density observed on transect N0 statistically differed from all other transects
- among the remaining transects there were several pairs of adjacent transects for which the probability of observing fish was not statistically significant
- the spatial pattern of the statistical significance of the observed relative fish density was more complex than the pairs noted in the probability of observing fish

Implications: The adjacent pairing of transects for which the probability of observing fish were not statistically different could provide guidance if a decision was at hand to adjust the survey design to include one of each pair rather than both. The transect groupings as produced by the relative fish density findings must also be considered. Given that these findings were based on highly summarized data (the full contemporary dataset summarized by transect), exploration of the statistical results at finer scales, such as transect data summarized by survey, may provide more robust guidance.

General observations

Within the contemporary dataset, where the number of categories within a predictor variable exceeded two, the statistical results of the presence:absence analysis generally differed from that of the relative fish density analysis in terms of which of the categories statistically differed or not. These findings suggest that the presence:absence ratio of observations was not an indicator of the relative density of fish passing under the transducer.

To gain insight into these findings, further analyses should be conducted on data summarized at finer scales. For example, is night the period of highest probability of observing fish when examined at the level of each monthly survey? Using the scripts provided (to prepare and analyze the data), FORCE personnel can apply the same approach to answer questions pertinent to the needs of FORCE.

It should be noted that echosounder gain settings used in the contemporary dataset were standardized rather than calibrated with each survey. This approach was used because data collection procedures for calibration data were insufficient to provide reliable calibration settings. Calibration procedures were subsequently updated starting with survey 15. For more information see the Calibration Quality Control Report (McGarry and Zydlewski, 2018) and the Notes for EK80 CW Calibration Settings (McGarry and Zydlewski, 2019).

Contents

Introduction

In preparation for tidal power development, FORCE-funded 24-hour mobile hydroacoustic surveys have been conducted since May 2016 to establish a baseline understanding of fish presence in the region of the Crown Lease Area (CLA) in Minas Passage. The ultimate goal of the hydroacoustic surveys was to collect sufficient data to document changes in fish presence that may be attributed to the presence of devices engineered to convert tidal energy to electricity (e.g. TISEC: Tidal In Stream Energy Conversion devices). Nine surveys were conducted between May 2016 and August 2017. Each survey traversed an established grid of transects, including six transects within the CLA and three reference transects located near the south shore of the channel, outside the region influenced by the presence of turbines in the CLA. Data collected in the reference region were to be used to help interpret changes in fish presence in the CLA. Herein, this dataset will be referred to as the "contemporary" dataset.

In addition to the contemporary dataset, an "historical" dataset of seven surveys was available for inclusion in analyses. The historical surveys, led by Dr. Gary Melvin (Melvin and Cochrane, 2014), were conducted between August 2011 and May 2012. Similar to the FORCE surveys, the historical surveys traversed an established grid: nine transects within the CLA and one reference transect located near the south shore of the channel.

There are four main objectives for this report:

1. provide FORCE personnel information relevant to understanding the historical and contemporary datasets, including technical guidance. That information is contained mainly in the Methods, Results, Appendix A, and Appendix B sections of this document. A glossary and list of abbreviations is available in Appendix E.

2. convey to FORCE a set of scripts that can be used to automate processing of hydroacoustic data (Appendix C), including export from Echoview in a variety of echo integration configurations and the preparation of the exported data for analyses. The processing steps are described in the Methods section of this report and are described more fully in comments internal to the processing scripts. In addition, scripts by which to generate visualizations of the data, and execute analyses are included (Appendix C).

3. provide examples of the data visualizations, including a case study example of drilling down into the data to gain insight into the summarized data. The TISEC presence:absence data were used in this example (see: Results – Data Visualizations).

4. provide an analytical approach to quantifying the relationship between the observed volume backscattering strength (S_v) and the variety of "predictor" variables in the dataset. Because of the large percentage of "zero" observations (59%) contained in the dataset, the analytical steps were separated into two steps: (1) implement a generalized linear model (GLM) to examine fish presence:absence in relation to the predictor variables, and (2) implement an analysis of

variance to examine fish density in relation to the predictor variables (see: Methods – Analytical Approach and Results – Analytical Approach). *This approach was designed and approved by a University of Maine statistician.*

Presented here is an overall approach to understanding the information contained within the hydroacoustic datasets. The scripts are provided such that deeper explorations of the data may be taken to investigate questions specific to the needs of FORCE.

Methods

Study Area

Minas Passage is located at the entrance to Minas Basin. The hydroacoustic survey design encompassed a rectangle approximately 5-km long by 2-km wide, reaching nearly shore-to-shore. A set of transects were executed in the Crown Lease Area (CLA) along with a set of reference transects near the southern shore (Figure 1).

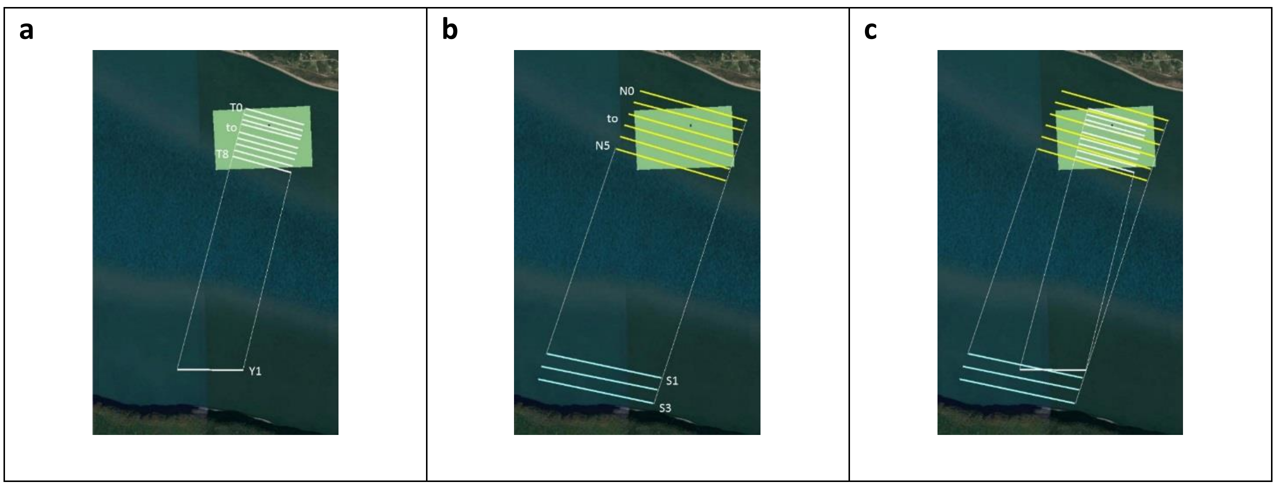


Figure 1. ***Grid Survey Plan for each Research Program****. CLA region is shown as green box for reference.* ***(a)*** *historical survey grid,* ***(b)*** *contemporary survey grid, and* ***(c)*** *historical survey grid superimposed on the contemporary grid to show match/mismatch in survey area.*

Historical Data: 2011-2012

In 2011 and 2012, seven mobile hydroacoustic surveys (Table 1) were conducted using a split-beam echosounder (Simrad EK60) operating at 120 kHz using the charter vessel FUNDY SPRAY (Melvin and Cochrane, 2014). The mobile surveys traversed an established grid (Figure 1a) with nine, generally east-west trending, transects (T0 – T8) executed in the CLA study area and one reference transect (Y1) positioned along the 30 m contour across the channel near the southern shore. Transects were nominally one kilometer in length. During each grid pass, the transects were traversed once, alternating survey direction on each successive line. That is, each transect was traversed in a direction either with the direction of tidal flow or against the direction of tidal flow. Data were collected during the cross-channel transits (X1 and X2) between the CLA and reference study areas. Data from the cross-channel transits are excluded from analysis. No TISEC devices were present during the seven surveys. Herein this dataset will be referred to as the "historical" dataset. Calibration settings provided as Cal2012_120.ecs by Gary Melvin were used as the calibration settings for the historical datasets.

See additional notes regarding the historical dataset in Appendix A: Historical Survey Detail.

Table 1. Historical Surveys. *Each survey consists of three to twelve repeats of the grid defined by the following transect lines: T0, T1, T2, T3, T4, T5, T6, T7, T8, X1, Y1, X2. Numbers in parentheses in the Month column indicate grid coverage (partial grids:completed grids). Historical survey data were time-stamped with Greenwich Mean Time (GMT). Time posted to the Start Time and End Time columns are shown in GMT with local time shown in parentheses. Only data collected from "T" and "Y" transects were included for analyses.* ***Please see extensive notations to this table in Appendix A (Table A1)****.*

Survey	Month	Start date	Start time	End date	End time	Day/ Night	Temperature (°C)	Turbine presence	Moon Phase Tide Range
1	Aug 2011 (4:3)	2011-08-22	11:45 (08:45)	2011-08-22	21:28 (18:28)	D	15.4	No	7 m
2	Sep 2011 (4:3)	2011-09-19	10:55 (07:55)	2011-09-19	20:23 (17:23)	D	15.7	No	8 m
3	Oct 2011 (4:3)	2011-10-03	09:53 (06:53)	2011-10-03	20:18 (17:18)	D	15.0	No	10 m
4	Nov 2011 (3:3)	2011-11-22	14:22 (10:22)	2011-11-22	22:32 (18:32)	D	10.3	No	/ 11 m
5	Jan 2012 (10:9)	2012-01-25	18:32 (14:32)	2012-01-26	16:15 (12:15)	D/N	3.6	No	11 m
6	Mar 2012 (12:11)	2012-03-19	14:23 (11:23)	2012-03-20	13:33 (10:33)	D/N	2.5	No	/ 9 m
7	May 2012 (5:4)	2012-05-31	12:09 (09:09)	2012-05-31	23:12 (20:12)	D	9.5	No	/ 10 m

Contemporary Data: 2016-2017

In 2016 and 2017, nine mobile hydroacoustic surveys (Table 2) were conducted by FORCE personnel using a split-beam echosounder (Simrad EK80) operating at 120 kHz using the charter vessel NOVA ENDEAVOR. Each survey traversed an established grid (Figure 1b) of transects similar to the historical survey-grid but differing in some fundamental ways. Transects were nominally two kilometers in length, generally trending east-west. Unlike the historical dataset, during each grid pass each transect was traversed twice before moving to the next transect line. That is, each transect was traversed once in the direction of tidal flow, “with”, and once in the direction counter to tidal flow, “against”. Six transects were located within the CLA study area and three reference transects were located near the south shore of the channel, outside the region influenced by the presence of turbines in the CLA. Data collected in the reference study area was intended to be used to help interpret changes in fish presence in the CLA. A TISEC device was present during four of the surveys: Nov 2016, Jan 2017, Mar 2017, and Jul 2017 (Table 2). Herein, this dataset collected by FORCE personnel will be referred to as the “contemporary” dataset. See Appendix A for a table summarizing the survey design characteristics for the historical and contemporary datasets (Table A3). Table A4 provides a visual display of the survey timing within each year. See additional notes regarding the contemporary dataset in Appendix A: Contemporary Survey Detail.

Calibration gain parameters for contemporary surveys 1 through 9 were “standardized” such that TransducerGain and SaCorrection were set to Simrad default settings of 27.00 dB and 0.0 dB. The remaining calibration parameters were in keeping with standard practice: Beam pattern settings were as per the Simrad factory transducer measurements upon delivery of the instrument. Survey environmental settings for salinity and temperature were as provided by FORCE personnel. Soundspeed and absorption coefficient were calculated from the measured salinity and temperature using the Echoview Sonar Calculator (Echoview Software Pty Ltd). This approach was used because data collection procedures for calibration data were insufficient to provide reliable calibration gain settings. While this approach is not in keeping with standard practices, and should not be relied upon for future surveys, it was a resolution settled upon after consultation with experts in the hydroacoustic community. Calibration data collection procedures were updated subsequently starting with survey 15. For more information see the Calibration Quality Control Report (McGarry and Zydlewski, 2018) and the Notes for EK80 CW Calibration Settings (McGarry and Zydlewski, 2019).

Table 2: Contemporary Surveys. *Each survey consists of 4 repeats of the grid defined by the following transect lines: N0, N1, N2, N3, N4, N5, South_CW, S1, S2, S3, North_FM with calibration files. Contemporary survey data were time-stamped with local time. Time posted to the Start Time and End Time columns are shown in local time. Only data collected from "N" and "S" transects were included for analysis. Please see extensive notations to the table below in Appendix A (Table A2).*

Survey	Month	Start date	Start time	End date	End time	Day/ Night	Temperature (°C)	Turbine presence	Moon Phase Tide Range
1	May 2016	2016-05-28	06:01	2016-05-29	05:35	D/N	7	No	10 m
2	Aug 2016	2016-08-13	09:09	2016-08-14	07:40	D/N	15	No	7 m
3	Oct 2016	2016-10-07	05:45	2016-10-08	04:21	D/N	15	No	8 m
4	Nov 2016	2016-11-24	08:38	2016-11-25	09:07	D/N	8.0	Yes	8 m
5	Jan 2017	2017-01-21	06:55	2017-01-22	05:55	D/N	1.5	Yes	7 m
6	Mar 2017	2017-03-21	08:24	2017-03-22	06:04	D/N	4	Yes	7 m
7	May 2017	2017-05-04	19:57	2017-05-05	18:21	D/N	5	Yes (free spinning)	9 m
8	Jul 2017	2017-07-03	21:34	2017-07-04	19:09	D/N	12	No	/ 8 m
9	Aug 2017	2017-08-30	18:53	2017-08-31	17:37	D/N	15.7	No	7 m

Data Processing

Echosounder data files were processed using Echoview ("EV"; Echoview Software Pty Ltd): Version 7 for the historical EK60 dataset and Version 8 for the contemporary EK80 dataset. Data were processed to remove backscatter from the region of the transducer nearfield and from non-biological sources (e.g. bathymetry and entrained air). Minimum thresholds were set for volume backscattering strength (S_V: -66 dB re 1 m^{-1}) and target strength (TS: -60 dB re 1 m^2) as described in Daroux and Zydlewski (2017). More information about processing details is available in the Historical Survey Detail section of Appendix A. Using Echoview, the processed data were then integrated and exported in a variety of echo integration configurations for use in analyses. Additional processing to prepare the exported data for analysis was conducted using R Software (R Core Team 2018), an integrated suite of software facilities for data manipulation, calculation, and graphical display.

Full Water Column x 20-m Along-Shiptrack data

To examine the distribution of relative fish densities throughout the study area, echo integration data were exported from Echoview binned in 20-m along-shiptrack distances, integrated over the full water column (Figure 2), herein referred to as the "20-m dataset". The 20-m along-shiptrack distance was selected to minimize autocorrelation (Daroux and Zydlewski, 2017). To place the echo integration data within a meaningful environmental context, using date and time as the common variable, the exported echo integration data were merged with a file holding the necessary metadata such as: diel state (i.e. dawn, day, dusk, night: a proxy for light level), tide phase (i.e. ebb, low-slack, flood, high-slack), study area (i.e. CLA, reference site), grid pass (i.e. "1", "2", "3", etc.), transect number (i.e. "T0", "T1", …"Y1", "N0", "N1", …"S1", "S2", "S3", etc.), data collection direction (i.e. "with" the direction of tidal flow or "against" direction of tidal flow or "along" designating the cross-channel transits), and turbine presence (true, false), among others. Additional steps were required for the historical dataset before the metadata merge could be executed.

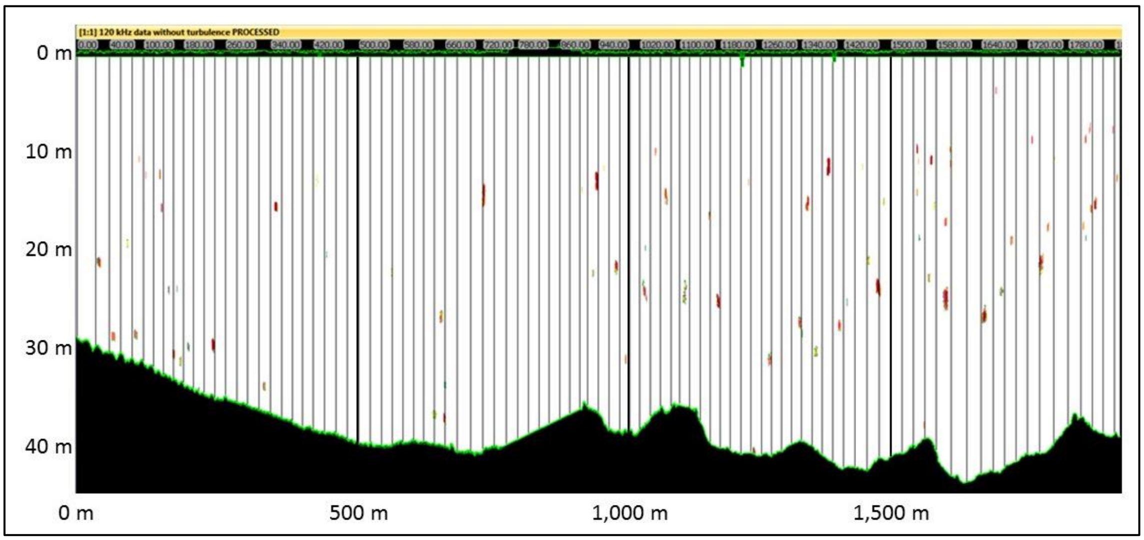


Figure 2. Configuration of the "20-m dataset". *Echogram shows data collected along one example transect. Colored marks indicate recorded backscatter from target in the water column. These are interpreted to be fish or fish aggregations. Vertical lines demarcate the "20-m along-shiptrack distances, over the full water column" integration cells used for analysis. x-axis is distance along transect. y-axis is range from transducer.*

Because the historical echosounder dataset and the associated historical datasheets were recorded with "time" set to Greenwich Mean Time (GMT), and because the contemporary data were recorded and will continue to be recorded with "time" set to local time, historical time in GMT was converted to local time for both the exported echo integrated data and the datasheet, paying special attention to periods crossing midnight such that the date was updated as needed, and paying attention to time of year as the conversion from GMT to local time is 3 or 4 hours depending on time of year. With the datasheet populated with local time, the appropriate diel, tide, and with/against phases were posted and the merge then generated a historical dataset with appropriately associated metadata. See notation "2" to Table A1 in Appendix A for more information regarding the conversion of the historical dataset time from GMT to local time.

In preparing the full dataset (historical plus contemporary) for analysis, in addition to appending columns associating the metadata to the echo integration data, columns were calculated from the data exported from Echoview to facilitate calculations (e.g. converting the backscattering strength (S_v) exported by Echoview to its linear form (s_v) using the relationship in Equation 1 (MacLennan *et al.*, 2002). Additionally, commands were executed to exclude data meeting certain criteria such as the "along" data collection direction in order to exclude the cross-channel transit data, and to exclude any "Interval" assigned a value of 0. (See the EV Exported Data Notes section of Appendix A for more information about "Interval".)

$$S_v = 10 * log_{10}(s_v) \qquad \text{Equation 1}$$

To standardize the number of grid passes per historical survey, which varied from three to twelve passes and not all of which were complete grid passes, three grid passes from each survey were selected for inclusion in analyses, excluding data from all other historical grid passes. See notation "3" to Table A1 in Appendix A for the list of grids included for each survey and notes regarding selection criteria. The derived "20-m dataset" of historical and contemporary data resulted in a dataset of 71,016 observations (11,347 and 59,669, respectively). The analytical variable of interest here is the mean volume backscattering strength (S_v).

Due to logistical difficulties and safety considerations, physical sampling was not available by which to confirm the identity and sizes of the fish generating the observed backscatter. Therefore, in this report we are using the observed mean volume backscattering strength (S_v) as a proxy for relative fish density. "S_v" and "relative fish density" are used interchangeably.

Analytical Approach

Data visualizations and analyses were produced using R Software (R Core Team 2018).

The 71,016 datapoints that constituted the "20-m dataset" included 29,105 non-zero S_v values (41%) and 41,911 zero values (59%). A zero-value indicated that no observations above the thresholds were observed in a cell size of 20-m along-shiptrack distance integrated over the whole water column (thresholds: S_v = -66 dB and TS = -60 dB; Daroux and Zydlewski, 2017). In order to implement analyses for a dataset of which zero values constituted 59% of the observations, a three-pronged approach to the analyses was undertaken. First, a set of data visualizations were constructed to explore the observations contained in the dataset. Second, to investigate the relationship between fish presence and the spatial and temporal variables, integrated S_v values were converted to fish "presence" and "absence" which was then used in the analyses. Third, because the relationships of the magnitude of relative fish density is also of interest, the non-zero S_v values were analyzed in relation to the spatial and temporal variables.

Exploratory Data Visualizations

To gain an understanding of the underlying historical and contemporary data available for spatial and temporal analyses, a series of data visualizations were produced in the form of histograms, boxplots, and frequency plots. The preponderance of zero observations created special challenges for data visualizations due to the data range when including zeros, and the overriding influence of the preponderance of zero-values. Therefore, the produced plots exclude zero values. In addition, the 29,105 non-zero backscattering values available to plot were distributed over a range of nearly 8 orders of magnitude, 3.6e-12 m^{-1} to 8.4e-5 m^{-1}. Calculations were done with the data in linear form (s_v: mean volume backscattering coefficient) and converted to the log form (S_v: mean volume backscattering strength) in order to display the full dynamic range of the data (S_v: -114.4 dB to -40.8 dB). In its log form, a change in S_v of 3 dB represents a doubling (+3 dB) or a halving (-3 dB) in linear terms. A change in S_v of 10 dB indicates a change of one order of magnitude, whereas 20 dB indicates a change of two orders of magnitude. R coding to produce these and other plots are included with this document (Appendix C), including coding to produce the plots with or without zeros and/or using the linear volume backscattering coefficient (s_v).

Boxplot Conventions

The central rectangle in the box plots include the 25th through 75th data percentiles. Thick line indicates median. The mean, calculated in its linear form (s_v), is indicated by an open square. Whiskers are placed at the minimum and maximum extremes of the data unless the range to the extremes are greater than 1.5 times the size range of the box. Where the range to the extreme is greater than 1.5 times the size range of the box, the whisker is placed at the 1.5 distance range from the edge of the box and any datapoints beyond the 1.5x are plotted individually as open circles. All calculations were made with the data in its linear form (s_v) and then converted to its log form (S_v) for display. The physical length of the whiskers, although symmetrical on a linear scale relative to the central rectangle, become asymmetrical when plotted on a log scale.

Analyses – Fish Presence:Absence

To investigate the relationship between the spatial and temporal distribution of the presence of fish and the predictor variables available in the “20-m dataset”, a binary logistic regression was implemented in R using the Generalized Linear Model (GLM) function. The binary logistic regression was chosen because the dataset contained a high percentage of zeros (59%). To prepare the dataset for analyses, a variable called “FishPresence” was created. The variable was populated with “zero” wherever the observed integrated mean volume backscattering strength (S_v) was zero. Wherever S_v was any value other than zero, the variable was populated with “one”. This created a binary “response” variable denoting presence (“1”) or absence (“0”) that could then be evaluated in relation to the “predictor” variables in the dataset. The predictor variables in the dataset were categorical: temporal (historical vs. contemporary, or by survey), spatial (CLA vs. reference study area, or by transect), and environmental (tide phase, diel state, or with and against the tidal flow). While a 50:50 ratio of fish presence:absence is not intrinsically of biological interest, the 50:50 ratio provides a baseline against which the significance of the differences in fish presence relationships to the explanatory variables can be measured.

The R output from the GLM model includes a table of the coefficients for each of the variables. The first variable listed is designated as the “reference” variable. The results reported in the reference row are in direct reference to the null hypothesis: “Is the ratio of fish-presence:fish-absence equal to 50:50?”. The table includes estimates of the y-intercept of the resulting model, the standard error, and associated p-value for each categorical variable. The closer the presence:absence ratio to 50:50 for the reference variable (the first variable in the list), the closer to zero is the estimated intercept. For a reference variable intercept that is not equal to zero, the p-value for that row will indicate whether the divergence from 50:50 is significant ($p < 0.05$). For all variables other than the reference variable, the values reported in the table are relative to the reference variable. Therefore, the p-values listed on each subsequent row of the table are a measure of the significance in the difference of the fish presence:absence ratio from that of the reference variable.

Given that each subsequent row is reported relative to the reference variable, the sign and magnitude of the values in the y-intercept column give an indication as to whether presence:absence ratio is very different from that of the reference variable and in which direction. The y-intercept for the 50:50 ratio for those variables reported relative to the reference variable is calculated by adding the value in the y-intercept column for the variable of interest to the value in the y-intercept column for the reference variable. The closer the summed y-intercept is to zero is an indication that the presence:absence ratio is closer to 50:50.

To determine whether the presence:absence ratios among the remaining variables statistically differ from each other, the values contained in the table were used to calculate the z-value (a measure of standard deviation providing guidance as to whether to reject the null hypothesis: “Is the presence:absence ratio for variable x equal to the presence:absence ratio for variable y?”) and from the z-values, the p-values were calculated (a measure of the probability that the null hypothesis is falsely rejected).

Example GLM modeling results for presence:absence based on individual predictive variables are included in the Results section of this report. Also included is an additive example where more than one predictor variable was included in the model, and an interaction model whereby the interactive effects of multiple predictor variables was modeled. The results from these examples can be used to guide deeper inquiry into the particulars generating the illustrated relationships at the aggregated levels used here.

Example R code to model FishPresence in the contemporary dataset:

1. using a single predictor variable: Tide Phase
   glm.TideC <- glm(FishPresence ~ Tide, data=data_contemporary, family=binomial)
2. using a single predictor variable: Diel State
   glm.DielC <- glm(FishPresence ~ Diel, data=data_contemporary, family=binomial)
3. using two additive predictor variables: Tide Phase and Diel State
   glm.Tide_DielC <- glm(FishPresence ~ Tide + Diel, data=data_contemporary, family=binomial)
4. using two interactive predictor variables: Tide Phase and Diel State
   *glm.Tide_DielCx <- glm(FishPresence ~ Tide * Diel, data=data_contemporary, family=binomial)*

Once the hierarchy of models was generated, an analysis of variance using the chi test was run to evaluate whether the higher complexity (modeling with more than one variable: additive or modeling with more than one variable: interactive) produced a better fitting model than those models lower in the hierarchy.

Example R code:

anova(glm.TideC, glm.Tide_DielC, glm.Tide_DielCx, test = "Chi")

R coding to implement the GLM modeling is included in the scripts with this document, along with an Excel spreadsheet containing the calculations to calculate the z-values required for calculating the p-value indicating statistical difference between variable pairs that do not include the baseline.

While using the GLM to generate a model by which to predict fish presence:absence using the spatial and temporal variables in the dataset, the relationship of the density of fish to these variables is also of interest. An analysis using the magnitude of S_v (non-zero) values was undertaken to gain insights into the relationship of our proxy for fish density (S_v) to the spatial and temporal variables.

Analyses – Fish Density (using S_v as proxy)

To investigate the relationship of the magnitude of relative fish density to the spatial and temporal variables, an analyses of variance (ANOVA) was implemented in R using the non-zero backscattering values of the "20-m dataset" (n=25,536). ANOVA is a robust statistical tool used to test whether there are statistical differences between the means of two or more independent groups. While the calculations for all previous analyses and visualizations were done with the data in its linear form, for the ANOVA analyses the data used in the calculations was the log form (S_v: mean volume backscattering strength) rather than the linear form (s_v: volume backscattering coefficient) (Equation 1). The distribution of the residuals of the linear-form did not approach normality, but the distribution of the residuals of the log-transformed data (S_v) approached normality and therefore more closely conformed to the normality assumption required by ANOVA.

The ANOVA was used to test the mean of the relative fish densities (non-zero S_v values) for statistical differences between the groupings within the predictor variables. The null hypothesis tested was whether the mean of the S_v values were equal as grouped within the predictor variable levels. The test statistic reported for ANOVA, the f-value, provided a statistical measure of whether the mean of each of those group levels were equal. The f-value tends to be greater when the null hypothesis is false. The p-value was used to assess whether the result was statistically significant. Whereas an ANOVA reports whether a significant difference is present, it doesn't report between which groups a significant difference was found. Therefore, a Tukey HSD (honestly significant difference) test was run. This multiple comparison procedure and statistical test is commonly run in conjunction with an ANOVA as a post-hoc analysis to find the means that are significantly different from each other.

To test that the ANOVA result was robust, a permutation test with 10,000 iterations was implemented in R in which the assignment of the S_v values to the categories within an explanatory variable were randomized. The resulting f-values were compared to the f-value garnered from the initial ANOVA run and the resulting p-value computed. The purpose of the permutation test was to evaluate the probability of observing an f-value magnitude equal to or greater than the f-value from the original ANOVA if the grouping labels were randomized. The null hypothesis tested here was: "There is no statistical difference in the group means of the observed S_v values than would be found in the group means if the assignment of the S_v values to the groups were randomized." Therefore, a resulting "significant" probability ($p < 0.05$) indicated that the original groupings of S_v values were not likely to occur by chance and that the ANOVA result from the original test was robust. A probability result (p-value) of 1e-4 (i.e. 0.0001 or 1 in 10,000) indicates that of the 10,000 permutations, in no case was the f-value equal to or greater than the original ANOVA f-value results.

Sample sizes with particularly pronounced imbalances in the count of observations can cause statistical testing to become sensitive to very small and inconsequential differences. Given that categories within some explanatory variables had very pronounced differences in the number of observations, an estimated marginal means (EMM) test was implemented in R. The purpose of calculating EMM was to mitigate the effects of imbalances in the number of observations by estimating what the marginal means would be if the number of observations had been balanced. The EMM were then tested for differences. A test for difference is then carried out with the equally weighted means. "emmeans" as implemented in R uses a confidence level of 0.95 and a significance level (alpha) of 0.05. The EMM model was populated with the model results from the ANOVA. The estimation of the marginal means and the subsequent test for differences will be referred to as the "estimated marginal means test" herein.

Note that while the EMM test addresses the mathematical imbalances, the underlying assumption is that there is no bias implicit in the difference in the amount of time spent sampling: for example, one hour of observations at slack tide (defined as the half hour preceding and following the predicted high or low slack) versus five hours of observations during the periods of flooding or ebbing flow. To test for the presence of bias versus biological differences inherent in the different flow regimes, these analyses can be run again using a selection of one hour's data from the flood and ebb time periods, such as the hour prior to and following the designated slack period and then run again using the hour at peak flow for the ebb and flood flow regime.

For the analyses implemented here, while the mean of the log-transformed data would be difficult to convert into biologically meaningful information, the calculated (or estimated) mean does provide a baseline against which the significance of the differences in relative fish density relationships between the explanatory group levels can be measured. In a log-transformed state, the high outliers (orders of magnitude) are de-emphasized such that the mean of the log-data is lower in magnitude than the mean of the linear-data log transformed (i.e. converted to S_v for reporting). Therefore, the log-means as reported in the tables associated with the fish density analyses were lower than the linear-means reported in the Data Visualizations section of this report.

A compact letter display (CLD) table was generated using the EMM results. For each of the categories within the predictor variable, the CLD tables report the EMMs, standard error, degrees of freedom, lower and upper boundaries of the confidence interval, and a "group number". (Note: although the command to display the table is "cld" (i.e. compact letter display), R uses group numbers rather than group letters.) If a group number appears in one row only, the EMM of the associated category is statistically different from the EMMs of all other categories. If a group number is reported in more than one row, the EMMs of the associated categories do not statistically differ.

The EMM along with bars indicating the confidence interval were plotted. Generally, when comparing two parameter estimates, in this case the EMMs, it is always true that if the confidence intervals do not overlap, then the difference between those statistics (the EMMs) will be significant. However, the converse is not true. That is, one cannot determine the statistical significance of the difference between two statistics based on overlapping confidence intervals. Therefore, to indicate the statistical comparison between the estimated marginal means, the comparison interval was also plotted. Where the comparison intervals overlap, the difference between the EMM was not statistically significant.

To provide a visual representation of the data underlying the ANOVA analyses, the data in their S_v (log) form were plotted as notched boxplots for each category within the explanatory variable examples.

Notched Boxplots

The boxplots included with the results for ANOVA analyses follow the conventions as described in the Exploratory Data Visualizations methods section of this report with the following exceptions. The "notched" form of the boxplot was used, and all calculations were made with the data in its log form (S_v). Relative to the exploratory data visualization boxplots that were calculated using the linear-form data (s_v) and plotted with a log-scale y-axis, calculating the boxplot using the log-form data does not affect the position of the upper and lower boundaries of the box or the placement of the minimum, maximum, or median magnitude, but calculating the mean and the length of whiskers using the log-data does affect their positions.

If you compare the plots calculated in log-form to the plots calculated in linear-form, you'll note that in the linear form there are far more outliers plotted beyond the ends of the whiskers and that the mean of the log-transformed data plots closer to the median. The mean on the notched boxplots are plotted as a star to indicate that the underlying calculation (mean of the log-transformed data) is different than the underlying calculation of the Data Visualizations boxplots for which the mean was calculated using the linear-form data and then transformed to the log form for display (plotted as an open square on those boxplots).

The notches included in the notched boxplots indicate the confidence interval around the median. Although not a formal test, if the notches of two boxes do not overlap it indicates that the plotted data were not from the same or similar populations and there is "strong evidence" (95% confidence interval) that their medians differ. Horizontal dashed lines were added to the boxplots outlining the boundaries of the notches to assist in visualizing the notch overlap or the lack thereof.

Results

The results presented here are examples of data inquiry that can provide information relevant to understanding the historical and contemporary datasets. The data used in these example data inquiries are highly aggregated (e.g. single-level aggregations by study area, survey, tide phase, or diel state). Deeper understanding of the data underlying these examples can be obtained with further exploration as aggregate-level relationships may differ markedly from less highly aggregated data (e.g., survey by tide phase or survey by transect or transect by tide phase).

Data Visualizations

The log-transformed non-zero relative fish density value (S_v) contained in the "20-m dataset" were nearly normal in their distribution with a few high values resulting in a mean (-87.1 dB) slightly greater (~12% in linear terms) than the median (-87.6 dB) (Figure 3). The distribution of the data in its linear form (s_v) does not approach normality (not shown). The mean of the data calculated in its linear form and converted to its log form (S_v) for reporting is -70.1 dB, greater than the median by a factor of 56.

Throughout the Results section, you'll note that outliers tend to be high values rather than low values. While this could be an artifact of setting a minimum threshold for the data integration, the abundance (and therefore density) distribution of marine animals tends to be clustered. Therefore, it might not be unexpected for the distributions of relative fish density to include large numbers of low-density values with proportionately less high-density values resulting in outliers at high values.

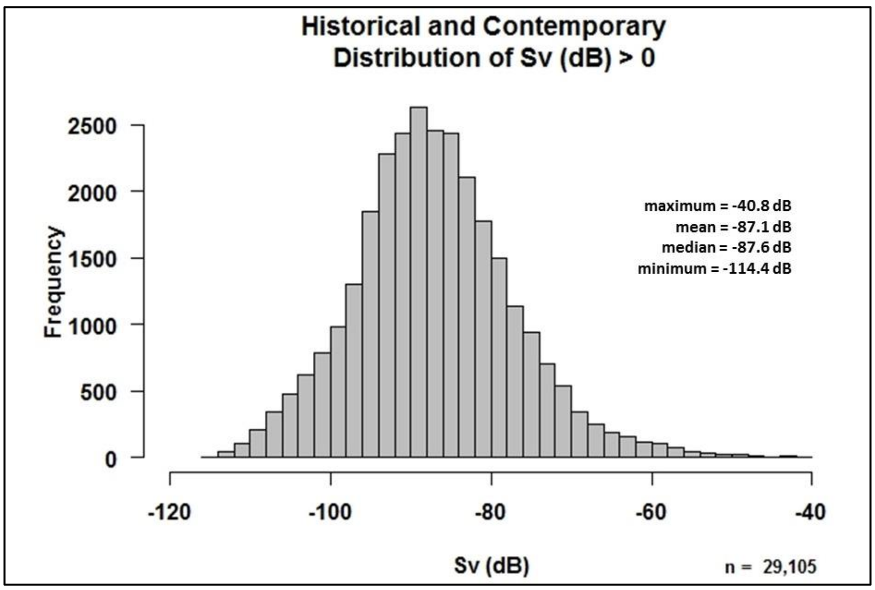


Figure 3. Frequency distribution of non-zero historical and contemporary S_v values*. Data shown are the data exported from Echoview integrated over the full water column binned in 20-m along-shiptrack horizontal bins for the historical and contemporary datasets (excluding zeros).*

Contemporary surveys generally recorded a wider range of S_v values than the historical surveys while possessing generally lower median values (Figure 4 and Figure 5). The distribution statistics for each survey are shown in Table 3.

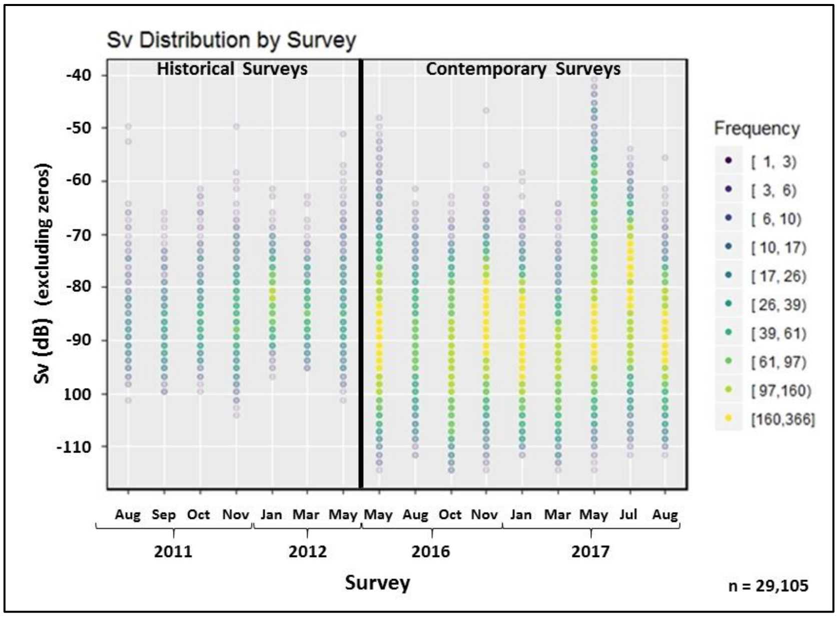


Figure 4. Distribution and frequency of historical and contemporary S_v values by survey*. Data shown are the data exported from Echoview (excluding zeros) integrated over the full water column binned in 20-m along-shiptrack distances for each survey.*

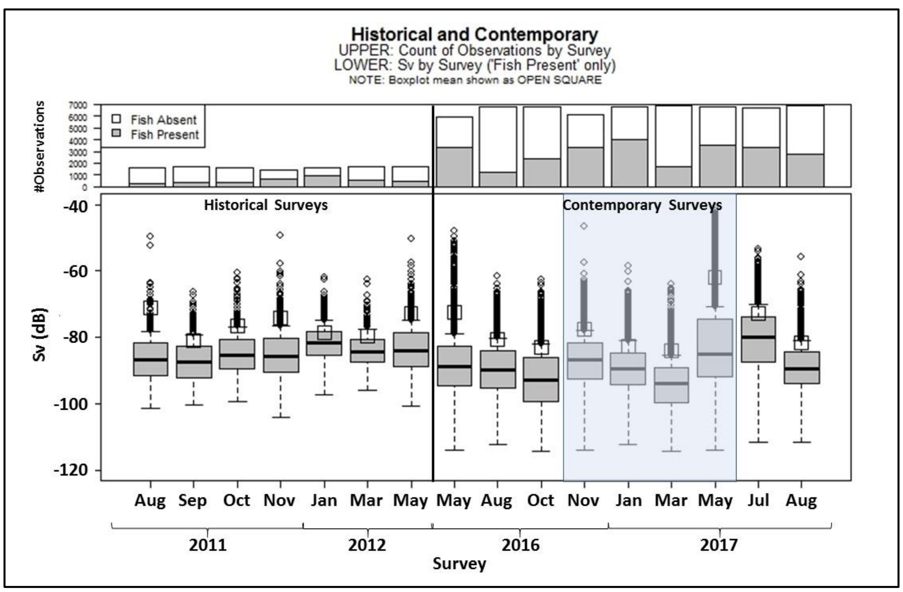


Figure 5. Distribution of historical and contemporary S_v values by survey*. Data shown are the data exported from Echoview integrated over the full water column binned in 20-m along-shiptrack distances for each survey. Data from all transects are aggregated within each survey.* ***Top****: Top of bar indicates the total number of observations for each survey. Axis range: 0 to 7,000. Shaded portion indicates number of non-zero observations (n = 29,105). White portion indicates number of zero observations (n = 41,911).* ***Bottom****: Boxplots of non-zero S_v observations by survey. See text for description of boxplot. Shaded portion of right-hand bottom plot demarcates surveys when a TISEC was in place in the CLA study area.*

***Table 3*. *Distribution statistics for each survey*.** *Statistics were calculated in linear form (s_v) and converted to the log form (S_v) for reporting. Non-zero datapoints only.*

	2011 Aug	2011 Sep	2011 Oct	2011 Nov	2012 Jan	2012 Mar	2012 May	2016 May	2016 Aug	2016 Oct	2016 Nov	2017 Jan	2017 Mar	2017 May	2017 Jul	2017 Aug
n (non-0)	267	343	393	678	911	526	451	3319	1190	2344	3371	3998	1698	3562	3322	2732
maximum	-49.5	-66.4	-60.7	-49.5	-62.0	-62.7	-50.4	-48.1	-61.7	-62.5	-46.5	-58.5	-63.8	-40.8	-53.5	-55.8
median	-86.9	-87.3	-85.5	-85.8	-81.8	-84.5	-84.1	-88.8	-89.7	-92.8	-86.6	-89.4	-94.0	-88.0	-80.0	-89.5
mean	-71.0	-81.1	-76.6	-74.2	-78.6	-79.8	-72.7	-72.5	-80.1	-83.1	-77.7	-82.8	-84.1	-62.0	-72.8	-81.8
minimum	-101.5	-100.4	-99.2	-104.1	-97.4	-95.9	-100.6	-113.8	-112.1	-114.4	-113.9	-112.1	-114.3	-113.8	-111.6	-111.5

Although differing in detail between surveys and the number of observations substantially differing between study areas, at this level of detail the data subsetted by Study Area (CLA and reference) suggests that as the range of S_v values vary from survey to survey in the CLA study area, the reference study area follows generally similar trends (Figure 6a and 6b).

a.

b.

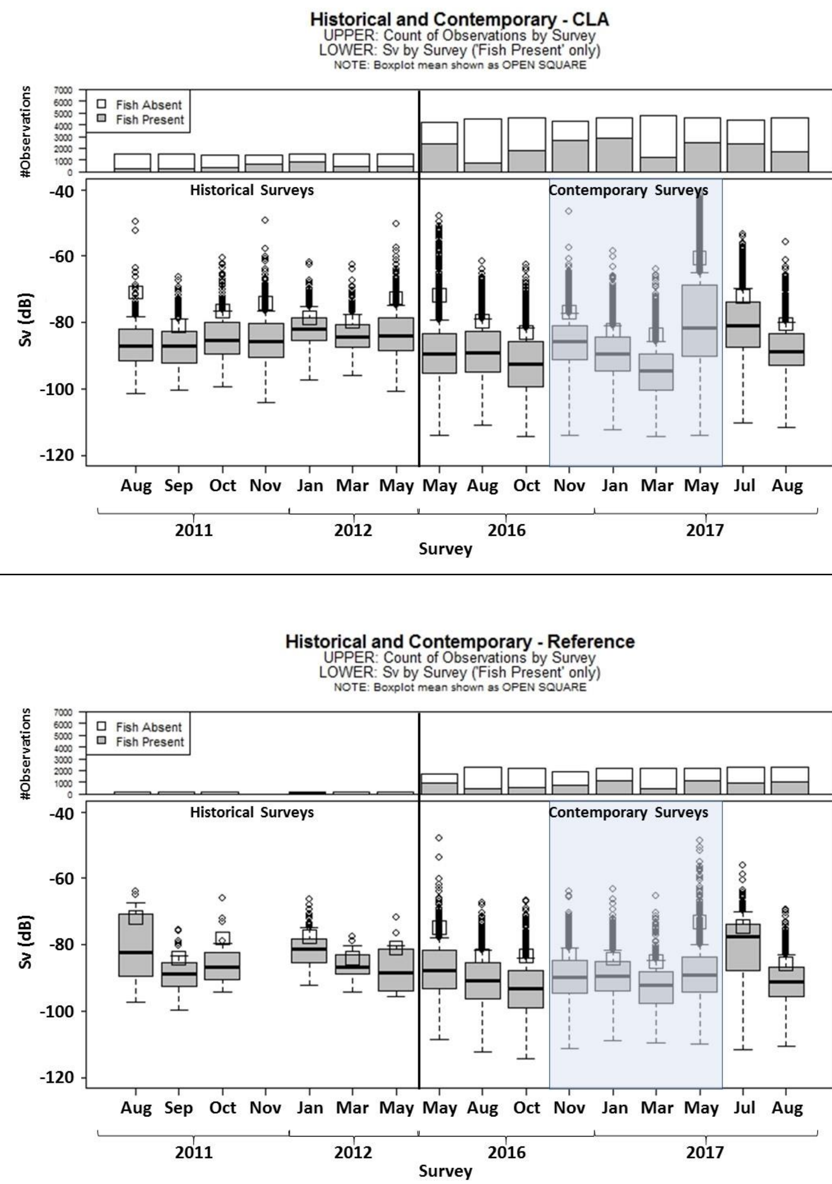


Figure 6. Distribution of historical and contemporary S_v values for the CLA and reference study areas grouped by survey*. Data shown are the data exported from Echoview integrated over the full water column binned in 20-m along-shiptrack distances for each survey. Data from all transects are aggregated within each survey.* ***(a)*** *S_v values observed in the CLA study area.* ***(b)*** *S_v values observed in the reference study area. No data was reported for the reference study area during the historical Nov 2011 survey. (See #6 to Table A1 in Appendix A for more information.)* ***Top****: Top of bar indicates the total number of observations for each survey in the respective study area. Axis range: 0 to 7,000. Shaded portion indicates number of non-zero observations (nCLA = 21,481, nRef = 7,624). White portion indicates number of zero observations (nCLA = 29,211, nRef = 12,700).* ***Bottom****: Boxplots of non-zero S_v observations by survey in the respective study area. See text for description of boxplot. Shaded portion of right-hand bottom plots demarcate surveys when a TISEC was in place in the CLA study area.*

Using the same “20-m dataset” as above, the data were aggregated and boxplots plotted to examine spatial groupings: Research Program (Figure 7a), and Study Area (Figure 7b), and environmental groupings: Tide Phase (Figure 7c), and Diel State (Figure 7d).

The range of backscattering values observed during the contemporary surveys was greater than those observed during the historical surveys (Figure 7a). While the median of the historical dataset was higher than that of the contemporary (Figure 7a), within the contemporary dataset there were a sufficient number of observations which were orders of magnitude greater than its median making the mean for the contemporary dataset higher than for the historical dataset. Using S_v as our proxy for relative fish density, these results indicated that backscatter from higher densities of fish were observed during the contemporary surveys than during the historical surveys. These observations gleaned from the highly aggregated data (by research program) were consistent with the lower-level aggregation by survey (Figure 5) and survey by study area (Figure 6).

There was a wider range of variability of the S_v values, and therefore inferred wider range of relative fish density, in the CLA than in the reference Study Area. In addition to the wider range of fish densities, the CLA Study Area had fish densities higher than that found in the reference Study Area. The differences in the ranges of the integrated mean volume backscattering strength and therefore differences in the ranges of fish densities found in both study areas was approximately a factor of 5.5. (Figure 7b).

The maximum S_v and range of S_v across the ebb, low, and flood tide phases were within less than 2 dB of each other. This indicates generally similar results during each of those tide phases relative to the high slack observations, which exhibited a maximum S_v value and range of S_v an order of magnitude less than the other three tide phases (Figure 7c).

There were much greater differences in the distribution of fish densities across diel states (Figure 7d) than was seen across the tidal phases. Maximum relative density of fish observed during day and night were approximately two orders of magnitude greater than the maximum recorded for dawn or dusk. In spite of those substantially greater maximum fish densities, the mean densities were within approximately one order of magnitude to each other, and the median densities were all within 3 dB (within a factor of two) across the diel states. Therefore, the diel states with the very high fish density observations included sufficient numbers of low fish density observations to reduce the variability in the measures of central tendency (mean and median) across the diel states.

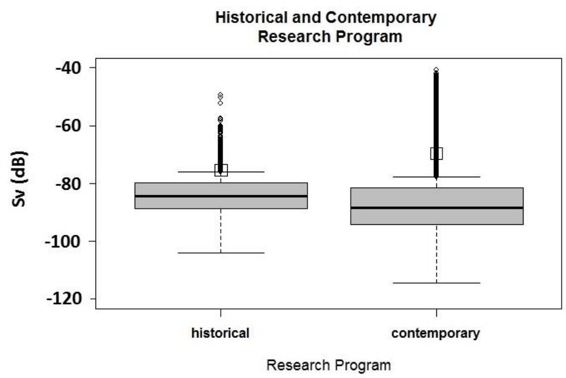
In the visualizations presented here, the data were highly aggregated such that the details of the range of S_v observations by spatial and temporal detail are hidden. The visualization of the data aggregated by diel state therefore, illustrates the value of examining the data in aggregated form (e.g. Figure 7) and the need to examine the data in lower-level aggregations in order to gain insight as to the source of distinctive range of S_v observations such as those at dawn and dusk.

Approximately 600 non-zero data points, each, comprised the dawn and dusk portions of the non-zero dataset whereas the day and night portions included ~14,000 non-zero datapoints each. Therefore, the question arises as to whether the difference in range of S_v observations by diel state was an artifact of sample size or a record of biological behavior. The dawn and dusk sampling periods constituted 4.3% of the total observations (71,016). The non-zero dawn and dusk observations (1,275) plotted here (Figure 7d) at 4.4% of the total non-zero observations (29,105) is proportional to the entire dataset. In other words, the presence:absence ratio within the dawn and dusk time periods is equivalent to the average presence:absence ratio for the entire dataset. While this fact is not sufficient to address the "artifact vs. biology" question, as first cut, it eliminates the question as to whether the dawn and dusk time periods were periods of particularly low fish presence, and thereby possibly periods of particularly low relative fish density.

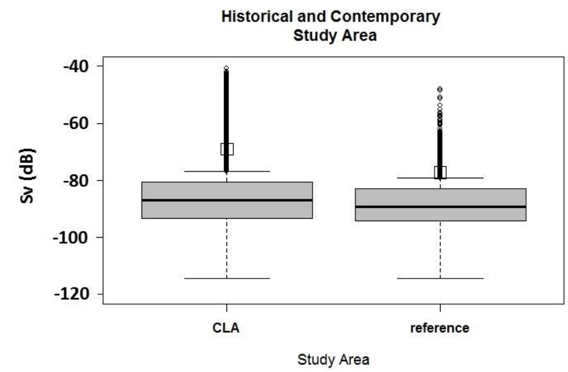
Remembering that non-zero observations encompassed nearly eight orders of magnitude, that the upper end of the dawn and dusk observations peak approximately two orders of magnitude below those for day and night warrants closer inspection. Given that the range of the distribution of S_v observations across tide phases do exhibit such a marked difference, at first glance there is not sufficient evidence to hypothesize that the dawn and dusk sampling periods were predominantly on a particular tide phase. Although the maximum observation during high slack was approximately an order of magnitude lower than the maximum of the other tide phases, the dawn/dusk maximum observations were lower still, by another order of magnitude.

Further exploration of the dawn/dusk data is warranted in order to gain insights as to whether the distinctive dawn and dusk range of S_v observations were an artifact of sampling or a biological signal. Did the sampling regime work out such that lower fish densities would be expected because of the environmental setting during which dawn and dusk sampling was executed? Possibilities to explore are: plotting the dawn/dusk data by survey (i.e. are the data predominantly recorded in a period of lower fish density?) or plotting by transect (i.e. are the data predominantly recorded on particular transects for which lower fish density dominates?) or given that the dawn/dusk time periods are seasonally approximately 30 to 40 minutes long, random sampling the dataset in 30 or 40 minute time-blocks may help shed light on the source of the distinctive ranges of S_v observations recorded during the dawn and dusk time periods.

a



b



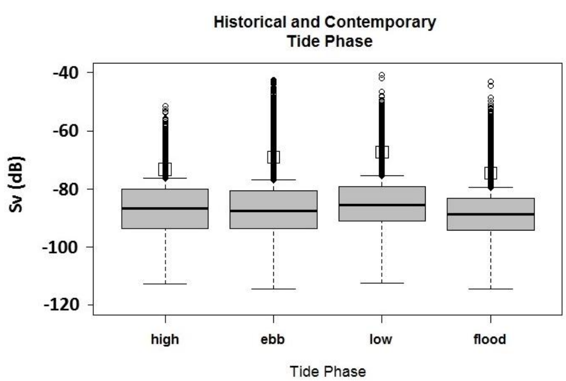
Distribution by Research Program.

	historical		contemporary	
n	3,569		25,536	
max (dB)	-49.5		-40.8	
median (dB)	-84.3		-88.3	
mean (dB)	-75.5		-69.5	
min (dB)	-104.1		-114.4	

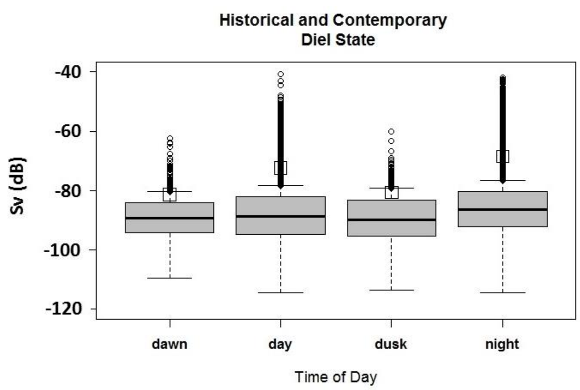
Distribution by Study Area.

	CLA		reference	
n	21,481		7,624	
max (dB)	-40.8		-48.1	
median (dB)	-87.0		-89.2	
mean (dB)	-69.0		-77.3	
min (dB)	-114.4		-114.3	

c



d



Distribution by Tide Phase.

	high	ebb	low	flood
n	2,136	15,285	2,931	8,753
max (dB)	-51.6	-42.4	-40.8	-43.1
median (dB)	-86.8	-87.4	-85.4	-88.7
mean (dB)	-73.2	-69.1	-67.3	-74.6
min (dB)	-112.5	-114.4	-112.2	-114.3

Distribution by Diel State.

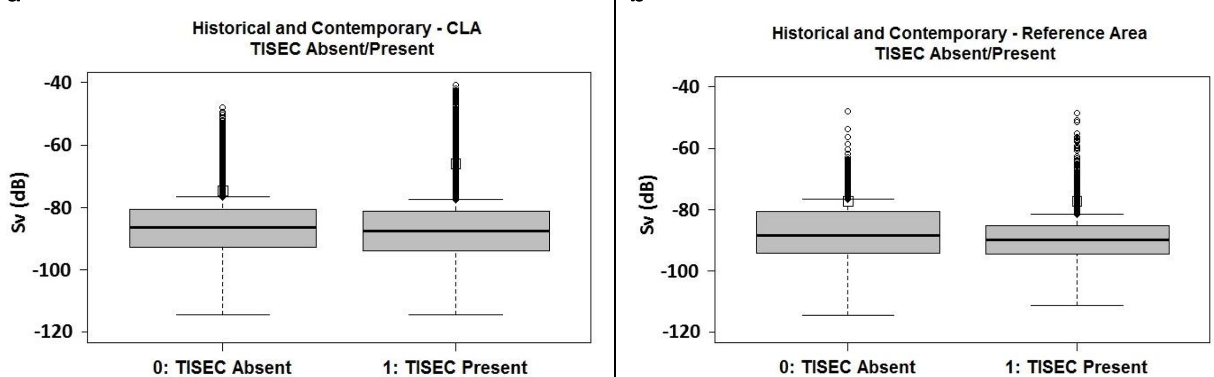
	dawn	day	dusk	night
n	665	13,592	610	14,238
max (dB)	-62.5	-40.8	-60.2	-42.0
median (dB)	-89.4	-88.7	-89.7	-86.4
mean (dB)	-81.4	-72.4	-80.6	-68.4
min (dB)	-109.6	-114.4	-113.5	-114.3

***Figure 7**. **Distribution of historical and contemporary S_v values for spatial and environmental groupings**. **(a)** by Research Program: historical and contemporary, **(b)** by Survey Area: CLA and reference, **(c)** by Tide Phase: ebb, flood, high, and low, **(d)** by Diel State: dawn, day, dusk, and night. Data shown are the data exported from Echoview (excluding zeros) integrated over the full water column binned in 20-m along-shiptrack horizontal bins. See text for description of boxplot.*

In addition to single-level groupings as illustrated in Figure 7, a multi-level grouping example is included here. The multi-level example included data aggregated so as to examine the S_v distribution in both the CLA and reference study areas while a TISEC device was in place in the CLA (Figure 8). Given that no TISEC device was in place during the time periods included in the historical dataset, the data associated with "TISEC Present" were observations from the contemporary dataset.

Although the minimum S_v value, the median, and interquartile range in the CLA site, with and without the presence of a TISEC device, are nearly indistinguishable (Figure 8a), the maximum observed S_v value in the CLA study area was greater by a factor of five while the TISEC device was present. In contrast, in the reference site (Figure 8b) there was a compression of the interquartile range and the maximum observed S_v was 10% less during the same time period. Changes in S_v values can be indicators of change in the aggregation densities of fish or change in the assemblages and therefore, changes in fish sizes passing under the echosounder. To gain insight into the potential influences that may have generated the observed changes in S_v, one must look both to changes in the behavior of the fish present (e.g. aggregating densities) and to the seasonal changes (e.g. fish assemblages or fish sizes) encapsulated by these time periods with and without the TISEC device. Some insights can be gained by examining the underlying data (such as transect-level which is explored further below) and *a priori* information for clues as to possible factors influencing the data at these summary levels.

a **b**



Distribution by Turbine Presence in CLA Area

	TISEC Absent		TISEC Present	
n	12,318		9,163	
max (dB)	-48.1		-40.8	
median (dB)	-86.5		-87.6	
mean (dB)	-74.6		-66.1	
min (dB)	-114.4		-114.3	

Distribution by Turbine Presence in Reference Area

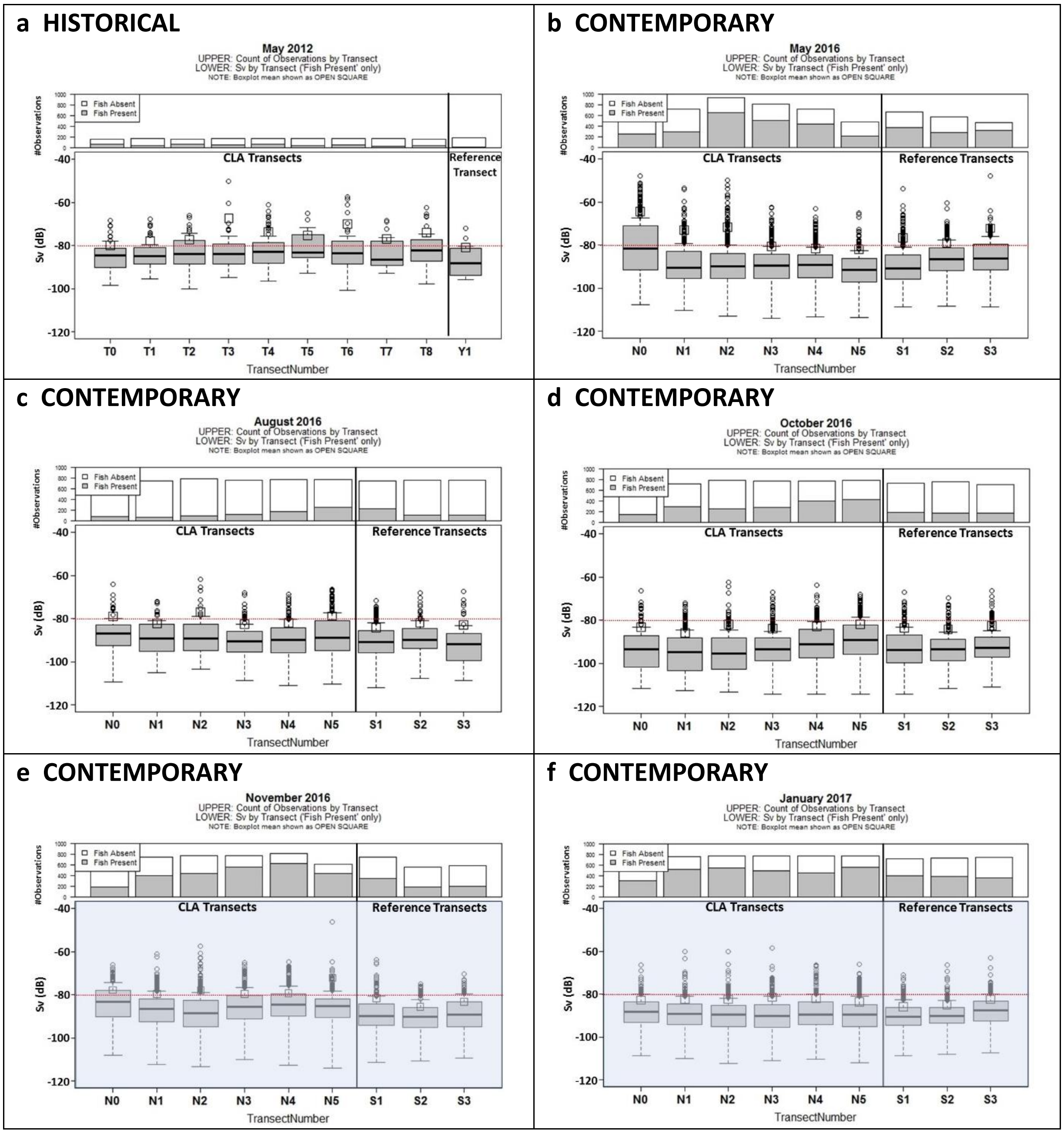
	TISEC Absent		TISEC Present	
n	4,158		3,466	
max (dB)	-48.1		-48.6	
median (dB)	-88.4		-90.0	
mean (dB)	-77.3		-77.2	
min (dB)	-114.3		-111.3	

Figure 8*. *Distribution of Sv values without and with a TISEC emplaced in CLA. *S_v distributions and distribution statistics for both the **(a)** CLA and **(b)** Reference study areas are shown. Data shown are the data exported from Echoview (excluding zeros) integrated over the full water column binned in 20-m along-shiptrack horizontal bins. See text for description of boxplot.*

To examine the underlying data for cues that suggest what data should be examined at finer scales, the distribution of S_v values by survey was scrutinized (Figure 5), as was the distribution of S_v values by survey within the CLA and within the reference study area (Figure 6). It is evident from those data that May 2017 had a strong influence on the maximum values observed in the aggregated “TISEC present in CLA” S_v distribution (Figure 8a). And it is noted that May appears to be a month with a notably wider range of observed S_v values both overall (Figure 5) and within the individual study areas (Figure 6), particularly during the contemporary surveys (May 2016 and May 2017). The S_v distribution for May 2012 in the historical dataset is not so markedly different from the remaining historical surveys as are the May surveys in the contemporary dataset, and certainly not in the reference study area. However, the maximum S_v observed in each of the three May surveys are among the five highest integrated values observed in the entire dataset. Nov 2011 and Nov 2016 exhibit the remaining two magnitudes of integrated S_v that make up the five highest. Aug 2011 is also within the top five, exhibiting a magnitude equivalent to Nov 2011.

Given the pattern of higher mean volume backscattering strength (S_v) and therefore potentially higher fish densities during May surveys, attributing the higher S_v observed during the “TISEC present” phase (Figure 8a) cannot be attributed to the presence of the TISEC without further investigation. Higher fish densities during May surveys as was noted by Daroux and Zydlewski (2017), may have been associated with adult alewife (*Alosa pseudoharengus*) spring spawning migrations and the presence of Atlantic herring (*Clupea harengus*) and striped bass (*Morone saxatilus*) (Baker *et al*., 2014). Striped bass are common in the Minas Passage along the shoreline and they spawn in the head of the tide in May-June (Rulifson and Dadswell, 1995). Spring variation may also be linked to other species migrating into the Basin for the summer, such as Atlantic sturgeon (*Acipenser oxyrhynchus*), American shad (*Alosa sapidissima*), American mackerel (*Scomber scombrus*), and rainbow smelt (*Osmerus mordax*) (Dadswell, 2010).

To place the “TISEC present” surveys in context, and particularly the May 2017 survey, the observed S_v for all of the contemporary surveys plus the May 2012 survey were aggregated by transect and plotted (Figure 9). Although there was variability from transect to transect and survey to survey, of note were the transects nearest to the north shore in the contemporary May surveys (May 2016 and May 2017). In both cases the data for the northern-most transects (N0, N1, N2) suggest a wider range of S_v values to include high values markedly different from other transects within the respective survey and among surveys. The segregation of the high values within those northern-most transects may indicate that the shallower areas are regions of easier fish movement. Within the “TISEC present” surveys (Nov 2016, Jan 2017, Mar 2017, May 2017) the distribution of S_v values for May 2017 stand out as distinctly different in range, magnitude, and spatial distribution from the other “TISEC present” surveys. This evidence suggests that the S_v values aggregated at the coarse scale of “TISEC present-absent” (Figure 8) are likely a function of seasonal and inter-annual variation. May, with its particularly high and wide-range of observed S_v values and apparently distinctive spatial pattern, appears to be an important month for surveying to continue to gather time-series data to help with interpretation.



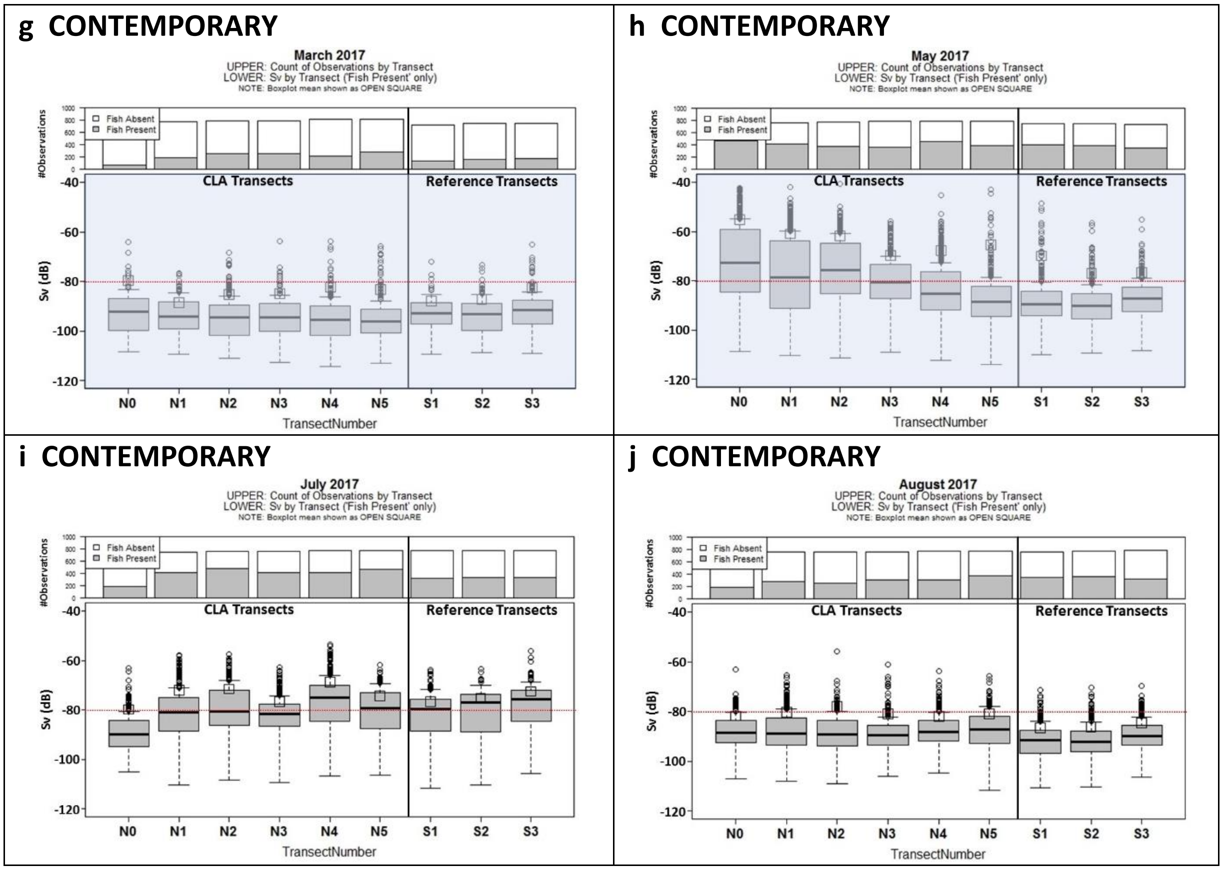


***Figure 9**. **Distribution of Sv values by Transect for Contemporary Surveys and May 2012**. S_v distributions by Transect for **(a)** May 2012, **(b)** May 2016, **(c)** August 2016, **(d)** October 2016, **(e)** November 2016, **(f)** January 2017, **(g)** March 2017, **(h)** May 2017, **(i)** July 2017, and **(j)** August 2017. **Top**: Top of bar indicates the total number of observations for each transect in the respective survey. Axis range: 0 to 1,000. Shaded portion indicates number of non-zero observations. White portion indicates number of zero observations. **Bottom**: Boxplots of non-zero S_v observations by transect in the respective survey. See text for description of boxplot. **Shaded Surveys** (Nov 2016, Jan 2017, Mar 2017, and May 2017) are surveys conducted while a TISEC was in place. **Red Line**: A red line is placed at -80 dB for reference across all surveys.*

Analytical Approach: Fish Presence:Absence

Given the preponderance of zeros in the dataset (59%) and the range of nearly eight orders of magnitude for the non-zero values, two separate analyses were selected to facilitate exploration of the dataset. The first analysis modeled fish presence:absence in relation to the spatial and temporal explanatory variables and is presented in this section, Analytical Approach: Fish Presence:Absence. The second analysis investigated the ranges of relative fish density (S_v values) in relation to the spatial and temporal explanatory variables and is presented in the next section of this report, Analytical Approach: Fish Density. For robust decision-making, we suggest that the results of both analyses (presence:absence and magnitude of relative fish density) be considered together with the characteristics of the underlying data.

Presented here are examples of the results of the modeling of fish presence when implementing the binary logistic model using the GLM command in R. The model was populated with counts of the fish presence:absence observations and a variety of categorical predictor variables. As noted above, the model output includes a table in which the first line of data is reported relative to a 50:50 presence:absence ratio. While a 50:50 ratio is not intrinsically of biological interest, the 50:50 ratio provides a baseline against which the significance of the differences in fish presence relationships to the explanatory variables can be measured.

It should be noted that large sample sizes can cause statistical testing to become sensitive to even very small, inconsequential differences resulting in statistical significance for small and uninteresting effects. The effect is particularly pronounced when there is an imbalance in the count of observations. Therefore, multiple views of the fish presence:absence data is provided in this section to provide the reader with a more robust understanding of the context in which to interpret the meaningfulness of the statistical results. For example, as shown below ("Survey" section) the fish presence:absence ratio for the May 2016 and Nov 2016 surveys was identified as not statistically different (presence: 56.2% and 54.9%, respectively) whereas the ratio for May 2016 and Jan 2017 surveys differed statistically (presence: 56.2% and 58.9%, respectively). In this case, depending on the question at hand, the statistical significance of the differences, 56.2% and 58.9%, may be deemed to be of practical importance or not. On the other hand, cases where differences do not reach the level of statistical significance may provide decisive insights. For example, the presence:absence ratio for three transect pairs (N2:N3, N4:N5, S2:S3) were not statistically different as shown below ("Transect" section). Therefore, if the decision at hand is an issue of cost savings on individual surveys (such as to increase survey frequency without adjusting the survey budget), the statistical results provide some guidance as to consider adjusting the survey design to include one of each pair rather than both.

Note that the underlying data used for these modeling examples are highly aggregated. Aggregate-level relationships of the fish presence:absence ratio to predictor variables such as Research Program, Study Area, Tide Phase, etc. may differ markedly from less highly aggregated data such as by transect within a survey, and thereby the strength of predictors may change. A script containing the R coding to implement the GLM modeling is included with this document along with sufficient commenting to allow one to explore relationships or interactive predictor variables.

GLM Output

For each GLM modeling result, three tables and one figure are presented. A stacked bar plot provides a visualization of the presence ("1") and absence ("0") counts for each category within the explanatory variable. Accompanying the bar plot is a table enumerating those counts and their associated percentages. The table is presented to assist in developing inquiries deeper into the data. Additionally, a table reporting the GLM output and a table summarizing the statistical significance levels are included.

To assist with understanding the GLM output, an explanation of the result table is provided here. The purpose of running the GLM was to produce the coefficients that describe the model of the data, and the parameters by which to quantify the statistical significance of the results. Thereby the GLM output table reports the data needed to *calculate* the estimated y-intercept for each category within the explanatory variable (the “estimate” column) and reports the associated standard error and p-value. A y-intercept of “0” indicates a presence:absence ratio of 50:50. For y-intercepts other than zero, the magnitude of the estimated y-intercept is indicative of how close to, or far from, 50:50 is the presence:absence ratio. The sign of the y-intercept indicates the direction of the ratio: ‘+’ indicates more “present” observations than “absent”, whereas ‘-’ indicates more “absent” observations than “present”.

The category listed in the first row of the table is referred to as the “baseline” category. The data in the remaining rows of the table are relative to the baseline category. The data reported for the baseline category are indicative of that category’s presence:absence ratio relative to 50:50. Therefore the estimate reported for the first row (baseline) is the y-intercept for that category. The estimate values reported in all remaining rows of the table are relative to the baseline. To determine the magnitude of the y-intercept for a category other than the baseline, the estimate reported for that category must be added to the y-intercept reported for the baseline. If the category’s reported estimate has the same sign as the baseline y-intercept, the sum will be greater than that reported for the baseline indicating the presence:absence ratio for the category is in the same direction as that of the baseline, but with a larger difference between presence and absence. Depending on the magnitude of the category’s reported estimate, if the sign is opposite to that of the baseline, the presence:absence ratio for the category may in the same direction as the baseline but closer to 50:50 (i.e. the sum of the y-intercept for the baseline and the category are closer to zero than the baseline y-intercept) or with its opposite sign, if the magnitude of the estimate for the category is greater than that of the baseline, the presence:absence ratio for the category will be reverse that of the presence:absence ratio of the baseline (i.e. greater “present” observations in the category if “absent” observations were greater than “present” in the baseline).

The p-value recorded in the first row of the R GLM-output table reports the statistical significance of the difference from 50:50 that is the fish presence:absence ratio for that category. All other p-values report the statistical significance of the difference of the fish presence:absence ratio for that category relative to the baseline category presence:absence ratio. A two-way table summarizing the p-values reported to two decimal places is also presented. For tables with more than two rows of categories, the statistical significance among the pairs of categories that do not include the baseline were calculated separately and are reported in the summary two-way table. See caption for Table 5 for more information.

Although the ratio of presence:absence for the “20-m dataset” was 41:59, finer-scale aggregations at the category level within the explanatory variables revealed the spatial and temporal categories for which the presence:absence ratio approached 50:50 or for which “absent” observations were exceeded by “present” (Table 4). Further examination of the data is warranted to determine if the reported presence:absence ratios hold at finer scales. For example: the presence:absence ratio for “night” as aggregated over all contemporary surveys was 52:48. Before generalizing this feature, aggregating the diel data by survey will provide the information to confirm whether or not “present” counts exceeded “absent” counts in every case, seasonally, or on some other time scale. Other finer-scale aggregations of data may also provide insights such as diel by transect.

Table 4: Fish Presence:Absence as Percentage by Explanatory Variable Category. *The fish presence ("1") and absence ("0") as percentage of total observations for each category within the explanatory variables included in this report. Column pairs are grouped with decreasing "presence" percentages moving left to right in the table.* ***Black bolded*** *percentages in left-most columns indicate presence:absence ratio ~50:50.* ***Red bolded*** *percentages in left-most columns indicate presence:absence ratio where presence > absence. Full dataset (n = 71,016) was used for research program explanatory variable. All remaining explanatory variables are based on contemporary data only (n = 59,669). See text under "Research Program" below for more information.*

Explanatory Variable	Category					>50%	50%	50%	51%-59%		60%-69%		70%-71%		80%-89%	
		0	1	100%	0	1	0	1	0	1	0	1	0	1	0	1
Research Program	historical	68.5%	31.5%	100.0%							68.5%	31.5%				
	contemporary	57.2%	42.8%	100.0%					57.2%	42.8%						
Study Area	CLA	55.1%	44.9%	100.0%					55.1%	44.9%						
	reference	61.6%	38.4%	100.0%							61.6%	38.4%				
Tide Phase	high	63.0%	37.0%	100.0%							63.0%	37.0%				
	ebb	**50.7%**	**49.3%**	**100.0%**			50.7%	49.3%								
	low	63.2%	36.8%	100.0%							63.2%	36.8%				
	flood	62.2%	37.8%	100.0%							62.2%	37.8%				
Diel State	dawn	54.5%	45.5%	100.0%					54.5%	45.5%						
	day	65.0%	35.0%	100.0%							65.0%	35.0%				
	dusk	62.8%	37.2%	100.0%							62.8%	37.2%				
	night	**47.5%**	**52.5%**	**100.0%**	47.5%	52.5%										
Survey	2016-May	**43.8%**	**56.2%**	**100.0%**	43.8%	56.2%										
	2016-Aug	82.4%	17.6%	100.0%											82.4%	17.6%
	2016-Oct	65.2%	34.8%	100.0%							65.2%	34.8%				
	2016-Nov	**45.1%**	**54.9%**	**100.0%**	45.1%	54.9%										
	2017-Jan	**41.1%**	**58.9%**	**100.0%**	41.1%	58.9%										
	2017-Mar	75.5%	24.5%	100.0%									75.5%	24.5%		
	2017-May	**47.9%**	**52.1%**	**100.0%**	47.9%	52.1%										
	2017-Jul	**50.5%**	**49.5%**	**100.0%**			50.5%	49.5%								
	2017-Aug	60.2%	39.8%	100.0%							60.2%	39.8%				
Transect	N0	69.1%	30.9%	100.0%							69.1%	30.9%				
	N1	57.6%	42.4%	100.0%					57.6%	42.4%						
	N2	53.4%	46.6%	100.0%					53.4%	46.6%						
	N3	53.2%	46.8%	100.0%					53.2%	46.8%						
	N4	**50.4%**	**49.6%**	**100.0%**			50.4%	49.6%								
	N5	**48.5%**	**51.5%**	**100.0%**	48.5%	51.5%										
	S1	58.7%	41.3%	100.0%					58.7%	41.3%						
	S2	63.1%	36.9%	100.0%							63.1%	36.9%				
	S3	63.2%	36.8%	100.0%							63.2%	36.8%				

Research Program

During both research programs (historical and contemporary) counts of fish presence observations were exceeded by counts of fish absence (Table 5 and Table 7). For the baseline category (historical), the deviation of the fish presence:absence ratio from 50:50 was statistically significant ($p < 2e\text{-}16$: Table 5). The fish presence:absence ratio during the contemporary program, although higher than during the historical was still less than 50:50 and was statistically different from the historical fish presence:absence ratio (Table 5).

When considered in conjunction with the S_v distributions within the historical and contemporary datasets (Figures 4 and 5), the data suggest that not only does fish presence differ between the two research programs (historical and contemporary), but also in range of values. This is explored in the next Results section (Analytical Approach: Relative Fish Density).

***Table 5*. *Generalized Linear Model Output of Fish Presence by Research Program*.** *The value in the "Estimate" column for row one indicates the estimate of the y-intercept coefficient for that category. "Zero" would indicate a fish presence:absence ratio of 50:50. A negative value indicates that the count of observations of fish absence was higher than the count of observations of fish presence. The corresponding p-value reported in the last column indicates the statistical significance level of the difference of the observed presence:absence ratio from 50:50 for that category. Any additional rows are reported relative to the category in the first row ("the baseline category"). In this example, we add the "estimate" of the baseline category ("historical" in this case) plus the estimate of the category of interest ("contemporary" is our only choice in this table): -0.77901 + 0.48884 = -0.29017. The resulting negative estimate of the y-intercept for the contemporary category (-0.29017) tells us that the presence:absence ratio for "contemporary" is still such that the number of "absent" observations exceeds the number of "present" observations, but at a lower ratio than the historical (i.e. -0.29017 (contemporary) is closer to zero than -0.77901 (historical)). The p-value reported for contemporary is a measure of the significance of the difference in the presence:absence ratio for that category (contemporary) relative to the baseline category listed in row one (historical). Summary for this table: "historical" presence:absence ratio is statistically different from 50:50 with more "absent" observations than "present" observations. The "contemporary" presence:absence ratio is statistically different than for the "historical" category, although like the historical category, there were more "absent" observations than "present" observations but by a lower ratio. This table description is referenced for all GLM Output tables in this section.*

Coefficients			
	Estimate	**Std. Error**	**Pr(>|z|)**
historical	-0.77901	0.02022	<2e-16
contemporary	+0.48884	0.02185	<2e-16

Table 6. ***Summary Two-way Table of the Statistical Significance of Differences in Pairs of Fish Presence:Absence Ratios*.** *Statistical significance is reported as p-value to two decimal places. Black box with white lettering in column one of row one is the p-value representing the statistical significance of the difference for that variable of the fish presence:absence ratio relative to 50:50. White (statistically significant: p < 0.05) and gray (not statistically significant: p > 0.05) cells contain the p-value quantifying the statistical significance of the difference in fish presence:absence ratios of the two categories (as listed in the corresponding column header and row name). Gray cells not shown here will be present in other two-way tables in this section. When there are only two categories in the table, the p-value quantifying the statistical difference for the second category relative to the first. When there*

are more than two categories in a table, the p-values reporting the statistical significance of the presence:absence ratios between all pairs that did not include the baseline category were calculated as a separate step. Solid black cells along the diagonal and below are cells of the redundant pairs. The two-way has been included for all explanatory variables, including those with only two rows, in order to standardize the reporting for all variables. This table description is referenced for all Summary Two-Way tables in this section.

	historical	**contemporary**
historical	0.00	0.00
contemporary		

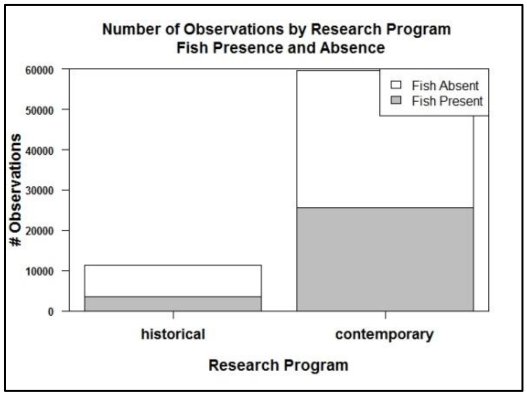


Figure 10. Fish Presence:Absence Observations by Research Program*. Top of bar indicates the total number of observations by research program. Shaded portion indicates the number of observations assigned a value of "1" (i.e. integrated S_v by "20-m" bin ≠ 0). White portion indicates number of observations assigned a value of "0" (i.e. integrated S_v by "20-m" bin = 0). Total number of observations = 71,016. See Table 7 for detailed quantification.*

Table 7*.* ***Research Program Fish Presence:Absence Observations****. Left side of the table contains the count of fish presence:absence observations by category. Right side of the table contains the corresponding percentages. Fish Presence = "0" designates "absence": no observations above the thresholds were observed in the "20-m" data bins (i.e. S_v = 0). Fish Presence = "1" designates "presence": observations above the thresholds were observed in the "20-m" data bins (i.e. S_v is any value other than zero). Where present,* ***bolded black values*** *(none shown here) highlight ratios ~50:50 and* ***bolded red values*** *(none shown here) highlight ratios where the count of "present" observations exceeded the count of "absent" observations.*

Fish Presence	**historical**	**contemporary**	**TOTAL**		**historical**	**contemporary**	**TOTAL**
0	7,778	34,133	41,911		68.5%	57.2%	59.0%
1	3,569	25,536	29,105		31.5%	42.8%	41.0%
TOTAL	11,347	59,669	71,016		100.0%	100.0%	100.0%

In light of the findings suggesting a significant difference in the proportions of fish presence in the contemporary vs. historical datasets further modeling of fish presence by spatial, temporal, and environmental variables was conducted using the ***contemporary dataset only***. In the R scripts included with this document, there is coding that can be used to drill down into the datasets to further explore whether the differences between the historical and contemporary datasets were a function of time or an artifact of the differing characteristics of survey design.

There were substantive differences in survey characteristics between the historical and contemporary surveys. In addition to the survey design characteristics outlined in the Methods section of this report, differences in the execution of the survey design may contribute to differences in the resulting recorded data. For example, there is evidence that suggests that the vessel speed during the historical surveys may have been substantively higher than during the contemporary surveys. At faster vessel speeds, it is possible to miss fish higher in the water column where the acoustic beam is narrower. To test this, the vessel speed can be calculated in Echoview for both the historical and contemporary datasets, and with the known beam widths of the transducers, and ping rates for each survey, one can calculate the depth in the water column where the beam swath overlaps. The volume of water above that depth between pings is not sampled. If that volume substantially differs between the research programs, it may indicate that artifacts due to the execution of the survey may have contributed to the statistically differing results.

Other differences between the research program surveys may have also contributed to differences in the recorded observations of backscatter: differences in the length of transect, single passes over transects during the historical surveys rather than the “with” and “against” passes during the contemporary surveys, or the distribution of observations particularly over the diel state. For more information concerning differing characteristics of survey design and execution, please see Tables A1, A2, A3, and A4 in Appendix A, particularly Table A3.

Study Area

In both study areas (CLA and reference), counts of fish presence observations were exceeded by counts of fish absence (Table 8 and Table 10). For the baseline category (CLA), the deviation of the fish presence:absence ratio from 50:50 was statistically significant ($p < 2e-16$: Table 8). Like the CLA, data collected in the reference Study Area recorded more counts of “absent” than “present”, but at a higher ratio than in the CLA (Table 10). The difference in the fish presence:absence ratio in the reference study area was statistically different from the ratio in the CLA study area ($p < 2e-16$: Table 8).

The number of observations and presence:absence ratios (Table 10) by study area provide a good case study for looking deeper into the question of whether an imbalance in the number of observations is unduly generating a statistically significant difference when the numerical differences are small and inconsequential. There were twice as many observations in the CLA study area than there were in the reference study area (Table 10 and Figure 11) and the difference in the presence:absence ratios were deemed statistically significant (Table 8). It is left to the reader to discern whether the difference in the presence:absence ratios (55:45 and

62:38, respectively) are of practical importance when considering the functioning of the ecosystem within Minas Passage.

***Table 8*. Generalized Linear Model Output of Fish Presence by Study Area – Contemporary Surveys Only**. *See text and the caption for Table 5 for more information.*

Coefficients			
	Estimate	Std. Error	Pr(>|z|)
CLA	-0.20398	0.01001	<2e-16
reference	-0.27042	0.01787	<2e-16

***Table 9*. Summary Two-way Table of the Statistical Significance of Differences in Pairs of Fish Presence:Absence Ratios – Contemporary Surveys Only**. *See caption for Table 6 for more information.*

	CLA	reference
CLA	0.00	0.00
reference		

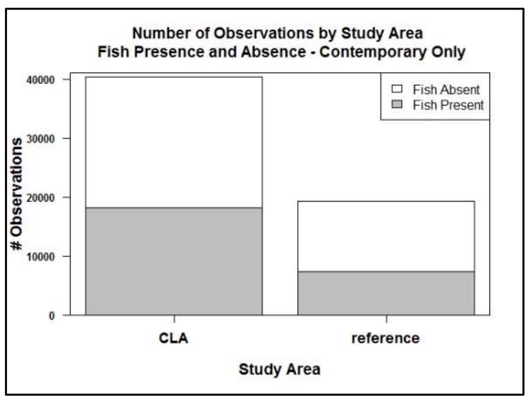


***Figure 11*. Fish Presence:Absence Observations by Study Area – Contemporary Surveys Only**. *Top of bar indicates the total number of observations by study area. Shaded portion indicates the number of observations assigned a value of "1" (i.e. integrated S_v by **"20-m"** bin ≠ 0). White portion indicates number of observations assigned a value of "0" (i.e. integrated S_v by **"20-m"** bin = 0). Total number of observations = 59,669. See Table 10 for detailed quantification.*

***Table 10*. Fish Presence:Absence Observations by Study Area – Contemporary Surveys Only**. *Left side of the table contains the count of fish presence:absence observations by category. Right side of the table contains the corresponding percentages. Fish Presence = "0" designates "absence": no observations above the thresholds were observed in the "20-m" data bins (i.e. S_v = 0). Fish Presence = "1" designates "presence": observations above the thresholds were observed in the "20-m" data bins (i.e. S_v is any value other than zero). Where present, bolded black values highlight ratios ~50:50 and bolded red values highlight ratios where the count of "present" observations exceeded the count of "absent" observations.*

Fish Presence	CLA	reference	TOTAL		CLA	reference	TOTAL
0	22,236	11,897	34,133		55.1%	61.6%	57.2%
1	18,133	7,403	25,536		44.9%	38.4%	42.8%
TOTAL	40,369	19,300	59,669		100.0%	100.0%	100.0%

Tide Phase

The counts of “absent” observations exceeded counts of “present” observations for the high-slack, low-slack, and flooding tide phases, whereas the presence:absence counts for the ebb tide phase was nominally 50:50. The presence:absence ratio for the high-slack tide phase (the “baseline” category in this case) significantly differed from 50:50 (p < 2e-16: Table 11 and Table 13). Among the four tide phases, the presence:absence ratios for the following pairs were not statistically different: high-low (p=0.83), high-flood (p=0.33), flood-low (p=0.43) (Table 12). The presence:absence ratio for the ebb tide phase statistically differed (p=0.00) from each of the other three tide phases: high-ebb, low-ebb, flood-ebb (Table 12).

The presence:absence ratios across the three statistically similar tide phases (high, low, and flood) are surprisingly close (“absent” = 63.0%, 63.2%, and 62.2%) (Table 13). The data, at this highly aggregated level suggest that the **highest probability of observing backscatter from fish occurs during the ebbing tide** when the presence absence ratio is nearly 50:50 (“absent”=50.7%) (Table 13). Before generalizing this finding from the highly aggregated data, exploration of the data aggregated at finer scales should be considered. In addition, if entrained air is particularly evident on specific tides, one must consider the challenges of recording backscatter from fish when entrained air is present in the water column. Further analysis at the detailed level of examining the echograms in Echoview for the influence of entrained air obfuscating backscatter from fish may be required. Included in the script for automating exports from Echoview is coding that exports the depth of the bottom line along the shiptrack as well as the depth of the turbulence line. These can be used to estimate the proportion of the water column lost to entrained air.

***Table 11*. *Generalized Linear Model Output of Fish Presence by Tide Phase – Contemporary Surveys Only*.** *See text and the caption for Table 5 for more information.*

Coefficients			
	Estimate	**Std. Error**	**Pr(>|z|)**
high	-0.53156	0.029431	<2e-16
ebb	+0.50300	0.031852	<2e-16
low	-0.00842	0.038541	0.827
flood	+0.03159	0.032721	<3.3e-01

***Table 12*. *Summary Two-way Table of the Statistical Significance of Differences in Pairs of Fish Presence:Absence Ratios – Contemporary Surveys Only*.** *See caption for **Table 6** for more information.*

	high	**ebb**	**low**	**flood**
high	0.00	0.00	0.83	0.33
ebb			0.00	0.00
low				0.43
flood				

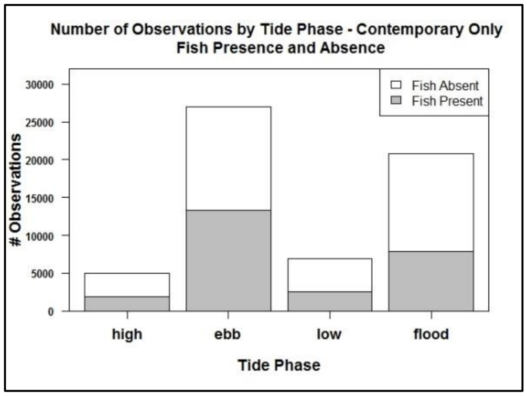


***Figure 12**. **Fish Presence:Absence Observations by Tide Phase – Contemporary Surveys Only**. Top of bar indicates the total number of observations by tide phase. Shaded portion indicates the number of observations assigned a value of "1" (i.e. integrated S_v by **"20-m"** bin ≠ 0). White portion indicates number of observations assigned a value of "0" (i.e. integrated S_v by **"20-m"** bin = 0). Total number of observations = 59,669. See Table 13 for detailed quantification.*

***Table 13**. **Fish Presence:Absence Observations by Tide Phase – Contemporary Surveys Only**. Left side of the table contains the count of fish presence:absence observations by category. Right side of the table contains the corresponding percentages. Fish Presence = "0" designates "absence": no observations above the thresholds were observed in the "20-m" data bins (i.e. S_v = 0). Fish Presence = "1" designates "presence": observations above the thresholds were observed in the "20-m" data bins (i.e. S_v is any value other than zero). Where present, bolded black values highlight ratios ~50:50 and bolded red values highlight ratios where the count of "present" observations exceeded the count of "absent" observations.*

Fish Presence	high	ebb	low	flood	TOTAL		high	ebb	low	flood	TOTAL
0	3,119	13,676	4,386	12,952	34,133		63.0%	**50.7%**	63.2%	62.2%	57.2%
1	1,833	13,291	2,556	7,856	25,536		37.0%	**49.3%**	36.8%	37.8%	42.8%
TOTAL	4,952	26,967	6,942	20,808	59,669		100.0%	**100.0%**	100.0%	100.0%	100.0%

Diel State

Night is the first category for which the count of fish "present" exceeds the count of "absent" (52.5%:47.5%) (Table 16). Counts of "absent" exceeded counts of "present" for the remaining three diel states (dawn: 54.5%:45.5%, day: 65.0%:35.0%, dusk: 62.8%:37.2%) (Table 16). The fish presence:absence ratio for night was statistically different (p = 0.00) than the fish presence:absence ratios for dawn, day, and dusk (Table 15). The fish presence:absence ratios for the following pairs were statistically different: dawn-day (p = 0.00) and dawn-dusk (p = 0.00) (Table 15). The presence:absence ratio for the following pair was not statistically different: dusk-day (p = 0.35, Table 15). The distribution of observations (Table 16) suggest that the **probability of observing fish presence is highest at night (52.5%) and dawn (45.5%)**. Before generalizing the finding of higher fish presence at night, aggregating the diel data by survey will provide information to confirm whether or not "present" counts exceeded "absent" counts in every case, seasonally, or on some other time scale. Other finer-scale aggregations of data may also provide insights such as diel by transect.

***Table 14**. Generalized Linear Model Output of Fish Presence by Diel State – Contemporary Surveys Only*. *See text and the caption for Table 5 for more information.*

Coefficients			
	Estimate	**Std. Error**	**Pr(>|z|)**
dawn	-0.18051	0.05771	0.00176
day	-0.43615	0.05890	1e-13
dusk	-0.34421	0.07962	<2e-05
night	+0.28122	0.05905	<2e-06

***Table 15**. Summary Two-way Table of the Statistical Significance of Differences in Pairs of Fish Presence:Absence Ratios – Contemporary Surveys Only*. *See caption for Table 6 for more information.*

	dawn	**day**	**dusk**	**night**
dawn	0.00	0.00	0.00	0.00
day			0.35	0.00
dusk				0.00
night				

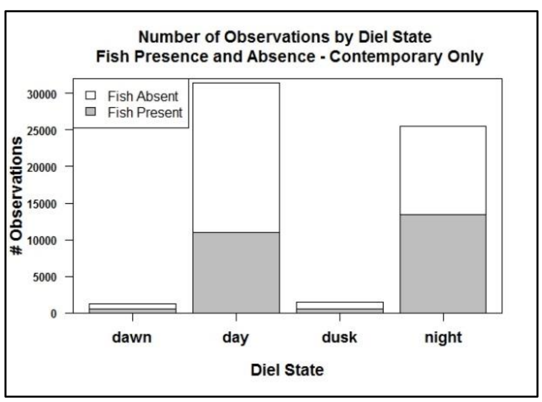


***Figure 13**. Fish Presence:Absence Observations by Diel State – Contemporary Surveys Only*. *Top of bar indicates the total number of observations by diel state. Shaded portion indicates the number of observations assigned a value of "1" (i.e. integrated S_v by **"20-m"** bin ≠ 0). White portion indicates number of observations assigned a value of "0" (i.e. integrated S_v by **"20-m"** bin = 0). Total number of observations = 59,669. See **Table 16** for detailed quantification.*

***Table 16*. Fish Presence:Absence Observations by Diel State – Contemporary Surveys Only.** *Left side of the table contains the count of fish presence:absence observations by category. Right side of the table contains the corresponding percentages. Fish Presence = "0" designates "absence": no observations above the thresholds were observed in the "20-m" data bins (i.e. $S_v = 0$). Fish Presence = "1" designates "presence": observations above the thresholds were observed in the "20-m" data bins (i.e. S_v is any value other than zero). Where present, bolded black values highlight ratios ~50:50 and bolded red values highlight ratios where the count of "present" observations exceeded the count of "absent" observations.*

Fish Presence	dawn	day	dusk	night	TOTAL		dawn	day	dusk	night	TOTAL
0	660	20,443	894	12,136	34,133		54.5%	65.0%	62.8%	**47.5%**	57.2%
1	551	11,034	529	13,422	25,536		45.5%	35.0%	37.2%	**52.5%**	42.8%
TOTAL	1,211	31,477	1,423	25,558	59,669		100.0%	100.0%	100.0%	**100.0%**	100.0%

Survey

There were four surveys for which the count of "present" observations exceeded the count of "absent" observations (May 2016, Nov 2016, Jan 2017, May 2017) and one survey (Jul 2017) for which the fish presence:absence ratio was nominally 50:50 (Table 19b). The presence:absence ratios for all surveys were statistically different from each other ($p < 0.05$) except for one pair of surveys for which the ratios were not statistically different: May 2016-Nov 2016 (p=0.16, Table 18).

The survey results provide a good case study for using ecosystem knowledge and additional analyses to inform interpretations of the statistical results. For example, although the May 2016 and Nov 2016 fish presence:absence ratios are not statistically different (Table 18), the assemblage of fish moving through Minas Passage during May will be very different from the fish assemblage during November. In addition, there are another two survey pairs for which the fish presence:absence ratios were statistically different (Table 18) but for which the p-values were greater than 0.00 unlike all remaining pairing of surveys and therefore may warrant further investigation to inform interpretations of the statistical results: Oct 2016-May 2017 (p=0.03) and May 2017-Jul 2017 (p=0.04). For example, the movement of diadromous fish into and out of Minas Passage in the spring and fall may influence these patterns (Baker *et al*., 2014; Dadswell, 2010; Rulifson and Dadswell, 1995).

The seasonal variation in the fish presence:absence ratio (Table 19b) may suggest that more frequent sampling (such as sequential days) is warranted during those months with higher fish presence in order to increase the likelihood of capturing the extremes of the fish movement.

***Table 17**. **Generalized Linear Model Output of Fish Presence by Survey – Contemporary Surveys Only**. See text and the caption for Table 5 for more information.*

Coefficients			
	Estimate	**Std. Error**	**Pr(>|z|)**
2016 May	+0.24916	0.02623	<2e-16
2016 Aug	-1.79009	0.04133	<2e-16
2016 Oct	-0.87936	0.03663	<2e-16
2016 Nov	-0.05208	0.03668	0.16
2017 Jan	+0.11131	0.03601	0.00
2017 Mar	-1.37565	0.03831	<2e-16
2017 May	-0.16577	0.03569	3e-6
2017 Jul	-0.27002	0.03583	5e-14
2017 Aug	-0.66120	0.03600	<2e-16

***Table 18**. **Summary Two-way Table of the Statistical Significance of Differences in Pairs of Fish Presence:Absence Ratios – Contemporary Surveys Only**. See caption for Table 6 for more information.*

	2016 May	**2016 Aug**	**2016 Oct**	**2016 Nov**	**2017 Jan**	**2017 Mar**	**2017 May**	**2017 Jul**	**2017 Aug**
2016 May	0.00	0.00	0.00	0.16	0.00	0.00	0.00	0.00	0.00
2016 Aug			0.00	0.00	0.00	0.00	0.00	0.00	0.00
2016 Oct				0.00	0.00	0.00	0.03	0.00	0.00
2016 Nov					0.00	0.00	0.00	0.00	0.00
2017 Jan						0.00	0.00	0.00	0.00
2017 Mar							0.00	0.00	0.00
2017 May								0.04	0.00
2017 Jul									0.00
2017 Aug									

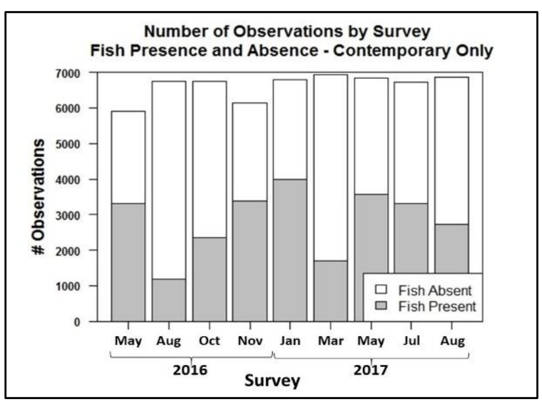


***Figure 14**. **Fish Presence:Absence Observations by Survey – Contemporary Surveys Only**. Top of bar indicates the total number of observations by survey. Shaded portion indicates the number of observations assigned a value of "1" (i.e. integrated S_v by 20-m bin ≠ 0). White portion indicates number of observations assigned a value of "0" (i.e. integrated S_v by 20-m bin = 0). Total number of observations = 59,669. See Table 19 for detailed quantification.*

Table 19. Fish Presence:Absence Observations by Survey – Contemporary Surveys Only. (a) *Count of fish presence:absence observations by category and **(b)** the corresponding percentages. Fish Presence = "0" designates "absence": no observations above the thresholds were observed in the "20-m" data bins (i.e. $S_v = 0$). Fish Presence = "1" designates "presence": observations above the thresholds were observed in the "20-m" data bins (i.e. S_v is any value other than zero). Where present, **bolded black values** highlight ratios ~50:50 and **bolded red values** highlight ratios where the count of "present" observations exceeded the count of "absent" observations.*

(a) Fish Presence	2016 May	2016 Aug	2016 Oct	2016 Nov	2017 Jan	2017 Mar	2017 May	2017 Jul	2017 Aug	TOTAL
0	2,587	5,556	4,402	2,768	2,788	5,238	3,277	3,392	4,125	34,133
1	3,319	1,190	2,344	3,371	3,998	1,698	3,562	3,322	2,732	25,536
TOTAL	5,906	6,746	6,746	6,139	6,786	6,936	6,839	6,714	6,857	59,669

(b) Fish Presence	2016 May	2016 Aug	2016 Oct	2016 Nov	2017 Jan	2017 Mar	2017 May	2017 Jul	2017 Aug	TOTAL
0	**43.8%**	82.4%	65.2%	**45.1%**	**41.1%**	75.5%	**47.9%**	**50.5%**	60.2%	57.2%
1	**56.2%**	17.6%	34.8%	**54.9%**	**58.9%**	24.5%	**52.1%**	**49.5%**	39.8%	42.8%
TOTAL	**100.0%**	100.0%	100.0%	**100.0%**	**100.0%**	100.0%	**100.0%**	**100.0%**	100.0%	100.0%

Transect

There was one transect for which the count of "present" observations exceeded the count of "absent" observations (N5: 51.5%:48:5%) and one transect for which the fish presence:absence ratio was 50:50 (N4) (Table 22b). Three pairs of adjacent transects (N2-N3, N4-N5, S2-S3) were not statistically different from each other ($p > 0.05$) in their fish presence:absence ratios (Table 21). The presence:absence ratio for N0 ($p = 0.00$) was statistically different from all other transects (Table 21). And the presence:absence ratios for one pair of cross-channel transects (S1:N1) were not statistically different ($p = 0.92$, Table 21). The presence:absence ratio of all other transect pairings were statistically different ($p < 0.05$, Table 21). These pairs were explored further in the results on relative fish density is examined later in this Results section.

The S1:N1 pair is of note given their placement across the channel from each other unlike the adjacent placement of all other pairs. The S1:N1 pairing may be of interest from a habitat-use perspective to explore further (e.g. are these transects in similar bathymetric settings such that the pairing provides insight into environmental characteristics that influence fish movement?). Investigation of the adjacent pairings may provide similar insights, but in addition may be of practical significance for survey design (e.g. when circumstances require excluding the time required to survey two transects or if in the interest of increasing survey frequency under the same budget, select one transect from two different pairs rather than two transects from the same pair).

The transect data is highly aggregated, i.e., all data collected for each transect during the first nine surveys of the contemporary research program are grouped for the following analyses. While caution is advised in using these results for decisions, the results from the highly aggregated data identify relationships warranting further investigation for ecosystem understanding.

***Table 20**. **Generalized Linear Model Output of Fish Presence by Transect – Contemporary Surveys Only**. See text and the caption for Table 5 for more information.*

Coefficients			
	Estimate	**Std. Error**	**Pr(>|z|)**
N0	-0.80471	0.02786	<2e-16
N1	+0.49911	0.03723	<2e-16
N2	+0.67069	0.03661	<2e-16
N3	+0.67618	0.03678	<2e-16
N4	+0.78895	0.03674	<2e-16
N5	+0.86519	0.03725	<2e-16
S1	+0.45274	0.03743	<2e-16
S2	+0.26843	0.03803	1e-12
S3	+0.26212	0.03822	7e-12

***Table 21**. **Summary Two-way Table of the Statistical Significance of Differences in Pairs of Fish Presence:Absence Ratios – Contemporary Surveys Only**. See caption for Table 6 for more information.*

	N0	**N1**	**N2**	**N3**	**N4**	**N5**		**S1**	**S2**	**S3**
N0	0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00
N1			0.00	0.00	0.00	0.00		0.37	0.00	0.00
N2				0.92	0.02	0.00		0.00	0.00	0.00
N3					0.03	0.00		0.00	0.00	0.00
N4						0.13		0.00	0.00	0.00
N5								0.00	0.00	0.00
S1									0.00	0.00
S2										0.92
S3										

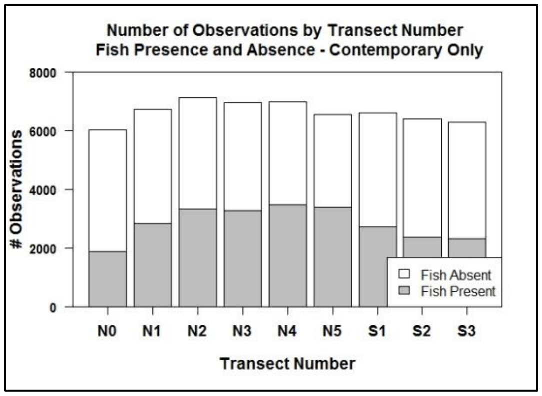


***Figure 15. Fish Presence:Absence Observations by Transect – Contemporary Surveys Only**. Top of bar indicates the total number of observations by transect. Shaded portion indicates the number of observations assigned a value of "1" (i.e. integrated S_v by **"20-m"** bin ≠ 0). White portion indicates number of observations assigned a value of "0" (i.e. integrated S_v by **"20-m"** bin = 0). Total number of observations = 59,669. See Table 22 for detailed quantification.*

***Table 22**. **Fish Presence:Absence Observations by Transect – Contemporary Surveys Only.** **(a)** Count of fish presence:absence observations by category and **(b)** the corresponding percentages. Fish Presence = "0" designates "absence": no observations above the thresholds were observed in the "20-m" data bins (i.e. S_v = 0). Fish Presence = "1" designates "presence": observations above the thresholds were observed in the "20-m" data bins (i.e. S_v is any value other than zero). Where present, **bolded black values** highlight ratios ~50:50 and **bolded red values** highlight ratios where the count of "present" observations exceeded the count of "absent" observations.*

(a) Fish Presence	N0	N1	N2	N3	N4	N5		S1	S2	S3	TOTAL
0	4,168	3,866	3,803	3,706	3,517	3,176		3,876	4,045	3,976	34,133
1	1,864	2,848	3,326	3,259	3,462	3,374		2,726	2,366	2,311	25,536
TOTAL	6,032	6,714	7,129	6,965	6,979	6,550		6,602	6,411	6,287	59,669

(b) Fish Presence	N0	N1	N2	N3	N4	N5		S1	S2	S3	TOTAL
0	69.1%	57.6%	53.4%	53.2%	**50.4%**	**48.5%**		58.7%	63.1%	63.2%	57.2%
1	30.9%	42.4%	46.6%	46.8%	**49.6%**	**51.5%**		41.3%	36.9%	36.8%	42.8%
TOTAL	100.0%	100.0%	100.0%	100.0%	**100.0%**	**100.0%**		100.0%	100.0%	100.0%	100.0%

Example: Modeling with Additive and Interactive Explanatory Variables

Presented here is an example of using an ANOVA with a Chi-Square test to discern whether more complexity in the modeling provides a better fit to the data. For example, to predict fish presence:absence by tide phase or by tide phase and diel state (where tide and diel effects add to one another) or by tide phase * diel state (where tide and diel interact). Reduction of the deviance is an indication of the improvement of the model fit obtained by adding additional terms.

Modeling fish presence:absence using tide phase + diel state (additive) provides a statistically significant better fit to the data ($p<2.2e-16$; Table 23) than by tide phase alone. Modeling fish presence:absence using tide phase * diel state (interaction) provides again, a better fit given that the deviance for the interactive model is less than that of the additive model (Table 23). The fit is a statistically significant better fit ($p<2.2e-16$; Table 23). Therefore, variance in the fish presence:absence data is better explained using the complexity of the interaction of the two explanatory variables. To investigate the influence of the variety of explanatory variables and their additive versus interactive impact, more explanatory variables can be added and the ultimate model could be made very complex.

Table 23*. *Analysis of Deviance Table – Contemporary Surveys Only. *Fish presence:absence during the contemporary surveys (n=59,669)was modeled using a single explanatory variable (Model 1: FishPresence ~ Tide Phase), model with two additive explanatory variables (Model 2: FishPresence ~ Tide Phase + Diel State), and modeled with two interactive explanatory variables (Model 3: FishPresence ~ Tide Phase * Diel State).*

	Resid. Df.	**Resid. Dev**	**Df**	**Deviance**	**Pr(>Chi)**
Model 1	59665	80626			
Model 2	59662	78799	3	1827.66	< 2.2e-16
Model 3	59655	78475	7	324.06	< 2.2e-16

Analytical Approach: Relative Fish Density (as inferred from S_v)

The analysis examples presented in this section investigated the mean of relative fish density in relation to the spatial and temporal explanatory variables. ANOVA, Tukey HSD, and permutation tests were implemented to test for significant differences in the mean of S_v as grouped by explanatory variable categories. An estimated marginal means test was used to mitigate the imbalance of number of observations between the groupings within an explanatory variable and a final ANOVA was then used to test for differences between those means. The log form of backscattering (S_v) was used in these analyses because the distribution of residuals of the log-transformed data more closely approached normality than did the residuals of the data in their linear form.

When using the log-transformed data, the influence of extreme outliers (orders of magnitude) is de-emphasized when calculating the mean relative to using the data in their linear form. In some cases, the apparent sequence of categories within an explanatory variable may shift when ordered by magnitude of the mean (EMM vs. linear mean). The means by Research Program is an example of the re-ordering (Table 26). The goal of this section of analyses was to demonstrate examples by which statistical differences or similarities in relative fish density could be identified over space and time. Although the linear mean has been included in the tables below for ease of comparison, the relationship between categories in terms of relative magnitudes of fish density are more appropriately addressed using the boxplots and tables in the Data Visualization section of this report.

For each explanatory variable example, three tables and two figures are presented. One table reports the f-value and p-value results from the initial ANOVA test and reports the f-value level tested with the permutation tests and the resulting p-value. The second table reports the p-value results from the Tukey HSD test portraying the categories within the explanatory variables for which the difference in the mean were statistically significant. The third table contains the R output from the estimation of the marginal means and their difference testing: the estimated marginal mean, standard error, degrees of freedom, lower confidence level, upper confidence level, and compact letter (number) display groupings. A graph onto which are plotted the estimated marginal mean, the confidence interval (95%), and the comparison range is included, along with notched boxplots providing a visual representation of the data used in these analyses.

To provide additional information, the range of the confidence interval and the number of observations within each grouping were added to the estimated marginal means tables. You'll note that the confidence interval is smaller when the sample size (n) is larger. As was noted with the presence:absence analyses, large sample sizes can cause statistical testing to become sensitive to very small differences resulting in statistical significance for small and uninteresting effects. The sample size effect is particularly pronounced when there is an imbalance in the count of observations. Therefore, the statistical results in this section should not be interpreted in isolation but considered in the context of other analyses in this report and the ecosystem questions at hand to determine the meaningfulness of the results in a biological context.

The notched boxplots were generated using the relative fish density data in its log-transformed state (S_v). The near co-location of the mean and the median on these plots is an indication that the distribution of the log-transformed data approaches normality, whereas the mean, when calculated in its linear form, was not co-located with the median, an indication of the influence of the extreme outliers in the linear data. (See boxplots generated from the data in its linear form in the Data Visualizations section of this report.)

Research Program

The mean of the relative fish density observations during the contemporary surveys differed significantly from the mean of the historical survey data (ANOVA: f=393.8, p=0.00; Table 24). This was corroborated by the permutation test (p=1e-4; Table 24) and the estimated marginal means test for which the results do not group the two categories indicating statistically significant differences in the estimated marginal means (Table 26). The 95% confidence intervals for the medians (Figure 17) and the comparison ranges (Figure 16) of each of the research programs do not overlap, providing additional evidence of differences between the data of the two research programs. The EMM (Table 26) for the historical dataset was higher than that of the contemporary dataset whereas as the inverse was true when evaluating the linear mean (Table 26). In light of the findings of a statistically significant difference in the observations of the relative fish densities in the contemporary vs. historical datasets, the analyses of relative fish density by spatial, temporal, and environmental variables was conducted using the ***contemporary dataset only***.

***Table 24**. **Statistical Results from ANOVA and Permutation Tests for Research Program**. (LEFT): The f-value and p-value reported under the ANOVA heading are the results from the ANOVA when run with non-zero S_v observations from the "20-m dataset". p-value < 0.05 indicates that there was a statistically significant difference in the mean of the S_v values by category within the explanatory variable. The ANOVA does not indicate between which set of means. When there are only two categories within the explanatory variable the two categories with statistically different means is self-evident. For explanatory variables with greater than two categories, finding the pairs of categories with and without statistical differences in the means can be found in the Tukey HSD results table. (RIGHT): The f-value threshold for the permutation test is reported as is the resulting p-value. For the 10,000 permutations of the observed S_v values randomly assigned to the categories within the explanatory variable, the p-value indicates the number of resulting f-values of equal or greater value than that listed as f-value test. A p-value of 1e-4 indicates that no resulting f-values were equal to or greater than the f-value results from the original ANOVA, providing confidence that the original ANOVA results are robust.*

ANOVA			**Permutation Test**	
f-value	**p-value**		**f-value test**	**p-value**
393.8	0.00		>= 393.8	1e-4

***Table 25*. *Tukey HSD Results for Research Program*.** *Whereas the ANOVA is not designed to indicate between which categories the means are or are not statistically significant, the Tukey HSD was implemented to provide that information. For the pair of categories indicated by the column header and the row name, the statistical significance of the difference of the means is reported. Boxes are shaded gray where the p-value is less than 0.05. Boxes are not shaded when the p-value is greater than 0.05. Black boxes are redundant pairs.*

	historical	contemporary
historical		0.00
contemporary		

***Table 26*. *Estimated Marginal Mean and Compact Letter Display for Research Program*.** *Results from the "emmean" computation as reported by R: The computed estimated marginal means (emmean) using the modeled data from the ANOVA are reported, along with the standard error (SE), degrees of freedom (df), lower confidence level (lower.CL), upper confidence level (upper.CL), and grouping (group). If a group number appears in one row only, the estimated marginal mean of the associated category (row name) is statistically different from the estimated marginal means of all other categories. If a group number is repeated in more than one row, the estimated marginal means of the associated categories do not statistically differ. To provide additional information, the range of the confidence interval and the number of observations (n) are included. Significance level used to determine group: alpha = 0.05. Confidence level used: 0.95. The rows are ordered from lowest estimated marginal mean to highest.*

RESEARCH PROGRAM	emmean (dB)	SE (dB)	df	lower.CL (dB)	upper.CL (dB)	group	CL range (dB)	n	min S_v (dB)	max S_v (dB)	mean S_v (linear)
contemporary	-87.53	0.06197	29103	-87.65	-87.41	1	0.24	25,536	-114.4	-40.8	-69.7
historical	-84.02	0.16576	29103	-84.35	-83.70	2	0.65	3,569	-104.1	-49.5	-75.5
TOTAL								**29,105**			

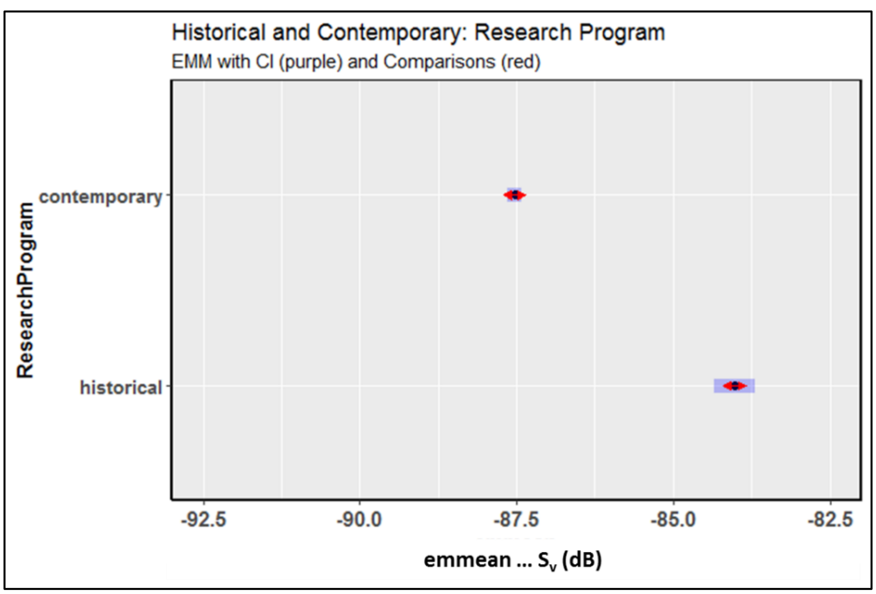


***Figure 16**. **Estimated Marginal Mean with Confidence Interval and Comparisons**. Graph displays results from the estimated marginal mean table. Black dot: estimated marginal mean. Purple bar: range from lower confidence level to upper confidence level (95% confidence interval). Red arrows: comparison range. Where the comparison levels (red arrows) overlap, the difference between the estimated marginal means is not statistically significant. Where the comparison levels (red arrows) do not overlap, the difference between the estimated marginal means is statistically significant. x-axis minimum and maximum has been standardized to encompass the full range necessary for all data reported in this section.*

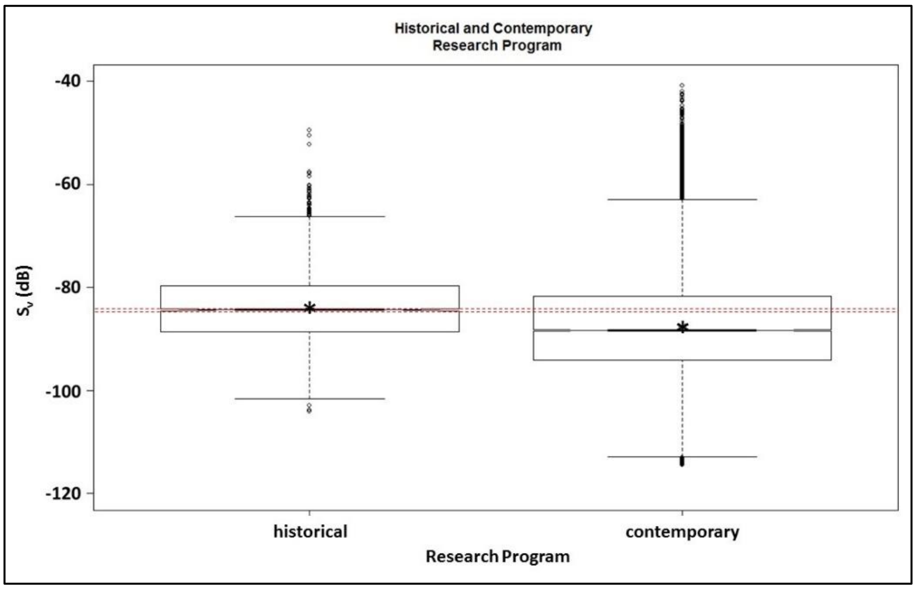


***Figure 17**. **Notched Boxplot for Research Program.** Notched boxplots of non-zero relative fish density values were calculated from "20-m dataset". Boxplot construction was the same as in the Data Visualizations section with the following exceptions. All calculations were done using backscatter in its log form (S_v). Mean of the log values is plotted as star. Median plus the confidence interval around the median is expressed as the notch. While not a formal test, if the notches from two boxes do not overlap it is "strong evidence" (95% confidence interval) that the medians differ. Colored horizontal dashed lines were added to guide the eye between box notches. Where the mean and median overlap, distribution of the data approaches normality. Therefore, note that the distribution of the data much more closely approaches normality when log-transformed (S_v) relative to the data in its linear form (s_v). (See boxplots of the linear data in the Data Visualizations section of this report.) Notched boxplots of similar ranges and similar shapes indicate that the variances are equal. The boxplots are presented in order to provide visual representation of the data used in these analyses.*

Study Area

Generally, relative fish density was greater in the CLA than the reference site. The mean of the relative fish density observations in the reference study area differed significantly from the mean in the CLA study area (ANOVA: f=164.4, p=0.00; Table 27). This was corroborated by the permutation test (p=1e-4; Table 27) and the estimated marginal means test for which the results do not group the two categories indicating statistically significant differences in the estimated marginal means (Table 29). The 95% confidence intervals for the medians (Figure 18) and the comparison ranges (Figure 19) of each of the study areas do not overlap providing additional evidence of differences between the observations in each study area.

***Table 27*. *Statistical Results from ANOVA and Permutation Tests for Study Area – Contemporary Only*.** *(LEFT): The f-value and p-value reported under the ANOVA heading are the results from the ANOVA when run with non-zero S_v observations from the "20-m dataset". p-value < 0.05 indicates that there was a statistically significant difference in the mean of the S_v values by category within the explanatory variable. The ANOVA does not indicate between which set of means. When there are only two categories within the explanatory variable the two categories with statistically different means is self-evident. For explanatory variables with greater than two categories, finding the pairs of categories with and without statistical differences in the means can be found in the Tukey HSD results table. (RIGHT): The f-value threshold for the permutation test is reported as is the resulting p-value. For the 10,000 permutations of the observed S_v values randomly assigned to the categories within the explanatory variable, the p-value indicates the number of resulting f-values of equal or greater value than that listed as f-value test. A p-value of 1e-4 indicates that no resulting f-values were equal to or greater than the f-value results from the original ANOVA, providing confidence that the original ANOVA results are robust.*

ANOVA			Permutation Test	
f-value	**p-value**		**f-value test**	**p-value**
164.4	0.00		>= 164.4	1e-4

***Table 28*. *Tukey HSD Results for Study Area – Contemporary Only*.** *Whereas the ANOVA is not designed to indicate between which categories the means are or are not statistically significant, the Tukey HSD was implemented to provide that information. For the pair of categories indicated by the column header and the row name, the statistical significance of the difference of the means is reported. Boxes are shaded gray where the p-value is less than 0.05. Boxes are not shaded when the p-value is greater than 0.05. Black boxes are redundant pairs.*

	CLA	**reference**
CLA		0.00
reference		

***Table 29*. *Compact Letter Display for Study Area – Contemporary Only*.** *Results from the "emmean" computation as reported by R: The computed estimated marginal means (emmean) using the modeled data from the ANOVA are reported, along with the standard error (SE), degrees of freedom (df), lower confidence level (lower.CL), upper confidence level (upper.CL), and grouping (group). If a group number appears in one row only, the estimated marginal mean of the associated category (row name) is statistically different from the estimated marginal means of all other categories. If a group number is repeated in more than one row, the estimated marginal means of the associated categories do not statistically differ. To provide additional information, the range of the confidence interval and the number of observations (n) are included. Significance level used to determine group: alpha = 0.05. Confidence level used: 0.95. The rows are ordered from lowest estimated marginal mean to highest.*

STUDY AREA	**emmean (dB)**	**SE (dB)**	**df**	**lower.CL (dB)**	**upper.CL (dB)**	**group**	**CL range (dB)**	**n**	**min S_v (dB)**	**max S_v (dB)**	**mean S_v (linear)**
reference	-88.81	0.11870	25534	-89.05	-88.58	1	0.47	7,403	-114.3	-48.1	-77.2
CLA	-87.01	0.07584	25534	-87.16	-86.86	2	0.30	18,133	-114.4	-40.8	-68.4
TOTAL								**25,536**			

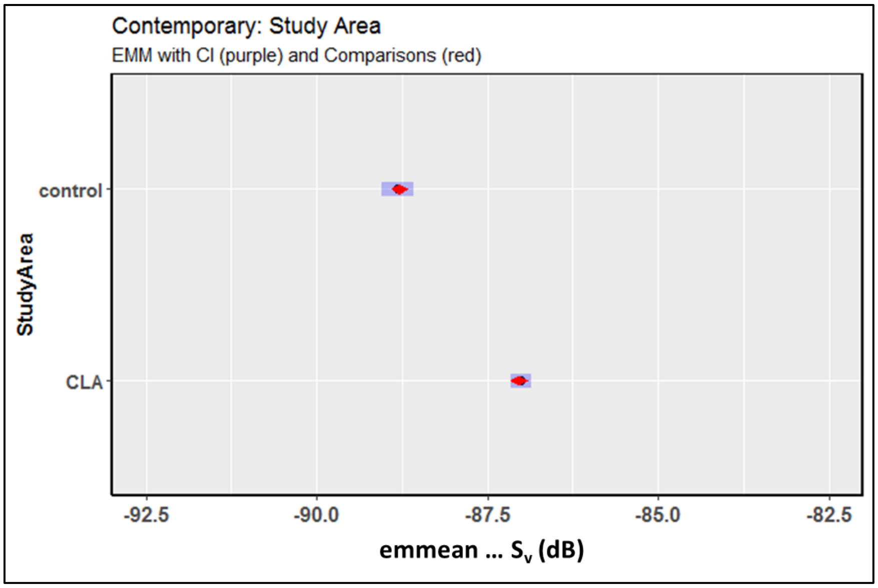


***Figure 18**. **Estimated Marginal Mean with Confidence Interval and Comparisons for Study Area – Contemporary Only**. Graph displays results from the estimated marginal mean table. Black dot: estimated marginal mean. Purple bar: range from lower confidence level to upper confidence level (95% confidence interval). Red arrows: comparison range. Where the comparison levels (red arrows) overlap, the difference between the estimated marginal means is not statistically significant. Where the comparison levels (red arrows) do not overlap, the difference between the estimated marginal means is statistically significant. x-axis minimum and maximum has been standardized to encompass the full range necessary for all data reported in this section.*

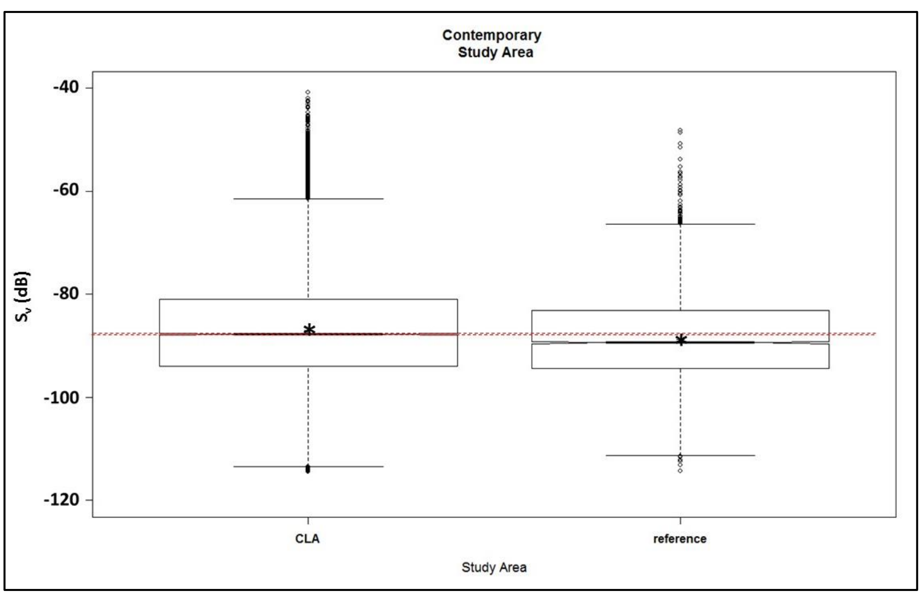


***Figure 19**. **Notched Boxplot for Study Area – Contemporary Only.** Boxplot of non-zero relative fish density data. Boxplot was calculated with the data in its log form (S_v) and are presented in order to provide visual representation of the data used in these analyses. See text and caption for Figure 17 for more information.*

Tide Phase

There was a significant difference between the mean S_v values for at least one pairing of tide phases (ANOVA: f=131.2, p=0.00; Table 30) which was corroborated by the permutation test (p=1e-4; Table 30). The Tukey HSD test elaborated as to which tide phase pairings were found to have statistical differences between their means: high:low, high:flood, ebb:low, ebb:flood, low:flood. The only pairing for which the difference in the mean of the S_v values was not significant was the ebb:high pair (Table 31). These findings were corroborated by the estimated marginal means test for which the results pair ebb:high (Group 2; Table 32). The 95% confidence intervals (boxplot notches) for the medians overlap for the ebb and high-slack tide phases (Figure 21) as do the comparison ranges (Figure 20) whereas the median notches and comparison ranges do overlap for any other tide phase pairings.

As was noted in the Methods section of this report, the purpose of calculating the EMMs was to mitigate the effects of imbalances in the number of observations between categories within an explanatory variable. The underlying assumption in that approach is that there is no bias implicit in the sampling. There may be such a bias in the tide data. No data was collected during low-slack or high-slack along certain transects during the contemporary surveys: S1, S2 and N4, N5, S2, S3 respectively (Figure A3 in Appendix A). All data from those respective transects were excluded from the low-slack and high-slack analyses.

***Table 30*. *Statistical Results from ANOVA and Permutation Tests for Tide Phase – Contemporary Only*.** *(LEFT): The f-value and p-value reported under the ANOVA heading are the results from the ANOVA when run with non-zero S_v observations from the "20-m dataset". p-value < 0.05 indicates that there was a statistically significant difference in the mean of the S_v values by category within the explanatory variable. The ANOVA does not indicate between which set of means. When there are only two categories within the explanatory variable the two categories with statistically different means is self-evident. For explanatory variables with greater than two categories, finding the pairs of categories with and without statistical differences in the means can be found in the Tukey HSD results table. (RIGHT): The f-value threshold for the permutation test is reported as is the resulting p-value. For the 10,000 permutations of the observed S_v values randomly assigned to the categories within the explanatory variable, the p-value indicates the number of resulting f-values of equal or greater value than that listed as f-value test. A p-value of 1e-4 indicates that no resulting f-values were equal to or greater than the f-value results from the original ANOVA, providing confidence that the original ANOVA results are robust.*

ANOVA			**Permutation Test**	
f-value	**p-value**		**f-value test**	**p-value**
131.2	0.00		>= 131.2	1e-4

***Table 31**. **Tukey HSD Results for Tide Phase – Contemporary Only**. Whereas the ANOVA is not designed to indicate between which categories the means are or are not statistically significant, the Tukey HSD was implemented to provide that information. For the pair of categories indicated by the column header and the row name, the statistical significance of the difference of the means is reported. Boxes are shaded gray where the p-value is less than 0.05. Boxes are not shaded when the p-value is greater than 0.05. Black boxes are redundant pairs.*

	high	ebb	low	flood
high		1.00	0.00	0.00
ebb			0.00	0.00
low				0.00
flood				

***Table 32**. **Compact Letter Display for Tide Phase – Contemporary Only**. Results from the "emmean" computation as reported by R: The computed estimated marginal means (emmean) using the modeled data from the ANOVA are reported, along with the standard error (SE), degrees of freedom (df), lower confidence level (lower.CL), upper confidence level (upper.CL), and grouping (group). If a group number appears in one row only, the estimated marginal mean of the associated category (row name) is statistically different from the estimated marginal means of all other categories. If a group number is repeated in more than one row, the estimated marginal means of the associated categories do not statistically differ. To provide additional information, the range of the confidence interval and the number of observations (n) are included. Significance level used to determine group: alpha = 0.05. Confidence level used: 0.95. The rows are ordered from lowest estimated marginal mean to highest.*

TIDE	emmean (dB)	SE (dB)	df	lower.CL (dB)	upper.CL (dB)	group	CL range (dB)	n	min S_v (dB)	max S_v (dB)	mean S_v (linear)
flood	-89.09	0.1147	25532	-89.32	-88.87	1	0.45	7,856	-114.32	-43.11	-74.28
ebb	-87.19	0.0882	25532	-87.37	-87.02	2	0.35	13,291	-114.40	-42.37	-68.65
high	-87.17	0.2375	25532	-87.65	-86.72	2	0.93	1,833	-112.53	-51.59	-72.72
low	-84.75	0.2011	25532	-85.14	-84.35	3	0.79	2,556	-112.25	-40.78	-66.79
TOTAL								**25,536**			

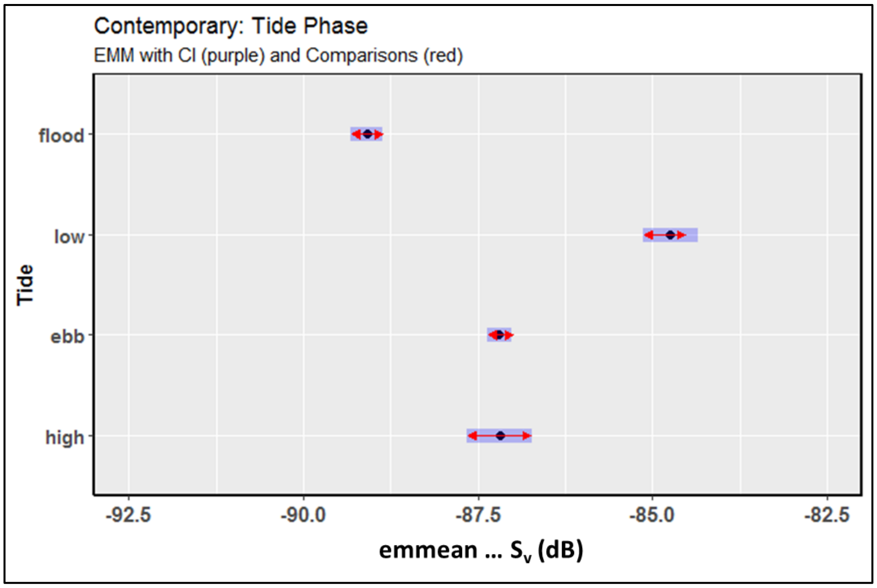


***Figure 20**. **Estimated Marginal Mean with Confidence Interval and Comparisons for Tide Phase – Contemporary Only**. Graph displays results from the estimated marginal mean table. Black dot: estimated marginal mean. Purple bar: range from lower confidence level to upper confidence level (95% confidence interval). Red arrows: comparison range. Where the comparison levels (red arrows) overlap, the difference between the estimated marginal means is not statistically significant. Where the comparison levels (red arrows) do not overlap, the difference between the estimated marginal means is statistically significant. x-axis minimum and maximum has been standardized to encompass the full range necessary for all data reported in this section.*

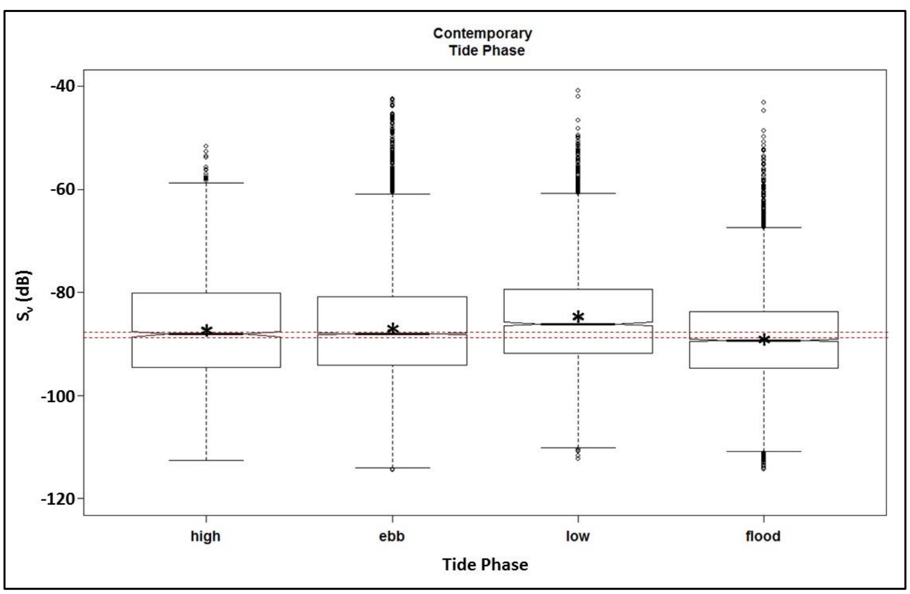


***Figure 21**. **Notched Boxplot for Tide Phase – Contemporary Only.** Boxplot of non-zero relative fish density data. Boxplot was calculated with the data in its log form (S_v) and are presented in order to provide visual representation of the data used in these analyses. See text and caption for Figure 17 for more information.*

Diel State

There was a significant difference between the mean S_v values for at least one of the pairing of the diel states (ANOVA: f=200.5, p=0.00; Table 33) which was corroborated by the permutation test (p=1e-4; Table 33). Pairwise statistical differences between their means: dawn: day, dawn:night, day:night, dusk:night (Table 34, Tukey HSD test). There were two pairings for which the difference of the means were not statistically significant: dawn:dusk, day:dusk (Table 34). These findings were corroborated by the estimated marginal means test for which the results group dawn:dusk (Group 1) and day:dusk (Group 2) (Table 35). The 95% confidence intervals (boxplot notches) for the medians (Figure 23) and the comparison ranges (Figure 22) overlap in the same pattern: dawn:dusk and day:dusk. There were no overlaps for any of the other diel pairings.

As was noted in the Methods section of this report, the purpose of calculating the EMMs was to mitigate the effects of imbalances in the number of observations between categories within an explanatory variable. The underlying assumption in that approach is that there is no bias implicit in the sampling. There may be such a bias in the diel data. No data was collected during dawn or dusk on a low-slack tide during the contemporary surveys (Figure A3 in Appendix A), thereby excluding all low-slack tide measurements from the dawn or dusk analyses. Similarly, certain transects were not traversed during dawn or dusk during the contemporary surveys: N3, N4 and N5, S1, S2 respectively (Figure A3 in Appendix A).

***Table 33**. **Statistical Results from ANOVA and Permutation Tests for Diel State– Contemporary Only**. (LEFT): The f-value and p-value reported under the ANOVA heading are the results from the ANOVA when run with non-zero S_v observations from the "20-m dataset". p-value < 0.05 indicates that there was a statistically significant difference in the mean of the S_v values by category within the explanatory variable. The ANOVA does not indicate between which set of means. When there are only two categories within the explanatory variable the two categories with statistically different means is self-evident. For explanatory variables with greater than two categories, finding the pairs of categories with and without statistical differences in the means can be found in the Tukey HSD results table. (RIGHT): The f-value threshold for the permutation test is reported as is the resulting p-value. For the 10,000 permutations of the observed S_v values randomly assigned to the categories within the explanatory variable, the p-value indicates the number of resulting f-values of equal or greater value than that listed as f-value test. A p-value of 1e-4 indicates that no resulting f-values were equal to or greater than the f-value results from the original ANOVA, providing confidence that the original ANOVA results are robust.*

ANOVA			**Permutation Test**	
f-value	**p-value**		**f-value test**	**p-value**
200.5	0.00		>= 200.5	1e-4

***Table 34*. *Tukey HSD Results for Diel State – Contemporary Only*.** *Whereas the ANOVA is not designed to indicate between which categories the means are or are not statistically significant, the Tukey HSD was implemented to provide that information. For the pair of categories indicated by the column header and the row name, the statistical significance of the difference of the means is reported. Boxes are shaded gray where the p-value is less than 0.05. Boxes are not shaded when the p-value is greater than 0.05. Black boxes are redundant pairs.*

	dawn	day	dusk	night
dawn		0.00	0.47	0.00
day			0.44	0.00
dusk				0.00
night				

***Table 35*. *Compact Letter Display for Diel State – Contemporary Only*.** *Results from the "emmean" computation as reported by R: The computed estimated marginal means (emmean) using the modeled data from the ANOVA are reported, along with the standard error (SE), degrees of freedom (df), lower confidence level (lower.CL), upper confidence level (upper.CL), and grouping (group). If a group number appears in one row only, the estimated marginal mean of the associated category (row name) is statistically different from the estimated marginal means of all other categories. If a group number is repeated in more than one row, the estimated marginal means of the associated categories do not statistically differ. To provide additional information, the range of the confidence interval and the number of observations (n) are included. Significance level used to determine group: alpha = 0.05. Confidence level used: 0.95. The rows are ordered from lowest estimated marginal mean to highest.*

DIEL	emmean (dB)	SE (dB)	df	lower.CL (dB)	upper.CL (dB)	group	CL range (dB)	n	min S_v (dB)	max S_v (dB)	mean S_v (linear)
dawn	-90.61	0.43144	25532	-91.46	-89.77	1	1.69	551	-109.6	-62.5	-82.1
dusk	-89.72	0.44032	25532	-90.59	-88.86	12	1.73	529	-113.5	-63.2	-81.6
day	-89.05	0.09641	25532	-89.24	-88.86	2	0.38	11,034	-114.4	-40.8	-72.0
night	-86.07	0.08742	25532	-86.24	-85.90	3	0.34	13,422	-114.3	-42.0	-68.2
TOTAL								**25,536**			

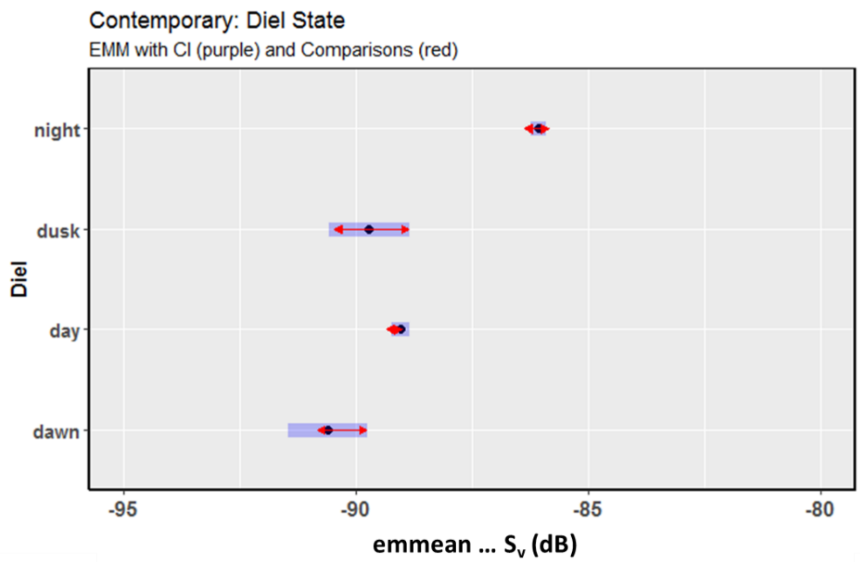


***Figure 22**. **Estimated Marginal Mean with Confidence Interval and Comparisons for Diel State – Contemporary Only**. Graph displays results from the estimated marginal mean table. Black dot: estimated marginal mean. Purple bar: range from lower confidence level to upper confidence level (95% confidence interval). Red arrows: comparison range. Where the comparison levels (red arrows) overlap, the difference between the estimated marginal means is not statistically significant. Where the comparison levels (red arrows) do not overlap, the difference between the estimated marginal means is statistically significant. x-axis minimum and maximum has been standardized to encompass the full range necessary for all data reported in this section.*

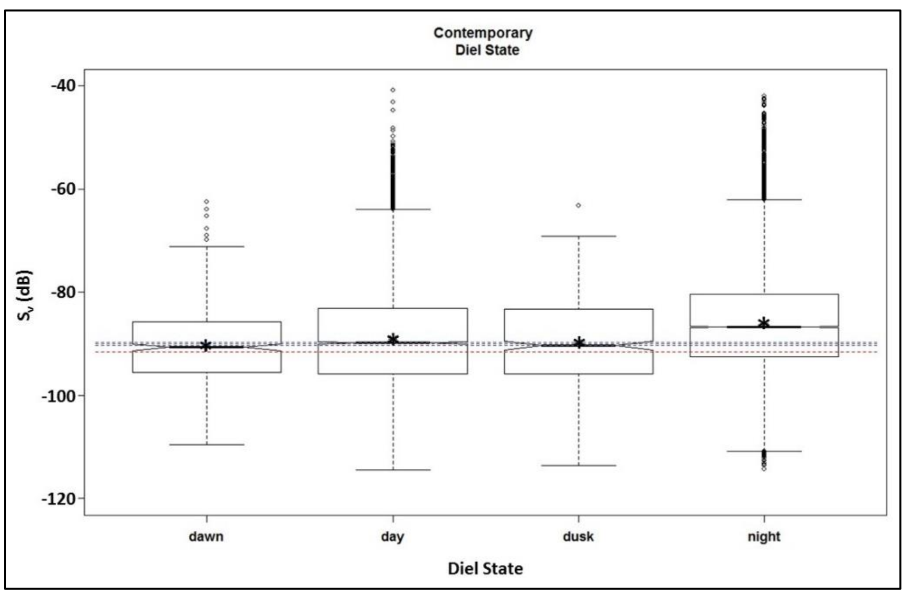


***Figure 23**. **Notched Boxplot for Diel State – Contemporary Only.** Boxplot of non-zero relative fish density data. Boxplot was calculated with the data in its log form (S_v) and are presented in order to provide visual representation of the data used in these analyses. See text and caption for Figure 17 for more information.*

Survey

There was a statistically significant difference between the mean of the S_v values for at least one pairing of the surveys (ANOVA: f=575.2, p=0.00 Table 36) corroborated by the permutation test (p=1e-4; Table 36). All survey pairings were found to have statistical differences between their means, except for three pairings for which the difference in the survey means was not statistically significant: Aug 2016:Jan 2017, Aug 2016:Aug 2017, Jan 2017:Aug 2017 (Table 37). These findings were corroborated by the estimated marginal means test for which the results group the Aug 2016, Jan 2017, and Aug 2017 surveys (Group=3, Table 38) and shown by the overlap in the comparison ranges (Figure 24). Whereas the analysis using the means and the estimated marginal means grouped the three surveys together, the 95% confidence intervals around the median suggest that May 2016 may be included in the grouping established by the means (Aug 2016, Jan 2017, and Aug 2017; Figure 25).

***Table 36*. *Statistical Results from ANOVA and Permutation Tests for Survey – Contemporary Only*.** *(LEFT): The f-value and p-value reported under the ANOVA heading are the results from the ANOVA when run with non-zero S_v observations from the "20-m dataset". p-value < 0.05 indicates that there was a statistically significant difference in the mean of the S_v values by category within the explanatory variable. The ANOVA does not indicate between which set of means. When there are only two categories within the explanatory variable the two categories with statistically different means is self-evident. For explanatory variables with greater than two categories, finding the pairs of categories with and without statistical differences in the means can be found in the Tukey HSD results table. (RIGHT): The f-value threshold for the permutation test is reported as is the resulting p-value. For the 10,000 permutations of the observed S_v values randomly assigned to the categories within the explanatory variable, the p-value indicates the number of resulting f-values of equal or greater value than that listed as f-value test. A p-value of 1e-4 indicates that no resulting f-values were equal to or greater than the f-value results from the original ANOVA, providing confidence that the original ANOVA results are robust.*

ANOVA			Permutation Test	
f-value	**p-value**		**f-value test**	**p-value**
575.2	0.00		>= 575.2	1e-4

Table 37*. *Tukey HSD Results for Survey – Contemporary Only. *Whereas the ANOVA is not designed to indicate between which categories the means are or are not statistically significant, the Tukey HSD was implemented to provide that information. For the pair of categories indicated by the column header and the row name, the statistical significance of the difference of the means is reported. Boxes are shaded gray where the p-value is less than 0.05. Boxes are not shaded when the p-value is greater than 0.05. Black boxes are redundant pairs.*

	2016 May	2016 Aug	2016 Oct	2016 Nov	2017 Jan	2017 Mar	2017 May	2017 Jul	2017 Aug
2016 May		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2016 Aug			0.00	0.00	1.00	0.00	0.00	0.00	1.00
2016 Oct				0.00	0.00	0.00	0.00	0.00	0.00
2016 Nov					0.00	0.00	0.00	0.00	0.00
2017 Jan						0.00	0.00	0.00	0.87
2017 Mar							0.00	0.00	0.00
2017 May								0.00	0.00
2017 Jul									0.00
2017 Aug									

Table 38. Compact Letter Display for Survey – Contemporary Only*. Results from the "emmean" computation as reported by R: The computed estimated marginal means (emmean) using the modeled data from the ANOVA are reported, along with the standard error (SE), degrees of freedom (df), lower confidence level (lower.CL), upper confidence level (upper.CL), and grouping (group). If a group number appears in one row only, the estimated marginal mean of the associated category (row name) is statistically different from the estimated marginal means of all other categories. If a group number is repeated in more than one row, the estimated marginal means of the associated categories do not statistically differ. To provide additional information, the range of the confidence interval and the number of observations (n) are included. Significance level used to determine group: alpha = 0.05. Confidence level used: 0.95. The rows are ordered from lowest estimated marginal mean to highest.*

SURVEY	emmean (dB)	SE (dB)	df	lower.CL (dB)	upper.CL (dB)	group	CL range (dB)	n	min S_v (dB)	max S_v (dB)	mean S_v (linear)
2017 Mar	-94.01	0.2289	25527	-95.54	-93.64	1	1.90	1,698	-114.3	-63.8	-84.1
2016 Oct	-92.55	0.1948	25527	-92.93	-92.16	2	0.77	2,344	-114.4	-62.5	-83.1
2017 Jan	-89.74	0.1492	25527	-90.04	-89.45	3	0.59	3,998	-112.1	-58.5	-82.8
2016 Aug	-89.59	0.2734	25527	-90.13	-89.05	3	1.08	1,190	-112.1	-61.7	-80.7
2017 Aug	-89.40	0.1805	25527	-89.75	-89.05	3	0.70	2,732	-111.6	-55.8	-81.8
2016 May	-88.36	0.1637	25527	-88.68	-88.04	4	0.64	3,319	-113.8	-48.1	-72.5
2016 Nov	-87.03	0.1625	25527	-87.35	-86.71	5	0.64	3,371	-113.9	-46.5	-77.7
2017 May	-82.39	0.1580	25527	-82.70	-82.08	6	0.62	3,562	-113.8	-40.8	-62.0
2017 Jul	-80.90	0.1636	25527	-81.22	-80.58	7	0.64	3,322	-111.6	-53.5	-72.8
TOTAL								25,536			

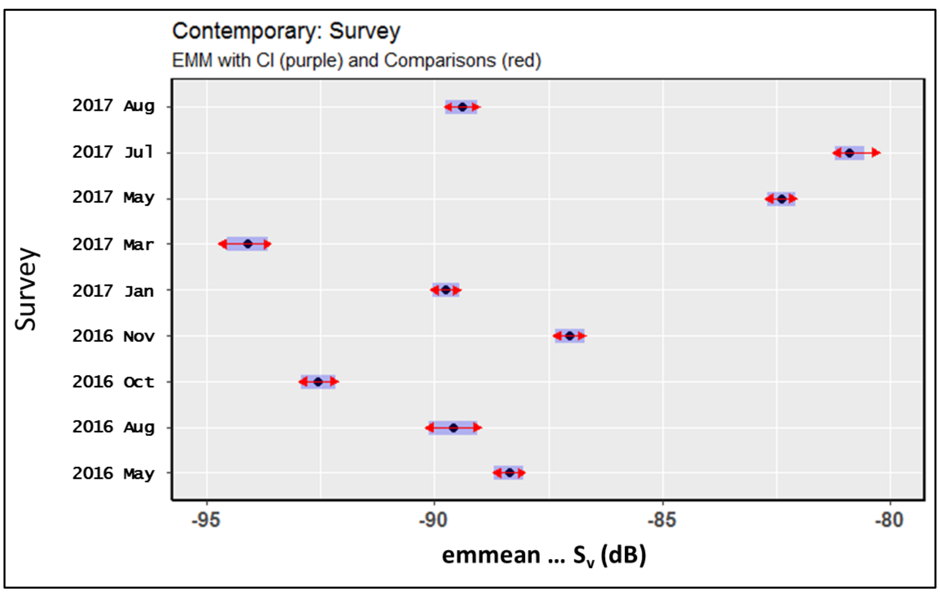


Figure 24. Estimated Marginal Mean with Confidence Interval and Comparisons by Survey – Contemporary Only*. Graph displays results from the estimated marginal mean table. Black dot: estimated marginal mean. Purple bar: range from lower confidence level to upper confidence level (95% confidence interval). Red arrows: comparison range. Where the comparison levels (red arrows) overlap, the difference between the estimated marginal means is not statistically significant. Where the comparison levels (red arrows) do not overlap, the difference between the estimated marginal means is statistically significant. x-axis minimum and maximum has been standardized to encompass the full range necessary for all data reported in this section.*

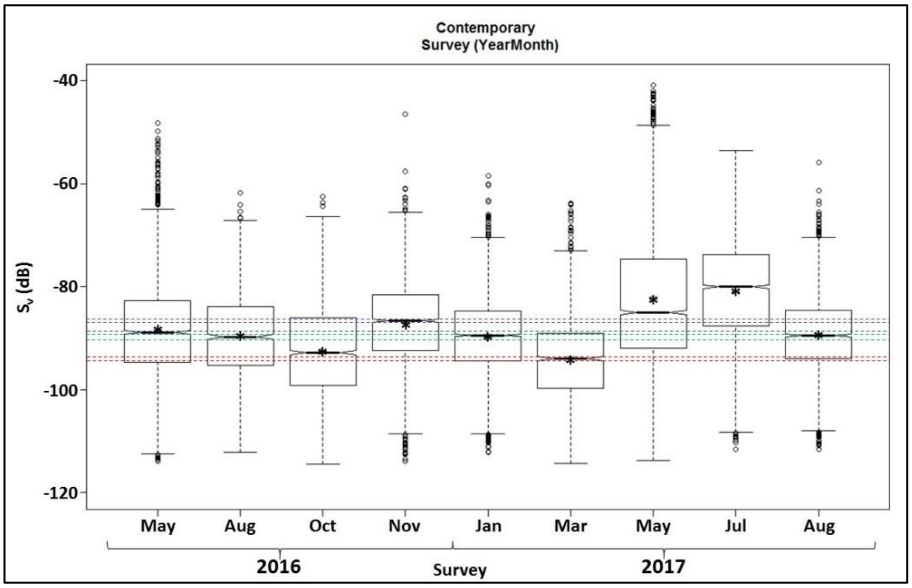


Figure 25*. *Notched Boxplot by Survey – Contemporary Only. *Boxplot of non-zero relative fish density data. Boxplot was calculated with the data in its log form (S_v) and are presented in order to provide visual representation of the data used in these analyses. See text and caption for Figure 17 for more information.*

Transect

There was a statistically significant difference between the mean of the S_v values for at least one pairing of the transects (ANOVA: f=61.9, p=0.00 Table 39) corroborated by the permutation test (p=1e-4; Table 39). The ANOVA f-value of 61.9 is the lowest of, and substantially lower than, the f-value results from the ANOVA tests for the variety of explantory variable examples included in these analyses. Of the 36 possible transect pairings, 11 pairings were not found to have statistical differences between their means: N1:N2, N1:N4, N1:S3, N2:N3, N2:N4, N2:S3, N3:N5, N3:S3, N4:S3, N5:S3, S1:S2 (Table 40). As with the fish presence:absence analysis, transect N0 alone is statistically different than all other transects (Table 40). The means for all remaining possible transect pairs were statistically different. These findings were corroborated by the estimated marginal means test for which the results group the following transect pairs: S1:S2 (Group 1), N3:N5:S3 (Group 2), N2:N3:S3 (Group 3), N1:N2:N4:S3 (Group 4) with the estimated marginal mean for transect N0 statistically different than the estimated marginal mean for all other transects (Table 41). The 95% confidence intervals around the medians suggest two groupings and two singular transects: S1:S2, N1:N2:N3:N5:S3, and N0 and N4 respectively (Figure 27).

***Table 39*. Statistical Results from ANOVA and Permutation Tests by Transect – Contemporary Only**. *(LEFT): The f-value and p-value reported under the ANOVA heading are the results from the ANOVA when run with non-zero S_v observations from the "20-m dataset". p-value < 0.05 indicates that there was a statistically significant difference in the mean of the S_v values by category within the explanatory variable. The ANOVA does not indicate between which set of means. When there are only two categories within the explanatory variable the two categories with statistically different means is self-evident. For explanatory variables with greater than two categories, finding the pairs of categories with and without statistical differences in the means can be found in the Tukey HSD results table. (RIGHT): The f-value threshold for the permutation test is reported as is the resulting p-value. For the 10,000 permutations of the observed S_v values randomly assigned to the categories within the explanatory variable, the p-value indicates the number of resulting f-values of equal or greater value than that listed as f-value test. A p-value of 1e-4 indicates that no resulting f-values were equal to or greater than the f-value results from the original ANOVA, providing confidence that the original ANOVA results are robust.*

ANOVA			**Permutation Test**	
f-value	**p-value**		**f-value test**	**p-value**
61.9	0.00		>= 61.9	1e-4

***Table 40*. Tukey HSD Results by Transect – Contemporary Only**. *Whereas the ANOVA is not designed to indicate between which categories the means are or are not statistically significant, the Tukey HSD was implemented to provide that information. For the pair of categories indicated by the column header and the row name, the statistical significance of the difference of the means is reported. Boxes are shaded gray where the p-value is less than 0.05. Boxes are not shaded when the p-value is greater than 0.05. Black boxes are redundant pairs.*

	N0	**N1**	**N2**	**N3**	**N4**	**N5**		**S1**	**S2**	**S3**
N0		0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00
N1			0.98	0.04	1.00	0.00		0.00	0.00	0.69
N2				0.40	0.86	0.01		0.00	0.00	1.00
N3					0.01	0.94		0.00	0.00	0.97
N4						0.01		0.00	0.00	0.41
N5								0.00	0.00	0.32
S1									0.96	0.00
S2										0.00
S3										

***Table 41*. *Compact Letter Display by Transect – Contemporary Only*.** *Results from the "emmean" computation as reported by R: The computed estimated marginal means (emmean) using the modeled data from the ANOVA are reported, along with the standard error (SE), degrees of freedom (df), lower confidence level (lower.CL), upper confidence level (upper.CL), and grouping (group). If a group number appears in one row only, the estimated marginal mean of the associated category (row name) is statistically different from the estimated marginal means of all other categories. If a group number is repeated in more than one row, the estimated marginal means of the associated categories do not statistically differ. To provide additional information, the range of the confidence interval and the number of observations (n) are included. Significance level used to determine group: alpha = 0.05. Confidence level used: 0.95. The rows are ordered from lowest estimated marginal mean to highest.*

STUDY AREA	emmean (dB)	SE (dB)	df	lower.CL (dB)	upper.CL (dB)	group	CL range (dB)	n	min S_v (dB)	max S_v (dB)	mean S_v (linear)
S1	-89.59	0.1944	25527	-89.97	-89.21	1	0.76	2,726	-114.3	-48.6	-76.4
S2	-89.25	0.2086	25527	-89.66	-88.84	1	0.82	2,366	-111.5	-56.6	-79.7
N5	-88.10	0.1747	25527	-88.45	-87.76	2	0.69	3,374	-114.2	-43.1	-73.1
N3	-87.78	0.1778	25527	-88.13	-87.44	23	0.69	3,259	-114.1	-56.0	-76.9
S3	-87.46	0.2111	25527	-87.88	-87.05	234	0.83	2,311	-111.1	-48.1	-76.5
N2	-87.23	0.1760	25527	-87.58	-86.89	34	0.69	3,326	-113.4	-40.8	-69.9
N1	-86.95	0.1902	25527	-87.33	-86.58	4	0.75	2,848	-112.5	-42.0	-68.7
N4	-86.87	0.1725	25527	-87.20	-86.53	4	0.67	3,462	-114.4	-45.4	-73.7
N0	-83.63	0.2351	25527	-84.09	-83.17	5	0.92	1,864	-111.5	-42.4	-60.9
TOTAL								25,536			

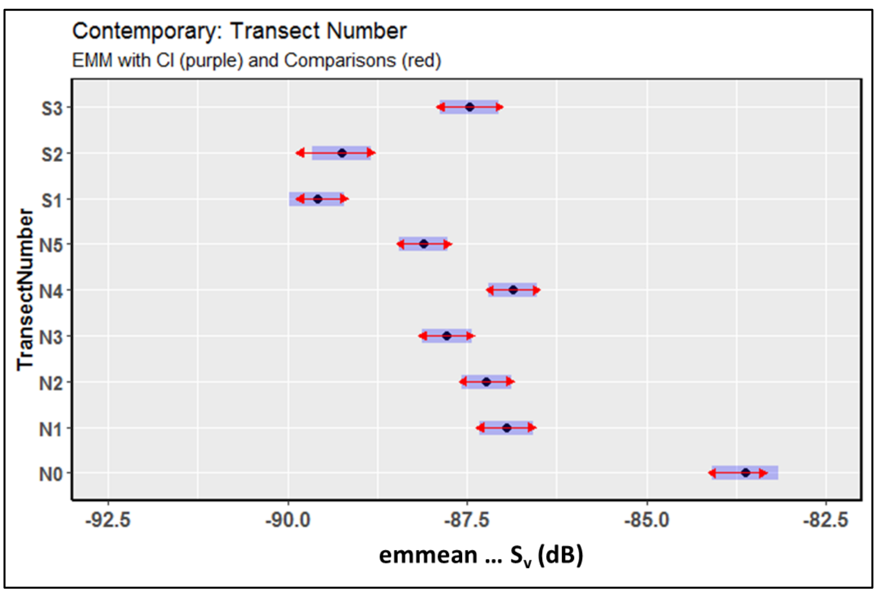


***Figure 26*. *Estimated Marginal Mean with Confidence Interval and Comparisons by Transect – Contemporary Only*.** *Graph displays results from the estimated marginal mean table. Black dot: estimated marginal mean. Purple bar: range from lower confidence level to upper confidence level (95% confidence interval). Red arrows: comparison range. Where the comparison levels (red arrows) overlap, the difference between the estimated marginal means is not statistically significant. Where the comparison levels (red arrows) do not overlap, the difference between the estimated marginal means is statistically significant. x-axis minimum and maximum has been standardized to encompass the full range necessary for all data reported in this section.*

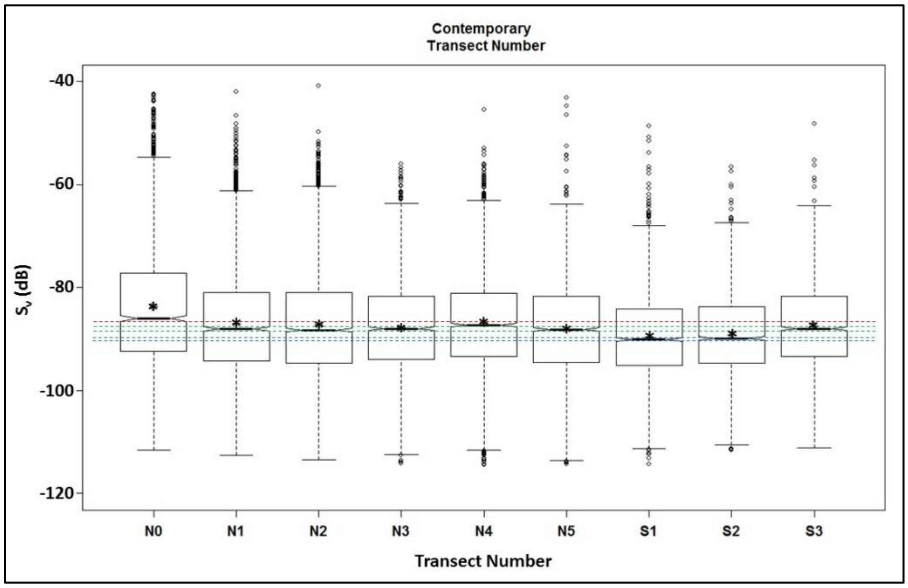


Figure 27. Notched Boxplot by Transect – Contemporary Only. *Boxplot of non-zero relative fish density data. Boxplot was calculated with the data in its log form (S_v) and are presented in order to provide visual representation of the data used in these analyses. See text and caption for Figure 17 for more information.*

Table 42*. *Results Summary Table – Contemporary Only. *Left three columns: number of observations (n) for fish absence (n:0), fish presence (n:1), category total (n:all). Percent columns: see caption to Table 4. P:A: Results from Presence:Absence analyses. Values indicate statistical groupings. "1" indicates the category with the highest percent of "present" observations within the Explanatory Variable. max S_v: checkmark indicates the category with the highest fish density observation (S_v) within the Explanatory Variable. data range: checkmark indicates the category with the widest range between the maximum and minimum fish density observation (S_v), "w" indicates the category with the widest range between the upper and lower whisker positions. EMM: compact letter display indicating groupings from the fish density analyses. "1" indicates category with the highest estimated marginal mean. Numeral groupings reported here are in opposite order as reported in the individual Results tables where "1" indicated category with the lowest estimated marginal mean. Median: compact letter display indicating groupings defined by overlap of the notch ranges in the notched boxplots. "1" indicates category with the highest median S_v.*

See table on the next page.

See previous page for Table Description.

Explanatory Variable	Category								>50%	50%	50%		< 50%	P:A	max S_v	data range	EMM	median
		n:0	n:1	n:all	0	1	100%	0	1	0	1	0	1					
Research Program	historical	7,778	3,569	11,347	68.5%	31.5%	100.0%					68.5%	31.5%	2			1	1
	contemporary	34,133	25,536	59,669	57.2%	42.8%	100.0%					57.2%	42.8%	1	√	√w	2	2
Study Area	CLA	22,236	18,133	40,369	55.1%	44.9%	100.0%					55.1%	44.9%	1	√	√w	1	1
	reference	11,897	7,403	19,300	61.6%	38.4%	100.0%					61.6%	38.4%	2			2	2
Tide Phase	high	3,119	1,833	4,952	63.0%	37.0%	100.0%					63.0%	37.0%	2		w	2	2
	ebb	**13,676**	**13,291**	**26,967**	**50.7%**	**49.3%**	**100.0%**			50.7%	49.3%			1		√	2	2
	low	4,386	2,556	6,942	63.2%	36.8%	100.0%					63.2%	36.8%	2	√		1	1
	flood	12,952	7,869	20,808	62.2%	37.8%	100.0%					62.2%	37.8%	2			3	3
Diel State	dawn	660	551	1,211	54.5%	45.5%	100.0%					54.5%	45.5%	2			3	3
	day	20,443	11,034	31,477	65.0%	35.0%	100.0%					65.0%	35.0%	3	√	√w	3,2	2
	dusk	894	529	1,423	62.8%	37.2%	100.0%					62.8%	37.2%	3			2	2,3
	night	**12,136**	**13,422**	**25,558**	**47.5%**	**52.5%**	**100.0%**	47.5%	52.5%					1			1	1
Survey	**2016-May**	**2,587**	**3,319**	**5,906**	**43.8%**	**56.2%**	**100.0%**	43.8%	56.2%					2			4	4
	2016-Aug	5,556	1,190	6,746	82.4%	17.6%	100.0%					82.4%	17.6%	8			5	5
	2016-Oct	4,402	2,344	6,746	65.2%	34.8%	100.0%					65.2%	34.8%	6			6	6
	2016-Nov	**2,768**	**3,371**	**6,139**	**45.1%**	**54.9%**	**100.0%**	45.1%	54.9%					2			3	3
	2017-Jan	**2,788**	**3,998**	**6,786**	**41.1%**	**58.9%**	**100.0%**	41.1%	58.9%					1			5	5
	2017-Mar	5,238	1,698	6,936	75.5%	24.5%	100.0%					75.5%	24.5%	7			7	7
	2017-May	**3,277**	**3,562**	**6,839**	**47.9%**	**52.1%**	**100.0%**	47.9%	52.1%					3	√	√w	2	2
	2017-Jul	**3,392**	**3,322**	**6,714**	**50.5%**	**49.5%**	**100.0%**			50.5%	49.5%			4			1	1
	2017-Aug	4,125	2,732	6,857	60.2%	39.8%	100.0%					60.2%	39.8%	5			5	5
Transect	N0	4,168	1,864	6,032	69.1%	30.9%	100.0%					69.1%	30.9%	5		w	1	1
	N1	3,866	2,848	6,714	57.6%	42.4%	100.0%					57.6%	42.4%	3			2	3
	N2	3,803	3,326	7,129	53.4%	46.6%	100.0%					53.4%	46.6%	2	√	√	23	3
	N3	3,706	3,259	6,965	53.2%	46.8%	100.0%					53.2%	46.8%	2			34	3
	N4	**3,517**	**3,462**	**6,979**	**50.4%**	**49.6%**	**100.0%**			50.4%	49.6%			1			2	2
	N5	**3,176**	**3,374**	**6,550**	**48.5%**	**51.5%**	**100.0%**	48.5%	51.5%					1			4	3
	S1	3,876	2,726	6,602	58.7%	41.3%	100.0%					58.7%	41.3%	3			5	4
	S2	4,045	2,366	6,411	63.1%	36.9%	100.0%					63.1%	36.9%	4			5	4
	S3	3,976	2,311	6,287	63.2%	36.8%	100.0%					63.2%	36.8%	4			234	3

DISCUSSION

Presented above was a data visualization and analytical *approach* designed to provide a methodology to explore the hydroacoustic data collected in Minas Passage to answer questions pertinent to the needs of FORCE personnel. It was 3-pronged:

1) *exploratory data visualization*: to gain an understanding of the underlying historical and contemporary data available for spatial and temporal analysis

2) *fish presence:absence*: to investigate the relationship between the spatial and temporal distribution of the presence of fish and the predictor variables

3) *relative fish density (using S_v as proxy)*: to investigate the relationship of the magnitude of relative fish density to spatial and temporal variables

This approach was undertaken to gain insights on the probability of recording observations of fish presence and to understand the relative density of fishes, in time and space. The predictor or explanatory variables for fish presence and density that could be evaluated were categorical: temporal (historical vs. contemporary, or by survey), spatial (CLA vs. reference study area, or by transect), and environmental (tide phase, diel state, or with and against predicted tidal flow).

The data used for the analyses included seven surveys conducted during the historical research program (Aug 2011 – May 2012) and nine surveys conducted during the contemporary research program (May 2016 – Aug 2017). The post-processed data was exported from Echoview in 20-m along-shiptrack distance bins integrated over the whole water column, the “20-m dataset”.

Data from the contemporary dataset were used for the analysis examples presented in this report. This approach was taken because statistical differences between the historical and contemporary dataset were found with both the presence:absence analysis and the relative fish density analysis. In addition, there were sufficient differences in the survey design and execution that deeper investigation into those differences is warranted before combining the datasets for analysis.

It should be noted that there were categorical gaps within the dataset. For example, there were no data collected during *dawn or dusk on a low-slack tide* during the nine contemporary surveys. Similarly, there were *transects* within the contemporary dataset for which no data was collected during dawn and dusk and during high and low slack tide periods. Should FORCE want to understand the dynamics of fish presence and relative density across these spatial and temporal categories, a detailed analysis of the data gaps in light of the questions pertinent to FORCE could help guide discussions concerning potential changes to the survey plan.

Within the contemporary dataset, where the number of categories within an explanatory variable exceeded two, the statistical results of the presence:absence analysis generally differed from that of the relative fish density analysis in terms of which of the categories statistically differed or not. These findings suggest that the presence:absence ratio of observations was not necessarily an indicator of the relative density of fish passing under the transducer. Selected findings are summarized in the Executive Summary.

The analysis examples included in this report were designed to provide an initial understanding of the data relative to the explanatory variables at a highly aggregated level and to demonstrate the approach. The results provide insights to form inquiries that could be conducted to dig deeper into the data at finer scales in order to answer pertinent questions. The results found using the data at finer scales can also be used to confirm whether insights from the highly aggregated data can be generalized or are a function of analyses using such highly aggregated data. For example: the results show that fish presence exceeds fish absence at night when the data is aggregated over the entire contemporary dataset, but this may not hold when examined on the finer levels of e.g. season, where behavior of the fish may differ on that temporal scale or on a spatial scale (e.g. transect). Much more investigation can and should be done using the scripts included with this document.

Further inquiries into the data that could be considered:

1) In both analytical approaches (fish presence:absence and relative fish density), the contemporary dataset was found to statistically differ from the historical dataset. The source of the difference may be one of natural variability or the difference may have also been influenced by the differences in survey design and execution. For example: historical transect length was nominally 1 km whereas the transect lengths during the contemporary surveys were nominally 2 km, during the historical survey each transect was traversed once during each grid pass (i.e. either with <u>or</u> against the direction of tide flow) whereas the transects were traversed twice during the each grid pass of the contemporary surveys (i.e. both with <u>and</u> against the direction of tide flow). In addition, between the two research programs there was a strong imbalance in the proportion of observations collected over the diel states (historic: day =78%, night=18%. contemporary: day=53%, night=43%). Deeper investigation into the sources influencing the statistical differences between the historic and contemporary datasets may provide insights as to whether analyses for the two datasets should remain separate.

2) Large imbalances in the count of observations between categories can cause statistical testing to become sensitive to very small, inconsequential differences resulting in statistical significance for small, uninteresting effects. In addition to the statistical effect is the question of whether the shortened window of observations that result in the smaller counts of observations generated a bias in the dataset. For example, the high and low slack tide periods are designated as the half-hour prior to and following the time of predicted slack for a total of one hour each occurring twice per day. Consequently, the count of observations for the two periods of running tides (ebb and flood) are an order of magnitude larger than the count of observations during the slack periods. One test that may be of interest is to select an hour of data from the ebb and flood

periods and rerun the tide phase analyses. The selected hour could be the hour adjacent to the slack period, and then again select an hour during the peak of flow.

3) For the 30 categories within the six variables analyzed, six categories had fish "present" counts that exceeded "absent" (night, May 2016, Nov 2016, Jan 2017, May 2017, N5) and for three categories the ratio of presence:absence was nominally 50:50 (ebb, Jul 2017, N4). "Absent" counts exceeded "present" counts for the remaining 21 categories (Table 4). Before generalizing findings from highly aggregated data, inquiries into the data at finer scales is warranted, including the night example referenced above and other distinctions in the dataset.

4) It was noted in the Analytical Approach: Fish Presence:Absence section of this report that there were pairings of transects for which there were not statistical differences in the presence:absence ratio. While it was suggested that those findings may provide guidance if transects need to be skipped for time or if there is an effort to increase survey frequency under the same budget constraints, the results in the Analytical Approach: Relative Fish Density suggest that there may be a different set of transect pairings for which there are or are not statistical differences. Given that there is the suggestion of transect pairings at this level of highly aggregated data, those findings should be investigated at finer scales.

5) The high value outliers should be explored to understand the particular states of an explanatory variable associated with observations of high fish density.

While the analytical approach presented here did provide insight using data summarized over spatial and temporal scales, and deeper inquiries at finer levels of summarized data will provide new understandings or confirmation of the findings at the summarized levels, more data needs to be collected to be able to draw larger inferences. In particular, the dataset needs to continue to be built such that multi-year data in comparable months are available. In addition, given the absence of a seasonal pattern and the preponderance of statistical differences between surveys, it may be advisable to increase sampling frequency within each month, sampling on consecutive days in order to get a finer scale understanding of the patterns and variability of fish presence and density in Minas Passage.

During the re-analysis of these data, it came to light that the echosounder gain settings during the contemporary surveys were not appropriately calibrated. After consultation with acoustic-community leaders it was determined that a post-hoc methodology by which to correct the calibrations was not available (McGarry and Zydlewski, 2018). Consequently, the echosounder gain settings have been standardized to the Simrad default settings (McGarry and Zydlewski, 2018; 2019). The calibration procedures were subsequently updated starting with Survey 15. Because appropriate calibration is fundamental to quantitatively compare survey results over time, distinguishing the contemporary dataset containing surveys with valid calibrations from those with the standardized calibration parameters is advised. Distinguishing the datasets will allow for the analyses to be combined or separated as appropriate.

Literature Cited

Baker, M., M. Reed, A. Redden (2014). "Temporal Patterns in Minas Basin Intertidal Weir Fish Catches and Presence of Harbour Propoise during April – August 2013." ACER, Wolfville, NS, Tech. Rep. 120.

Dadswell, M. (2010). Occurrence and migration of fishes in Minas Passage and their potential for tidal turbine interaction. Technical Report.

Daroux, A., G.B. Zydlewski (2017). Final Report: Marine Fish Monitoring Program Tidal Energy Demonstration Site – Minas Passage. Submitted to Fundy Ocean Research Center for Energy, 16 October 2017. 34 pp.

Daroux, A., L.P. McGarry, G.B. Zydlewski (2017). Marine Fish Monitoring at FORCE: Report on processing and analysis of surveys from May, July and August 2017. Submitted to Fundy Ocean Research Center for Energy, 29 December 2017. 16 pp.

MacLennan, D.N., P.G. Fernandes, J. Dalen (2002). A consistent approach to definitions and symbols in fisheries acoustics. ICES Journal of Marine Science, 59: 365-369.

Melvin, G.D., N.A. Cochrane (2014). Investigation of the Vertical Distribution, Movement and Abundance of Fish in the Vicinity of Proposed Tidal Power Energy Conversion Devices. Final Report, OEER/OETR Research Project 300-170-09-12.

McGarry, L.P., G.B. Zydlewski (2019). Notes for EK80 CW Calibration Settings for FORCE. Submitted to Fundy Ocean Research Center for Energy, 02 February 2019. 4 pp.

McGarry, L.P., G.B. Zydlewski (2018). Calibration Quality Control. Submitted to Fundy Ocean Research Center for Energy, 22 June 2018. 10 pp.

Rulifson R., M. Dadswell (1995). Life history and population characteristics of striped bass in Atlantic Canada. Transactions of the American Fisheries Society, 124(4), 477-507.

APPENDIX A: Technical Notes

Historical Survey Detail

Table A1: Historical Surveys. *Each survey consists of three to twelve repeats of the grid defined by the following transect lines: T0, T1, T2, T3, T4, T5, T6, T7, T8, Y1, X1, Y2*[7]. *Only data collected from "T" and "X" transects were included for analysis. Additional notations are listed below.*

Survey [9,11,12,13]	**Month**[0,3]	**Start date**	**Start time**[2,5]	**End date**	**End time**[2,5]	**Day/ Night**[5,7]	**Temperature (°C)**	**Turbine presence**	**Moon Phase Tide Range**
1	Aug 2011 (4:3)	2011-08-22	11:45 (08:45)	2011-08-22	21:28 (18:28)	D	15.4	No	7 m
2	Sep 2011 (4:3)	2011-09-19	10:55 (07:55)	2011-09-19	20:23 (17:23)	D	15.7	No	8 m
3	Oct 2011 (4:3)	2011-10-03	09:53 (06:53)	2011-10-03	20:18 (17:18)	D	15.0	No	10 m
4[6]	Nov 2011 (3:3)	2011-11-22	14:22 (10:22)	2011-11-22	22:32 (18:32)	D	10.3	No	/ 11 m
5[10]	Jan 2012 (10:9)	2012-01-25	18:32 (14:32)	2012-01-26	16:15 (12:15)	D/N	3.6	No	11 m
6[4]	Mar 2012 (12:11)	2012-03-19	14:23 (11:23)	2012-03-20	13:33 (10:33)	D/N	2.5	No	/ 9 m
7[10]	May 2012 (5:4)	2012-05-31	12:09 (09:09)	2012-05-31	23:12 (20:12)	D	9.5	No	/ 10 m
8	Jun 2012[1]								

[0] **September 2010:** a datasheet for a September 2010 survey is included in the historical datasheets provided by Dr. Melvin. But no echosounder data for that survey was delivered to UMaine. See Melvin and Cochrane (2014) for reporting that includes that survey.

[1] **June 2012:** according to the datasheet, 10 grids were executed on June 25-26, 2012, but only data for one partial grid was received with the transfer of the "Melvin" data to UMaine. Because only one partial grid of data is available, the survey has been excluded from the analytical work. Note that an event that may be of interest: "extremely high fish concentrations" is noted on page 16 of Melvin and Cochrane (2014) but unavailable for inclusion in the work herein.

[2] Echosounder data and the associated datasheets were recorded with **"time" set to GMT** (Greenwich Mean Time). Testing of tide height change per echograms and predicted tide height change confirmed that GMT is the correct designation of time for the historical "Melvin" data and is also consistent with reporting in Melvin and Cochrane (2014). "Time" associated with the echo integrated data was converted to local time after export from Echoview. "Time" as designated on the datasheet was converted to local time to ensure that tide, diel, and "with/against" stages were appropriately assigned. Time in parentheses in Table A1 is Local Time.

Note: If further analysis requires subsequent metadata merges with the historical data, use the "Time Offset" feature in the Echoview Fileset Properties to set the conversion (3 or 4 hours depending on time of year) from GMT time at which the data was collected to local time before executing the exports. Then any additional exports of the historical datasets can be directly merged with the historical datasheet for which "time" has already been converted to local time.

[3] Numbers in parentheses in the Month column are the **number of partial:complete grids** executed for the survey. Three was the minimum number of complete grid passes in the historical surveys. To standardize the number of grid passes per historical survey, three grid passes were included in the data used for analyses. Complete grid passes were evaluated and three were selected to maximize "good" data (e.g. selecting grids with lesser entrained air where possible, etc.). One exception is the Nov 2011 survey during which no data were reported in the reference study area. So although three complete grid passes were executed, incomplete passes without the reference study area data have been included in the analyses. (Data were collected in the reference Study Area but not transferred to the University of Maine.) The grids selected for inclusion in the analyses in this report are the following:

Aug 2011: 1,2,3
Sep 2011: 1,2,3
Oct 2011: 1,2,3
Nov 2011: 1,2,3 <- NOTE: no data was available from the reference study area during this survey[6]
Jan 2012: 5,8,9
Mar 2012: 7,10,11
May 2012: 2,3,4

Detailed notes regarding data quality of grid passes can be found in the "20170519 Datasheet_Melvin WORKG" tab of the *20170519 Datasheet_Melvin WORKING20190115LPM.xlsx* spreadsheet. See also the "Datasheet Documentation" tab in the *20190115 AllDataParseToMonthly_MelvinONLY_20mBinsFullWaterColumn.xlsx*. Final grid selection notes are in: *MelvinGridNOTES.xlsx*.

[4] For survey **March 2012**, the datasheet lists 12 grid passes but T5, T7, and T8 were populated with zeros on the datasheet for particular grids (#3 and #12). Therefore, Grids #3 and #12 are **not complete grids** and should not be included in the grids selected for analysis. See *MelvinGridNOTES.xlsx* for more detailed information for grids in **all historical surveys**.

[5] **Start Time, End Time, and Day/Night** are reported for the entirety of each survey dataset, whereas a subset of the data (three complete grid passes for each survey) were used in analyses. Therefore, Start Time, End Time, and Day/Night as represented in the analyses may differ from what is reported here. See *MelvinGridNOTES.xlsx* for more detailed information for grids in **all historical surveys**.

[6] **Nov 2011:** No .raw data were provided for the **reference transect**. Raw data for the cross-channel transects (X1 and X2) were included – but not the connecting reference transect. Therefore, although 3 complete grid passes were executed during data collection, the data provided for analyses were not 3 complete grid passes for Nov 2011. Given that analyses presented in this report did not include historical data aggregated at scales finer than "research program", the data for the CLA study area for November 2011 were included in the analyses.

[7] Note the **predominance of day coverage**. Percentages reported here for CLA and referece transects only.
Detail for entire historical dataset… Dawn: 3%, Day: 69%, Dusk: 2%, Night: 26%
Detail for (complete) grids included in 2019 analysis… Dawn: 3%, Day: 78%, Dusk: 1%, Night: 18%

[8] There is an **inconsistency in the reference and cross-channel transect notations** in Melvin and Cochrane (2014). Table 4 therein refers to the reference transect as Transect X1 and the cross-channel transects as Transect Y1 and Transect Y2. Table A5-9 and the narrative on page 16 refer to the reference transect as Transect Y1 and the cross-channel transects as Transect X1 and Transect X2. It appears that the X1 in Table 4 was in error. Given that

the datasheets used for these analyses were constructed from the Table A5-9, the **reference transect is referred to as Transect Y1 in this document**.

[9] In the historical dataset, **passes over each transect** were in **one direction only** ("with" OR "against" tidal stream flow). Whereas passes over each transect in the cotemporary dataset were executed twice; once "with" AND once "against".

[10] Surveys in **Jan 2012** and **May 2012** include transects during which the vessel traversed partway across the transect length and then **returned to the start of the transect** following which the transect was surveyed in its entirety. The data from the initial partial transect were excluded from analyses via edits to the start and end times of the transects in the datasheet Excel file used for the metadata merge. Given that these exceptions were not noted in the datasheet .pdfs received with the historical dataset, and that the start and end times on those datasheets encompassed the whole effort for that transect rather than limiting the start and end times to the one clean traverse across the transect, the partial transect would not have been excluded from the analytical dataset if the cruise tracks hadn't been plotted confirming the spatial extent of the data. If one doesn't catch this particular exception, then both "with" and "against" data would be included although the metadata merged would have labeled the direction as either "with" or "against" as specified by time-of-day.

[11] **GPS values** are particularly erratic in the historic dataset. Caution should be exercised when analyzing data specified by the recorded latitude and longitudes.

[12] The historical data used in the analyses included in this report were as exported from Echoview 7 by Aurelie Daroux.

[13] In the historical dataset, the **data collection file** was run as **one long file** for the entire survey. Implications: (a) the .raw files imported to Echoview define the data for the entire survey, precluding the ability to generate Echoview files for individual transects or grids, (b) for processing historical data by transect once the data is exported from Echoview requires accuracy in the start and end of each transect line as defined by the time entries in the datasheets, (c) "along" data is included in the Echoview exports and therefore needs to be explicitly excluded at a later point in the processing. (For the contemporary data exported from Echoview, the Echoview files for the "along" transects were excluded from the export process thereby eliminating the necessity to explicitly exclude "along" later in the processing.) Note that the time entries in the original historical datasheets were not sufficiently accurate to exclude data from the transits between transects. LPM produced a new datasheet for the historical data with more tightly defined start and end times for the transects. LPM made two passes at this process for all 7 of the historical surveys. The results are visualized here:

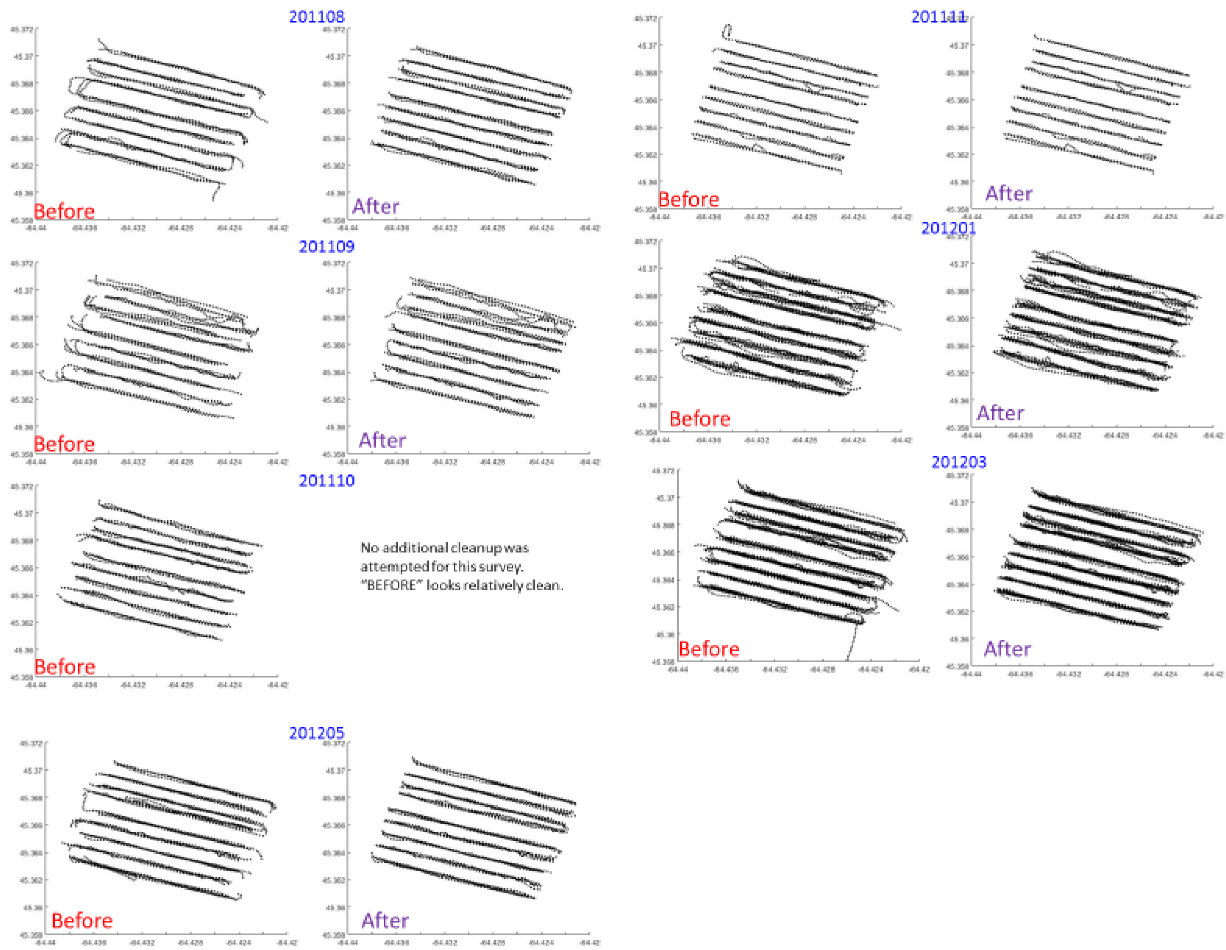


Figure A1. CLA Transect Lines in the Historical Dataset. *Because the echosounder data was collected in one long .raw file, the definition of the ends of the transect lines were completely dependent on the start and end times defined in the datasheet. Shown here for the "20-m dataset" are the definitions of the transect lines based on an unedited datasheet ("Before") and the edited datasheet ("After") after two passes at refining start and end times. Each dot representing a 20-m along-shiptrack distance bin.*

Summary of Historical Survey Dataset Reprocessing since December 2017 report (Daroux *et al.* 2017)

1. no changes were made to the data exported from Echoview (**#12** above)
2. definition of the ends of the transects in the metadata file were adjusted to remove "20-m bins" associated with transits between transects (**#13** and **Figure A1** above).
3. time recorded at GMT in the metadata file was converted to local time and environmental metadata (tide phase, diel state, "with/against", etc.) reassigned based on local time (**#2** above)
4. time in the exported Echoview files was converted from GMT to local time (**#2** above)
5. partial Jun 2012 survey data was excluded from analyses (**#1** above)
6. complete versus incomplete grid passes were documented and in conjunction with review of echograms for bad data and excessive turbulence, three grids for each survey were selected for inclusion in the analyses (**#3** above)
7. partial repeats of transects were identified and the metadata file edited in order to exclude the partial passes (**#10** above)

Contemporary Survey Detail

Table A2: Contemporary Surveys. *Each survey consists of 4 repeats of the grid defined by the following transect lines: N0, N1, N2, N3, N4, N5, South_CW[L], S1, S2, S3, North_FM with calibration files. Only data collected from "N" or "S" transects were included in analyses. Additional information and notes regarding where data differs from the standardized grid are included in the notations are below.*

Survey	Month[Z]	Start date	Start time[A]	End date	End time	Day/ Night[T]	Temperature (°C)	Turbine presence	Moon Phase Tide Range
1	May 2016 [B,C,D,E,O,Q,R,Y]	2016-05-28	06:01	2016-05-29	05:35	D/N	7	No	10 m
2	Aug 2016 [C,E,R,Y]	2016-08-13	09:09	2016-08-14	07:40	D/N	15	No	7 m
3	Oct 2016 [C,D,F,M,P,R,Y]	2016-10-07	05:45	2016-10-08	04:21	D/N	15	No	8 m
4	Nov 2016 [E,G,H,HH,P,R,QQ,S,Y]	2016-11-24	08:38	2016-11-25	09:07	D/N	8.0	Yes	8 m
5	Jan 2017 [C,G,H,HH,R,RR,Y]	2017-01-21	06:55	2017-01-22	05:55	D/N	1.5	Yes	7 m
6	Mar 2017 [P,R,Y]	2017-03-21	08:24	2017-03-22	06:04	D/N	4	Yes	7 m
7	May 2017 [I,R,Y]	2017-05-04	19:57	2017-05-05	18:21	D/N	5	Yes (free spinning)	9 m
8	Jul 2017 [I,J,R,Y]	2017-07-03	21:34	2017-07-04	19:09	D/N	12	No	/ 8 m
9	Aug 2017 [I,R,Y]	2017-08-30	18:53	2017-08-31	17:37	D/N	15.7	No	7 m

[A] **Time** recorded here and in the .raw files is local time at Minas Passage.

[B] Raw data collected in passive for **AC/DC test** in addition to standard grid

[C] Raw data collected for **stationary data** in addition to standard grid

[D] Raw data collected for **transects in addition to** standard grid

[E] Raw data collected for **unspecified test** in addition to standard grid

[F] Raw data collected for **ping rate test** in addition to standard grid

[G] **Both South and North transects** (CW and FM respectively) were done on the east side of the grid

[H] Raw data collected for **"T" transect** between N2 and N3 in addition to standard grid

[HH] **"T" transect** looks like it means: **"Turbine"**. As of April 24, 2018, how the "turbine" data is labeled in the digital datasheets has not been standardized (i.e. should they be classified as "with/against" or "stationary" or simply "turbine" given that in some cases they weren't stationary – but weren't doing actual transects in an effort to run over the turbine multiple times).

[I] Transect execution for **South CW and North FM were reversed** (i.e. the actual transects were: South FM and North CW). See note "L" below for more information.

[J] **Survey 8** raw data collected for **GR1-N0** and the **first file of GR1-N1**: The set of transect coordinates that were stored in the ship's plotter were deleted before the start of this survey. "Helper" gave the captain old coordinates. GR1-N0 is indeed offset (Figure A2a). It falls about midway between where it should be (N0) and the next transect N1. Both GR1-N0 transects ("with" and "against") were excluded from analytical processing. The N1 segment falls within the cloud of N1 transects for Grid 1 from each of the contemporary Surveys 1 through 9. So all the GR1-N1 data is included for processing.

[L] To alleviate the nightmare that is the **North/South CW/FM** issue and filenames, the "along" transects have been ***excluded from analytical processing*** (i.e. the "along" were not intended to be included in analyses and therefore were not included in the "alldata" files.) NOTE that North/South confusion for surveys 7,8, and 9 cascades throughout the processing steps. i.e. .raw file names = South when they were actually traveling North. The EV file names were corrected to correspond to the actual direction of travel, but when the .vbs script exports the data from EV, the script incorporates the .raw file name rather than the EV filename, and hence we're back to files with names that don't correspond to the actual direction of travel.

[M] Survey 03 has both **North/South CW**. As of January 2018, only the South CW files have been imported into Echoview. Therefore, there is more "along" CW data available than is indicated by the number of "along" EV files. (But again, "along" data were excluded from analytical processing for this report.)

[N] **Oct 2016** includes one transect labeled **Grid 5.** This data is excluded from analysis.

[O] **Entrained air** so severe and persistent through the water column that sections of transects are assigned as bad data regions and eliminated from any effort to ascertain whether fish are present. (e.g. Survey1_GR3_N1W.EV and Survey1_GR3_N1A.EV) Note: To distinguish passive data regions from actual bad data regions, each were assigned a different type of bad data region within Echoview.

[P] During transect **GR4_N2A** for surveys **Oct 2016, Nov 2016, Mar 2017**, there is a feature that **looks non-biological** (derelict gear? a tether?). In all 3 cases the time is within 30 minutes of slack (low) and is located at 45 22.166'N 64 26.195'W or thereabouts. Given the consistency of location and feature, the signals were designated as a bad data/no data region in each of the 3 surveys. 45 22.166' N 64 26.195' W 45 22.161' N 64 26.201' W 45 22.170' N 64 26.213' W The ?tether? rises about 8 m off the seafloor and is 50+ m long.

[Q] **Survey 1 skipped transects.** There are a lot of .raw files per transect (20+). Datasheet .docx has a note that the transects should be shortened. Taking too long. Specifically GR2_N2A took 3 hours (cruisetrack looks like they got caught in an eddy). Lots of entrained air. There are 746 .raw files associated with that transect. GR2_N4A is missing from .raw files and is not listed on the datasheet. There are no notes on the datasheet as to why that transect was skipped. However, they went directly into South_CW therefore must be skipped transects to make up time. They did one transect (S1W) on the control site and then went into North_FM and started GR3. See *SimradFilesPerTransect.xlsx* for details as to which were skipped, etc.

[QQ] Survey 4 **didn't end file for GR2_N5A** and began heading south to control site **skipping N5W**. In Echoview, only one file is included for N5A so as to exclude the cross-channel transit. No Echoview file was created for the skipped N5W.

[R] Data collection procedures for **calibration** data were insufficient to provide reliable calibration parameters. For more information see the **Calibration Quality Control Report issued June 22, 2018** and the **Notes for EK80 CW Calibration Settings issued February 2, 2019**.

[RR] Special consideration for **this calibration**: Collection range window included nearfield. (Min range set at 1.2 m whereas the nearfield/farfield boundary used is 1.7 m.)

[S] **Nov 2016**. **Skipped transects for time** – and **started GR4 still in FM**.

[T] **Day/Night coverage** detail for CLA and control transects: Dawn: 2%, Day: 53%, Dusk: 2%, Night: 48%

[Y] Note that for Survey 10 and prior of the Contemporary surveys, the **files names** (e.g. GR1_N0A and GR1_N0W) were *labeled "against" and "with" respectively based on the <u>conceptual</u> plan for the grid*, rather than based on whether the tide was flooding or ebbing. The merge of the EV exported data with the datasheet has corrected the data internal to the "alldata" files (i.e. the WithAgainst column is consistent with direction of predicted tide), leaving the file names disconnected from the physical attributes of the tidal flow.

[Z] Review the **trackline positions and direction of travel** for the **"along" transects** before you use them for analyses. Somewhere along the way the grid lost its shape in terms of the locations at which the "along" transects are recorded. (See screenshots in "A2b" below.)

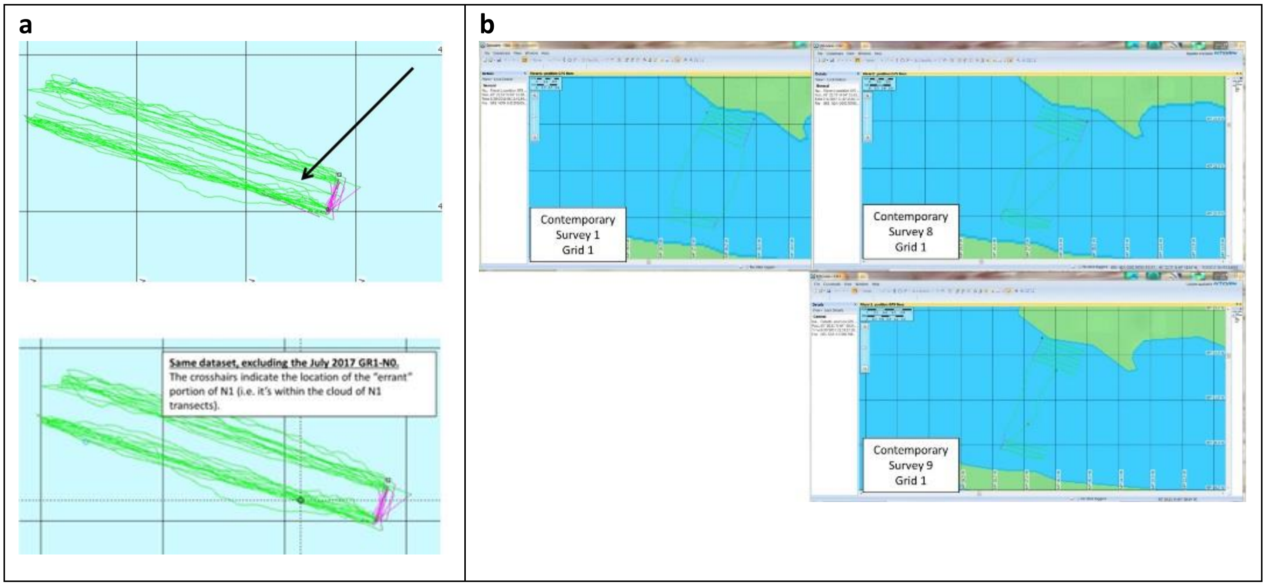


***Figure A2**. **Contemporary Dataset Notes.** **(a)** Physical location of the two transects (N0 and the first pass of N1) for survey 8 (July 2017) that were executed under old coordinates, relative to plotted transects N0 and N1 from grids one through four for surveys one through nine. The arrow points to the "with" and "against" transects of the errant N0. N0 is distinctly between its assigned location and the location of N1. If the transects are nominally assigned 200 m apart, this one is 100 m apart. N0 for that grid pass is therefore excluded from analysis. The errant N1 plots within the cloud of N1 locations from surveys 1 through 9 and therefore was included in analyses. Notation **"J"** above. **(b)** Grid shape as defined in survey 1 is not held for surveys 8 and 9. Given that "along" transects are outside the scope of the analytical work described in this document, no further action or investigations was required (e.g. determining how many of the 9 contemporary surveys have compromised grid shapes). However, if the "along" is used in subsequently analyses, be sure to examine the position of the "along" transects to determine whether the data is appropriate to include. Notation **"Z"** above.*

<u>Summary of Contemporary Survey Dataset Reprocessing (since December 2017 report (Daroux *et al.* 2017))</u>

1. data was exported from Echoview 7 where processing for EK80 data was still in beta testing to Echoview 8 as recommended by Echoview
2. top and bottom lines within Echoview were adjusted to eliminate gaps that cause spurious data in the export
3. within Echoview, "bad data" regions were redefined for consistency: passive data was defined as bad data-empty water and regions of backscatter from non-biological targets were defined as bad data-no data. Use of the two definitions specifies which portions of the transect were excluded due to passive data collection versus portion of the transects lost to entrained air or other non-biological targets
4. all Echoview files (~72 per survey) were reviewed and corrected for errors in .raw data inclusion (e.g. "with" and "against" within the same EV file, data from more than one transect within the same EV file, etc.)
5. upon discovery of the calibration issues, extensive testing and then consultation with acoustic-community leaders was undertaken in order to determine if a post-hoc solution was available. Where one was not forthcoming, worked extensively with Echoview in an effort to create and test a post-hoc solution. When it was deemed that a final solution was not imminent, worked

with acoustic-community leaders to settle on an approach that would allow analyses to move forward (McGarry and Zydlewski, 2018).

6. once the approach to standardize the calibration parameters was identified (McGarry and Zydlewski, 2019) new Echoview calibration files (.ecs) were created for all nine contemporary surveys
7. Echoview export script was extensively updated to include export of EV file metadata with the export of the EV data for analyses
8. Echoview exports for all transect data for all nine surveys was executed including data, EV metadata, and EV files for archiving the data in the state used for these analyses
9. metadata (datasheet) files were completely reworked to correct errors and to reassign "with/against" based on predicted tidal phase (original entered was a combination of predicted tidal phase and perceived tidal phase in the field)
10. developed and tested new scripts to automate steps to prepare EV exported data for analyses (see Appendix C for more description of the scripts)
11. incorporated some data quality control tests into the scripts based on issues found in the December 2017 data and scripts. Some of these are articulated in "Notes" and "Cautions" below. See scripts for complete list of data quality control tests
12. worked with University of Maine statistician to develop a statistically rigorous approach to analyzing the hydroacoustic data from Minas Passage

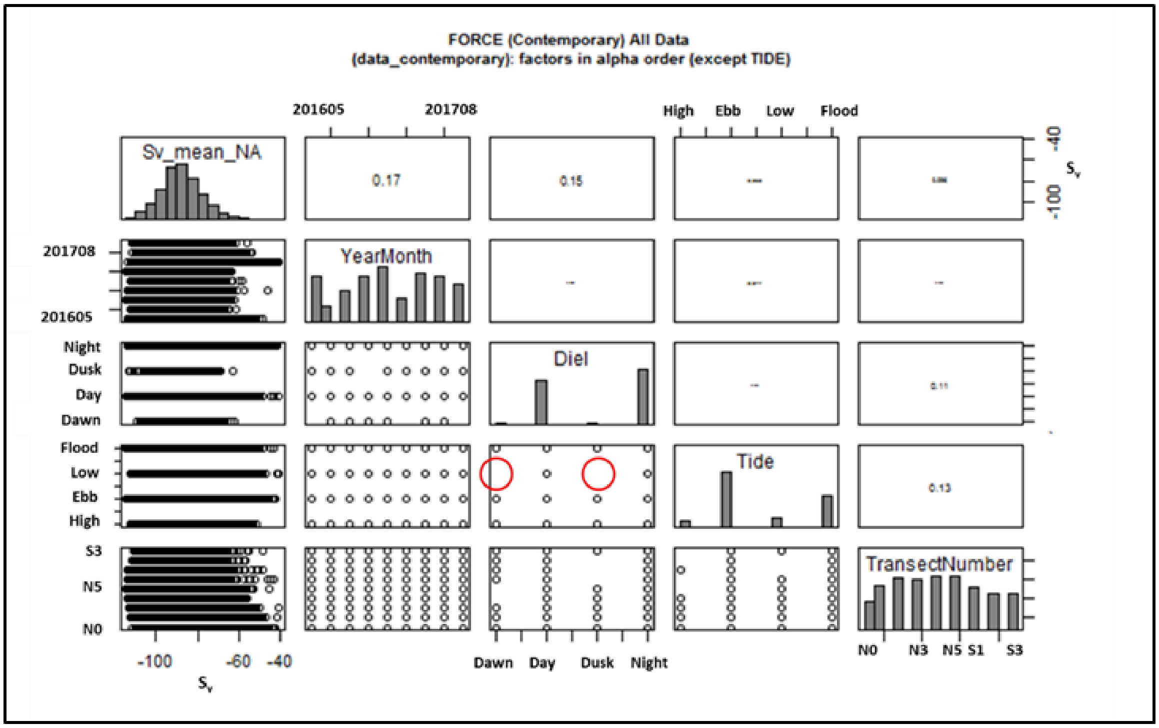


Figure A3: Panel Plot - Contemporary Surveys Only. *Panel plot of non-zero S_v values and selected variables contained within the contemporary portion of the "20-m dataset". Histograms on the diagonal: only the x-axis is associated with the histograms (i.e. the heights are relative heights of the number of non-zero S_v observations within the individual variable). Dot plots below the diagonal are read with both x and y axes as indicated by the categories in the histograms of the associated row and column. For example: the plot with the two red circles is a Diel-by-Tide plot. As per the x-axis labels, from left to right: dawn, day, dusk, night. As per the y-axis labels, from bottom to top: high, ebb, low, flood. Therefore, the two red circles highlight that no data was collected during low slack at either dawn or dusk. Note that there are gaps in transects by tide phase, and gaps in transects by diel state. Correlations among the pairs are posted in panels above the diagonal. Font size is indicative of magnitude of each correlation. Coding to generate this plot is included in the R scripts: SCRIPT2019 20mBinAnalysis.R.* ***Caution:*** *There may be an upper limit to the number of categories that can be shown within a variable. For example: when this plot was generated for the full dataset (historical plus contemporary), the histogram values for the first two surveys (YearMonth) were summed and shown as one bar.*

Survey Characteristics – Historical and Contemporary Surveys

Table A3: Contemporary and Historical Surveys. *Survey characteristics.*

	Historical/Melvin	**Contemporary/FORCE**
Number of Surveys	7	9
Target	1 or 2 full tide cycle: 12-24 hours	2 full tide cycles: 24 hours thereby encompassing tide cycles both day and night
Vessel Speed-Over-Ground (Nominal or Actual)	should be investigated	should be investigated
Number of Complete Grid Passes per Survey	3 – 11 (one survey also includes a partial 12th grid pass)	4
Grid Length	~1 km	~ 2 km
Number of Transects	CLA: 9 reference: 1	CLA: 6 reference: 3
Transect Direction Relative to Tidal Flow	data collection for every transect was executed once: "with" tidal flow **or** "against"	data collection for every transect was executed twice: "with" tidal flow **and** "against"
Diel Distribution of Data Collection *(Historical "Analytical" describes the distribution of data in the grids selected for inclusion in analyses.)*	…Dataset… Entire Analytical Dawn: 3% 3% Day: 69% 78% Dusk: 2% 1% Night: 26% 18%	…Dataset… Entire = Analytical Dawn: 2% Day: 53% Dusk: 2% Night: 43%
Tide Distribution of Data Collection *(Historical "Analytical" describes the distribution of data in the grids selected for inclusion in analyses.)*	…Dataset… Entire Analytical Low: 11% 17% Flood: 34% 27% High: 6% 6% Ebb: 49% 50%	…Dataset… Entire = Analytical Low: 12% Flood: 35% High: 8% Ebb: 45%

Table A4: ***Contemporary and Historical Surveys – Year, Month, Tide Range.*** *Shaded and hatched boxes indicate months in which hydroacoustic survey data was collected. Hatched shading indicates the surveys during which a turbine was present in the CLA study area. Number within the shaded box indicates the tide range (in meters) predicted for survey days.*

	2011	2012	2016	2017
January		11		7
February				
March		9		7
April				
May				9
June		10	10	
July				8
August	7		7	7
September	8			
October	10		8	
November	11		8	
December				

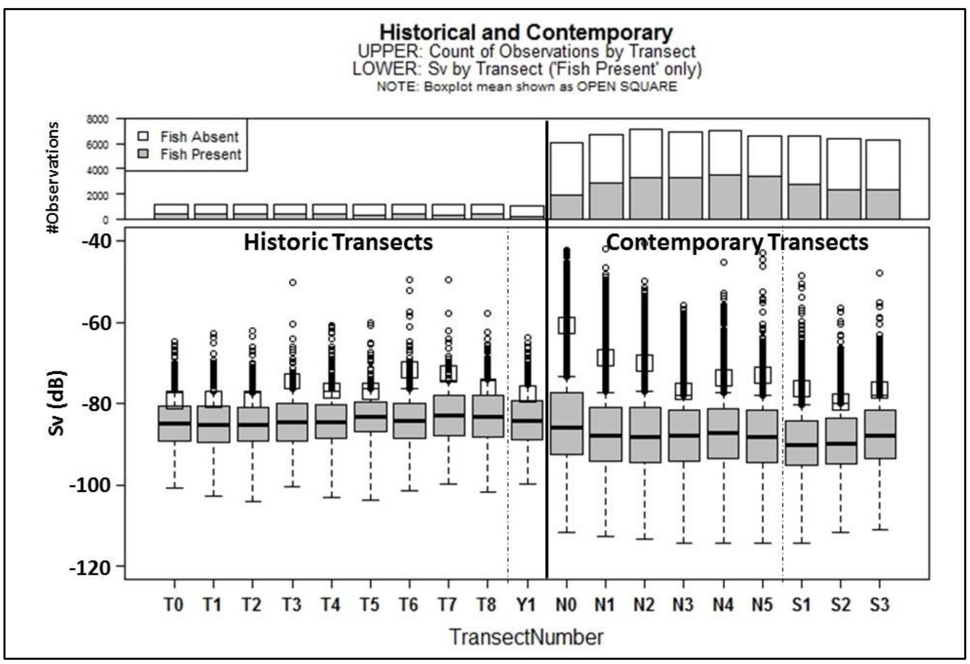


Figure A4: Distribution of S_v values by Transect for Historical and Contemporary Datasets. *Data shown are the data exported from Echoview integrated over the full water column binned in 20-m along-shiptrack distances (the "20-m dataset"). The data is highly aggregated. Each transect represents all data in the "20-m dataset" used in the analyses contained in the report. Top: Top of bar indicates the total number of observations for each survey. Axis range: 0 to 8000. Shaded portion indicates the number of non-zero observations (n = 29,105). White portion indicates the number of zero observations (n = 41,911). Bottom: Boxplots of non-zero S_v observations by transect. See text in the Explanatory Data Visualizations section of report for description of boxplot.*

Survey Design Notes – Contemporary Surveys

The original grid plan was designed such that four complete grid passes were completed within 24-hours resulting generally in two grid passes during day, executing one full grid during ebbing tide and one full grid during flooding tide, and two grid passes during night, again with one full grid during ebbing tide and one full grid during night. For each grid, every 1.8-km transect was traversed twice, once “with” the direction of tidal flow, and once “against” the direction of tidal flow, before moving to the next transect. The surveys were scheduled to begin on the ebbing tide with the EK80 echosounder set to record in “continuous wave” (CW) mode, starting by traversing transect N0 in the direction “with” the ebbing tide. Each successive transect, N0 to N5, were occupied in order (both “with” and “against”). Then a southward across-channel transect was executed terminating near Passage’s southern coastline. This cross-channel transect was designated “South_FM” to indicate that the direction of travel was southward across the channel and that the data would be collected with the EK80 echosounder set to record in “frequency modulated” mode. Upon completion of the southward transect, the EK80 echosounder was returned to its “continuous wave” mode, and three reference transects, S1 to S3, were each executed twice: once “with” and once “against” the direction of tide flow. To finish the grid, a northward return transect “North_CW” returned the vessel to N0. One grid pass consisted of one full set of all transects.

Note that the original grid plan called for the south across-channel transect to be conducted in continuous wave mode (South_CW) and the northward return transect in frequency modulated mode (North_FM). However, this convention was not consistently met during surveys 1 through 9. For example: in survey 3 both North and South transects were executed in CW mode and the two modes (FM and CW) were interchanged for surveys conducted in May, July, and August 2017. **In future surveys the convention should be standardized.** Note that there is now a mismatch between the echosounder .raw filenames and the contents of the .raw files. When constructing the Echoview files for the cross-channel data, extra caution is required to ensure that the intended data (regardless of the .raw filename) is included.

EV Exported Data Notes – Historical and Contemporary

- Interval: start ping differs between historical (Melvin) and contemporary (FORCE) surveys
    - Historical: Interval starts with the first ping regardless of whether there was a good GPS location associated with it
    - Contemporary: Interval starts with first ping for which good GPS location was available. So some pings may be skipped
    - SIGNIFICANCE: Where there were missing GPS locations for those start pings in the historical data, the GPS location for the initial Interval(s) (e.g. 20-m along-shiptrack bins) were populated with 999 for both longitude and latitude. For the contemporary dataset, the initial few pings without a good GPS location were designated as Interval 0, allowing the EV export to calculate the appropriate GPS location for the remaining contemporary Intervals
- Interval: 0
    - Interval "0" only occurs if there is no GPS associated with the very first ping recorded in the Echoview file AND the "Start interval numbering from the first ping in the echogram" box is checked in the Grid tab of the Variable Properties.
- Passive and Bad Data definitions
    - were assigned two different bad data types so that the portions of the echogram excluded from analysis could be distinguished between exclusion of noise/turbulence vs. passive data collection
- "Along" data for the contemporary dataset was excluded from analysis at a different processing stage than for the historical dataset.
    - contemporary: new data files were created for each transect during a survey, thereby facilitating the creation of individual Echoview files for each transect. Therefore, "along" and any other non-standard grid data were excluded simply by not exporting any echo integration data from EV files associated with any non-standard grid data
    - historical: generally survey data was collected in one file, thereby making it impossible to segregate standard grid data from non-standard grid data at the Echoview stage. Therefore, "along" data and partial grid data were included in the echo integration exports from Echoview and succeeding steps (e.g. the merge with the metadata, and included in the appending of the historical dataset with the contemporary dataset to create the "alldata" file). Exclusion of the "along" and unwanted grids was executed by explicit command in the R scripting resulting in the "data_subsetMaster" dataframe used for analysis.

APPENDIX B: Notes Going Forward

Cautions – Analytical

- **threshold settings in EV** were -66 dB (minimum integration (S_v) threshold) and -60 dB (minimum target strength (TS) threshold), changes to these settings for future surveys will alter the comparability of the data.
- **depth of transducer** (historical and contemporary) is listed at 0 m (i.e. no offset for the depth at which transducer is deployed).
    - Therefore "depths" reported are "range from transducer" unless some offset is applied to data outside of EV. (No offset has been applied to the data processed here.)
    - Therefore, in order to keep the datasets consistent in future analysis, if the same deployment configuration is used (boat and pole mount) depth of transducer should continue to be reported at 0 m.
    - Best Practice: the depth of the transducer should be recorded on the data sheet for each survey, and that offset from the surface entered into Echoview. By doing so, data recorded using differing deployment methods (e.g. different boat, different polemount) can be directly compared by depth. "Range" from transducer face is still available in Echoview even when offset for the depth of the transducer has been entered.

Cautions – Data Processing Procedures

- need to be super cautious **working in .csv or .xlsx**
    - "time" as exported from Echoview includes hh:mm:ss.SSS. However, it is not uncommon for the hours component to be lost when using .csv files (number formatting is not embedded in a .csv file).
    - also found that the decimal seconds got dropped in the .csv file – those decimal seconds can be important for getting lines of data in chronological order
    - Excel will sometimes split the contents of the "EV_filename" column across two columns which results in an offset of the contents of all following columns relative to the column headers
- it's an **easy check in EV** to plot the cruise track within each EV file.
    - This can function as a quality control that you've imported only the .raw files actually associated with that transect regardless of the filename, etc.
    - Also serves as a quality control that .raw files were appropriately labeled. (Example: Nov 2016 .raw file for GR2_N5A wasn't closed and renamed before turning south for the "along" transect. Therefore when creating the N5A EV file including .raw files by filename only, "along" data would be included in the N5A transect.)
    - Also serves as a quality control for any exceptions to the standardized survey plan
        - for example: two of the historical surveys include transects during which the vessel traversed partway across the transect length and then returned to the

start of the transect, following which the transect was surveyed in its entirety. These exceptions were not noted in the datasheets and the start and end times in the datasheets encompassed the whole effort for that transect rather than limiting the start and end times to the one clean traverse across the transect. The results of not discovering this excursion from the survey plan is that for each of those transects, both "with" and "against" data would be included in analysis although the metadata merge would assign the direction as either "with" or "against".

- The **EV export** scripts are written such that the export includes all data categories available (making for lots of columns not used in our analyses), the logic is that if we ever want to do analyses using additional data columns, they'd already be in the exported files thereby eliminating the "version control" issue (i.e. any additional analysis would be done with equivalently processed data)
  - REMINDER: the number of columns exported, when all columns is selected in EV, changes from EV version to EV version. So processing the data based on column header names rather than column position is vital to keep consistency in the data.
  - REMINDER: you need to select ALL columns in EV (it's not the default)
    - See the Export tab in the EV File Properties
  - Be sure to include the ALL columns setting in the EV template to ensure that those settings are in place for all EV files created for the project.
- **Minimum Surface Exclusion Line** was set to 1.7 m for all surveys (historical and contemporary) except for May 2012 which was set to 1.5
- The **deadzone** for a 7^{o} 120 kHz echosounder with a 1.024 ms pulse length operating in seawater is 0.8 m at 10 m depth and 1.0 m at 100 m depth. When defining the bottom line within Echoview, a minimum of a 1-m stepback is recommended in order to exclude the deadzone from the data used for analyses.
- Make sure the **bottom line (and top line)** has no gaps in EV. Otherwise, "data" gets included in the automated exports from below bottom (or above top). (See next comment.)
- The early versions of the **Echoview template** introduced a "**smoothing**" to the bottom (or top) line after bottom (or top) edits. This generated two challenges (1) the smoothing algorithm commonly introduced erratic behavior in the bottom (or top) line and (2), the result commonly introduced data to the analysis that in truth we were trying to exclude (a "+1" depth bin - "data" that we don't want)
  - In August 2018, UMaine sent FORCE a revised Echoview template that incorporates the smoothing before the manual edits of the bottom and top lines thereby eliminating the extraneous "+1 m" "data"

Cautions – Datasets

- Historical (Melvin)
    - datasheets and raw files are recorded in GMT (whereas contemporary (FORCE) are recorded in local time)
        - LPM proofed GMT by comparing tide height change for each of the Melvin surveys (2011-2012) per echogram (assuming recorded time was GMT and then assuming recorded time was Local) against predicted tide height change, and found that when assuming GMT, the tide height change per echogram closely corresponded to predicted tide height change (within +/- 2 m) whereas the tide height change differed substantially from predicted tide height change when assuming the recorded time was local (up to +/- 16 m)
- Contemporary (FORCE)
    - despite settings of 4 pings per second, apparently only 2 pings per second are being recorded
        - make sure any reporting reports the actual pings-per-second achieved

Cautions – Data Collection (Simrad) and Data Processing (Echoview) Software

As with the release of all new hydroacoustic scientific instruments such as the EK80, there is a lag between the release of the instrument and the time at which the research community has vetted its operation, including the operation of its data collection and calibration software and resultantly the updates required of the processing software. Therefore, it is not unexpected that new releases including corrections and refinements will be forthcoming for the EK80 software provided by Simrad and Echoview. Please keep your software up-to-date with the latest releases.

APPENDIX C: Data Export and Processing Scripts

Please review the extensive comments in the script files for more detail than is listed here.

Export Scripts

20190225 Script_exportEV8_MASTER run20190225.vbs

Purpose: export the following

a. data
b. metadata
c. jpgs
d. archiving files from Echoview

Output:

a. .csv data files of the full suite of Echoview variables in a variety of user-defined cell sizes (e.g. 20-m along-shiptrack distance integrated over the full water column). REMINDER: exclude the EV files associated with “along” at this stage.
b. .jpg files of the raw and processed echograms associated with the data contained in the .csv data files
c. .txt files documenting approximately 140 Echoview settings associated with each of the .csv data files (e.g. colorbar settings for the echogram .jpgs among others)
d. .csv files of the line depth of the “turbulence line” and the “bottom line” by which to calculate the proportion of the water column lost to turbulence
e. .jpg of the cruise track over which the data in the .csv files were collected showing S_v mean alongtrack distinguishing regions designated as bad-data/no-data (designates bad-data/turbulence regions) and bad-data/empty-water (designates for passive data region)
f. .evd files: an Echoview format that hardcodes data values based on settings at the time of export. The script is set to export the “data without turbulence” variable (i.e. the variable we use for analysis). Can be used to hardcode an archive of the data in the condition it was at the time of the automated export
g. .ev files: a copy of the Echoview file for archiving. Gives you a fully operational copy of the Echoview file and all its variables to archive in its state at the time of the export in case future edits are required. Leaving you the original EV file for “exploring” if necessary. (i.e. the archived EV file leaves no question as to which EV files (and therefore settings) generated the exports)

Processing Scripts

SCRIPT2019 AppendCSV_20m.R

Purpose: To prepare for merge with metadata ("datasheet"),

a. import and append the "20m dataset" transect .csv data files exported from Echoview for a single survey
b. populate "Transect" metadata column (i.e. "GR1_N0A" which is read from the names of each of the imported .csv files)
c. populate DateTime columns as yyyymmdd.decimaltime to mitigate the .csv issue where "hour" gets dropped from the time column
d. perform critical data quality control tests (failed status requires script termination)
e. populate the data quality checks variable

Output:

a. data file of the appended .csv data files for the survey
b. .csv file holding the data quality checks for review
c. archive the workspace

SCRIPT2019 MergeCSV_20m.R

Purpose: To prepare for analyses by associating (merging) metadata with S_v thereby placing S_v in a meaningful ecological context

a. import appended survey data file (.csv) and metadata file ("datasheet" .csv)
b. populate additional data file columns (e.g. linear version of S_v, FishPresence, etc.)
c. using start and end date-time as the parameters by which to match metadata to data file, populate the metadata columns in the data table
d. perform data quality control tests (e.g. test for failed matches, for "along" (excluded from these analyses, and for misspellings coming from the user produced metadata (datasheet) file

Output:

a. count of failed matches signifying that start and end times in the metadata files ("datasheet" .csv) may need attending to (script prints instructions on how to find the data lines that failed to match with the metadata start/end times)
b. count of the data bins that matched to "along" (rather than CLA or reference) signifying that start and end times in the metadata file may need attending to (check_along variable holds the detail by which to find those lines that matched to "along")
c. .csv file of the survey data merged with metadata (diel state, tide phase, study area, "with/against", etc.)
d. print to Console the unique values of a variety of columns within the newly merged dataset for user's review (1) for confirmation that the merged data is as expected (i.e. all from the correct survey, etc.), (2) for missing data (e.g. no "day" etc.), or (3) for misspellings
e. .csv file holding the merged data for the survey
f. archive the workspace

SCRIPT2019 AppendAlldata.R

Purpose: To append the merged survey data onto the alldata .csv file

a. import newly appended survey data and the alldata .csv files
b. append survey data to alldata

Output:

a. "alldata" .csv with latest merged survey data appended at the bottom
b. archive the workspace

SCRIPT2019 20mBinAnalysis.R

Purpose: Generate data visualizations and execute analyses using S_v and Presence:Absence data

Output:

a. plots ready for examination and/or saving as .jpg, .png, .tiff, or .pdf
b. analytical results printed to the Console and available as a dataframe for export

Processing Notes:

a. some data quality checks are performed here for which output is printed to Console rather than to a file. These should be reviewed at time of running the script.

APPENDIX D: Additional Files Transferred with this Report

Files referenced in Appendix A

- 20170519 Datasheet_Melvin WORKING20190115LPM.xlsx
- 20190115 AllDataParseToMonthly_MelvinONLY_20mBinsFullWaterColumn.xlsx
- MelvinGridNOTES.xlsx
- SimradFilesPerTransect.xlsx

Merged and Appended "alldata" File for Import into R Processing Scripts

- 20190304_alldata_Grid_20mBinFullWaterColumn_thruSurvey09_201708.csv

Files Documenting Calculations for Significance Between Categories that do not include the baseline ("reference") category in presence:absence analyses

- 20190501_significanceCalcs.xlsx

Data Files

- Echoview post-processing files for historical and contemporary surveys
- Echoview calibration files (.ecs) for historical and contemporary surveys
- Echoview export (data) files for historical surveys
- Echoview export (data, metadata, and archiving) files for contemporary surveys
- metadata ("datasheet") files for historical and contemporary surveys

APPENDIX E: Glossary and Abbreviations

ANOVA: analysis of variance
a statistical procedure used to analyze the differences among group means in a sample

CLA: Crown Lease Area
located in Minas Passage

CLD: compact letter display
a compact presentation of the results of multiple comparisons
although the name references "letter", R uses numbers to indicate groupings

dB: decibel

3 dB: a 3 dB change in S_v (the log form) is equivalent to a doubling or halving in linear terms
10 dB: a 10 dB change in S_v (the log form) is equivalent to a change by an order of magnitude in linear terms
20 dB: a 20 dB change in S_v (the log form) is equivalent to a change by two orders of magnitude in linear terms

EMM: estimated marginal mean
when there are pronounced differences in the number of observations, the estimated marginal mean is a way to estimate what the mean would be if the number of observations were balanced

EV: Echoview Software
industry-standard software used to post-process hydroacoustic data preparing it for analyses

GLM: Generalized Linear Model implemented in R for binary logistic regression

GMT: Greenwich Mean Time
used interchangeably with UTC (Coordinated Universal Time)

R: an integrated suite of software facilities for data manipulation, calculation, and graphical display

S_v: mean volume backscattering strength
the log form of a fundamental hydroacoustic measurement
unit: dB re 1 m^{-1}
a proxy for relative fish density
relationship to s_v is shown in Equation 1

s_v: **volume backscattering coefficient**
the linear form of a fundamental hydroacoustic measurement
unit: m^{-1}
relationship to S_v is shown in Equation 1

TISEC: **Tidal In Stream Energy Conversion device**
an device engineered to convert tidal energy to electricity