Title
A comparison of acoustic and visual metrics of sperm whale longline depredation

Journal
The Journal of the Acoustical Society of America, 135(5)

ISSN
0001-4966

Authors
Thode, Aaron M
Wild, Lauren
Mathias, Delphine
et al.

Publication Date
2014-05-01

DOI
10.1121/1.4869853

Peer reviewed

A comparison of acoustic and visual metrics of sperm whale longline depredation

Aaron M. Thode[a)]
Marine Physical Laboratory, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093-0238

Lauren Wild[b)]
Sitka Sound Science Center, 834200 Lincoln Street, Sitka, Alaska 99835

Delphine Mathias[c)]
Marine Physical Laboratory, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093-0238

Janice Straley
University of Alaska Southeast, Sitka, Alaska, 99835

Christopher Lunsford
Auke Bay Laboratories, NOAA, 17109 Point Lena Loop Road, Juneau, Alaska 99801

(Received 7 June 2013; revised 23 January 2014; accepted 6 March 2014)

Annual federal stock assessment surveys for Alaskan sablefish also attempt to measure sperm whale depredation by quantifying visual evidence of depredation, including lip remains and damaged fish. A complementary passive acoustic method for quantifying depredation was investigated during the 2011 and 2012 survey hauls. A combination of machine-aided and human analysis counted the number of distinct "creak" sounds detected on autonomous recorders deployed during the survey, emphasizing sounds that are followed by silence ("creak-pauses"), a possible indication of prey capture. These raw counts were then adjusted for variations in background noise levels between deployments. Both a randomized Pearson correlation analysis and a generalized linear model found that noise-adjusted counts of "creak-pauses" were highly correlated with survey counts of lip remains during both years (2012: $r(10) = 0.89$, $p = 1e\text{-}3$; 2011: $r(39) = 0.72$, $p = 4e\text{-}3$) and somewhat correlated with observed sablefish damage in 2011 [$r(39) = 0.37$, $p = 0.03$], but uncorrelated with other species depredation. The acoustic depredation count was anywhere from 10% to 80% higher than the visual counts, depending on the survey year and assumptions employed. The results suggest that passive acoustics can provide upper bounds on depredation rates; however, the observed correlation breaks down whenever three or more whales are present.
 [http://dx.doi.org/10.1121/1.4869853]

PACS number(s): 43.80.Nd [ANP] Pages: 1–15

I. INTRODUCTION AND BACKGROUND

A. Sperm whale depredation and SEASWAP

In the eastern Gulf of Alaska (GOA) a demersal longline fishery for sablefish (*Anoplopoma fimbria*) occurs about 8.5 months a year. Sablefish (also called blackcod and butterfish) reside on the continental slope, and most commercial longliners operate in water depths between 400 and 1000 m.

Sperm whales (*Physeter macrocephalus*) are a cosmopolitan species distributed throughout the world's oceans (Whitehead, 2003; Barlow *et al.*, 2008; Gosho *et al.*, 1984; Rice, 1989). While females and immature individuals generally reside at low latitudes, adult males also travel and forage at higher latitudes (Jaquet, 1996; Whitehead *et al.*, 1992; Teloni *et al.*, 2008). In the U.S., these whales are listed as an endangered species, and are also "vulnerable" on the International Union for the Conservation of Species (IUCN) red list, but their current population in the North Pacific is unknown.

Sperm whales are known to take fish from fishing gear, a behavior known as "depredation." Although quantitative data are limited, sperm whale depredation appears to be increasing worldwide (Ashford *et al.*, 1996; Capdeville, 1997; Nolan and Liddle, 2000; Purves *et al.*, 2004). Perez (2006) estimated that marine mammal depredation on the combined longline fisheries in Alaska caused a loss of about 2.2% of the total fishery groundfish catch during 1998–2004, based on visual evidence of torn or partial fish.

Since 1987 the Alaska Fisheries Science Center has conducted annual longline surveys of the upper continental slope, referred to as domestic longline surveys (Sigler and Zenger, 1989). The domestic longline survey began annual

[a)]Author to whom correspondence should be addressed. Electronic mail: athode@ucsd.edu
[b)]Present address: Scottish Oceans Institute, School of Biology, University of St. Andrews, KY16 8LB, United Kingdom.
[c)]Present address: Grenoble Images Parole Signal Automatique (GIPSA-lab) CNRS: UMR5216 - Université Joseph Fourier - Grenoble I - Université Pierre Mendès-France - Grenoble II - Université Stendhal - Grenoble III - Institut Polytechnique, Grenoble, France.

sampling of the Gulf of Alaska in 1987, biennial sampling of the Aleutian Islands in 1996, and biennial sampling of the eastern Bering Sea in 1997. The domestic survey also samples major gullies of the Gulf of Alaska in addition to sampling the upper continental slope. At present the survey is divided into five "legs" that cover five distinct geographic regions in the Gulf of Alaska. Along with the stock assessment data, the survey also gathers data related to depredation: Counts of lips or other unidentifiable remains, as well as counts of damaged fish, identified to species. A previous study reviewed data from the domestic surveys from 1999 to 2001 for all sets with sperm whales present; they compared sets with and without physical evidence of depredation and found a 5% lower catch rate in sets with depredation evidence (Sigler *et al.*, 2008).

B. Background on sperm whale foraging and acoustic behavior

A deep-diving species, sperm whales regularly descend to depths greater than 400 m for periods ranging between 30 and 45 min and rest at the surface for periods ranging between 5 and 10 min (Mullins *et al.*, 1988; Watkins *et al.*, 1993; Jaquet *et al.*, 2000; Wahlberg, 2002; Drouot *et al.*, 2004; Watwood *et al.*, 2006). The few data available from higher latitudes indicate shallower dive depths than what is measured in temperate or tropical latitudes (Whitehead *et al.*, 1992; Teloni *et al.*, 2008).

Sperm whales are vocally active underwater, and during a single dive, an individual can generate thousands of impulsive sounds, called clicks (Goold and Jones, 1995; Worthington and Schevill, 1957; Madsen *et al.*, 2002; Wahlberg, 2002). In the Gulf of Alaska (GOA), click sounds from sperm whales have been detected throughout the year on bottom-mounted recorders, revealing a year-long presence in the region (Mellinger *et al.*, 2004).

Another distinctive acoustic feature of sperm whales is the existence of "creak" (or "buzz") sounds, a sequence of pulses produced at a rate of 10 per second or faster (Madsen *et al.*, 2002), and often characterized by a decrease in the pulse interval and (occasionally) amplitude over the 5-to-10 s duration of the sound (Whitehead and Weilgart, 1990; Whitehead, 2003). A typical creak rate during dives is about 10 creaks per hour per animal (Miller *et al.*, 2004). Previous bioacoustic tagging work on sperm whales has shown that most creaks are associated with prey capture attempts (Miller *et al.*, 2004; Watwood *et al.*, 2006). Creaks are often followed by a few seconds of silence before the animal resumes "usual" clicking, defined here as a "creak-pause" event.

C. Observations leading to present study

In 2003, the Southeast Alaska Sperm Whale Avoidance Project (SEASWAP) was created to investigate this issue with the long-term goal of reducing depredation. A collaborative study between fishermen, scientists, and managers, SEASWAP works with both the coastal fishing fleet and the federal sablefish survey to collect various quantitative data on longline depredation. Using the shape of the flukes as a unique identifier, SEASWAP has found that at least 106 individual sperm whales have been involved in depredation.

In May 2006, SEASWAP deployed an underwater videocamera on a longline during an active haul, with the cooperation of the F/V Cobra. The resulting video and audio (Mathias *et al.*, 2009) revealed that a sperm whale was making creak sounds while depredating fish, even under good visual conditions. The whale managed to remove the fish from the hook without leaving behind any visual evidence. Thus, the idea arose that acoustic monitoring for creak sounds might provide a metric of depredation activity, complementing standard methods of estimating depredation rate, which involve counting fish remains on a hook, a time-intensive and expensive process that may undercount depredation rates if fish are removed from a hook completely.

A more recent SEASWAP study used bioacoustic tags to confirm that creak rates produced by individual sperm whales during depredation conditions could exceed creak rates during natural foraging conditions, sometimes by as much as a factor of 3 (Mathias *et al.*, 2012). Furthermore, the tagging study found that the relative fraction of creaks that were followed by pauses (a "creak-pause fraction") was quite low in the Gulf of Alaska tagging sample, when compared with published reports from the Gulf of Mexico and Ligurian Sea.

There are several possible interpretations of creak-pause events. One is that these intervals are used to recycle air within the sound production system (Wahlberg, 2002). Because creaks generally have lower received levels on hydrophones than the usual clicks, it may also be possible that some clicks at the end of a creak become masked by noise, creating a false impression of silence. Still another interpretation is that certain individuals are more likely to be silent after generating creaks. Finally, the silences may be indicative of prey capture. A substantial literature has argued that sperm whale creaks are echolocation signals (Gordon, 1987; Jaquet *et al.*, 2001; Wahlberg, 2002; Madsen *et al.*, 2002), and periods of time where creaks are detected have been described as prey capture attempts (Miller *et al.*, 2004; Watwood *et al.*, 2006). Miller *et al.* (2004) found that the majority of creaks produced by sperm whales in the Ligurian Sea and the Gulf of Mexico are followed by pauses of about 5 s. Analogous signals, with pauses, have been observed in other species. For example, laboratory studies on bat echolocation have found that post-buzz pause durations were longer after successful captures (e.g., Surlykke *et al.*, 2003). Beaked whales and porpoises often pause for less than a second when creaking or buzzing (Johnson *et al.*, 2009; DeRuiter *et al.*, 2009).

If the presence of extended silences following creaks is a valid indication of prey capture, then perhaps successful prey capture attempts can be distinguished from unsuccessful ones on the basis of acoustic data. The lower creak-pause fraction measured from Alaskan sperm whales may suggest that these animals generally had lower prey acquisition success rates than whales in the Gulf of Mexico or Ligurian Sea (Watwood *et al.*, 2006); i.e., the Alaskan whales required more creaks per capture, since they include fish as a natural part of their diet. Thus, as the current research began it was recognized that distinguishing these so-called "creak-pause" sounds from creak events may be an important step in quantifying an acoustic depredation metric.

This paper describes how passive acoustic measurements of creak and creak-pause rates around the NOAA sablefish survey can be adjusted for variations in background noise levels between sets to yield acoustic metrics of depredation that can be highly correlated with certain visual estimates of depredation. Section II describes the logistics of the sablefish survey, the passive acoustic equipment, deployment configurations, acoustic signal processing methods, and statistical analysis procedures for both the survey depredation databases and acoustic measurements. The section focuses on the use of correlograms to facilitate rapid review of creak signals in the acoustic data, as well as the use of randomized correlation trials and a simple generalized linear model to conduct statistical comparisons between the visual and acoustic depredation metrics. Section III evaluates the performance of correlogram-based processing and displays the results of the correlative statistical analysis for field data collected in 2011 and 2012. Section IV discusses the findings, examines the implications, and suggests future work and further improvements for both the data collection and analysis.

II. DATA COLLECTION AND ANALYSIS PROCEDURE

A. Relevant facts about the sablefish survey

The NOAA sablefish survey takes place annually. In 2011 the F/V Ocean Prowler and in 2012 the F/V Alaskan leader were chartered to deploy demersal longline sets at a total of 65 stations, or geographic locations. At each station two "sets" of gear were deployed, roughly in tandem. Each set consisted of 80 "skates" of gear, with a skate being 100 m long with 45 hooks spaced 2 m apart. Each set was 8 km long, with 1 km or less separation between the end of the first set and beginning of the second set. The sets were hauled on a regular schedule, with the first set generally hauled beginning at around 09:30 local time and the second starting around 13:00. Retrieval times generally lasted between 3 and 4.5 h per set. Each set of gear was attached to varying lengths of running line, dependent on the bottom depth, and a flag and buoy array on the surface, providing useful deployment locations for autonomous acoustic recorders (Fig. 1). For every hauled skate, a 100% hook census records the number of baited hooks, damaged hooks, and the number of undamaged and depredated fish, enumerated by species.

B. Passive acoustic recorders and deployment strategy

Several custom-built autonomous acoustic recorders were used to obtain the acoustic data from the NOAA federal sablefish survey. The recorders could be programmed with an internal duty cycle, which was used to minimize the amount of time a given recorder would be acquiring data while not deployed.

In 2011, these "ADIOS" recorders used a Persistor CF2 data acquisition system and a HTI-96 min hydrophone

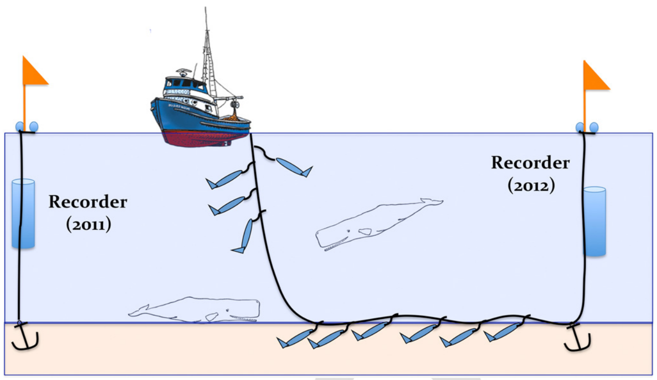

FIG. 1. (Color online) Schematic of how autonomous acoustic recorders are attached to buoy lines of longline fishing gear over a 2-year period. In 2011 a recorder was deployed independently next to each set; in 2012 the recorder was deployed on a buoyline of the set itself.

(High-Tech Inc.) with a sensitivity of -172 dB re 1 V/μPa. The calibration values were obtained from High Tech Inc. measurements of the individual sensors before shipping. These laboratory values were checked in the field by comparing the received level of a FM sweep of known source level at known range in deep water with theoretical predictions during previous field seasons (Thode *et al.*, 2010). After a 26-fold analog voltage amplification (i.e., 28 dB gain) the data were written to a 0–2.5 V range A/D converter as 16-bit samples. Thus, the system could record peak-to-peak impulses of 153 dB re 1 μPa (pp) without clipping. The Persistor system would log data at 50 kHz sampling rate to a 4 Gb flash memory card for 10 h, then stop sampling for 2 h to transfer the data to a hard disk.

In 2012, a new type of custom-built recorder ("ADIOS2") eliminated the hard disk and wrote the data directly and continuously to four 32 Gb flash memory cards, using a 100 kHz sampling rate and an internal amplification of 20 (26 dB).

In 2011, two recorders were deployed at each station, one for each set, during three legs of the sablefish survey. Each recorder was deployed on an extra anchored buoyline deployed roughly 1.6 km off to the side of the midpoint of a given set (Fig. 1). In 2012, each recorder was deployed directly on a buoy line connected to the end of each set, during the same three legs as in 2011. The deployment depths during both years were standardized at 100 fathoms (182 m).

C. Machine-aided creak detection

Before this project began, SEASWAP acoustic data were traditionally reviewed for creak sounds by listening to 30 s of data at a time while simultaneously viewing spectrograms of the data. The spectrogram was useful for noting times when a whale's "usual" click rate started to segue into a creak, but a spectrogram would generally not reveal a creak's presence over a minute timescale [Fig. 2(a)]. A review of a 3-h haul generally took 6 to 12 h, depending on the quality of the recording and the number of whales present.

The current study conducted 57 deployments that collected acoustic data during times of sperm whale depredation, yielding over 170 h of raw data to review. In 2011, the

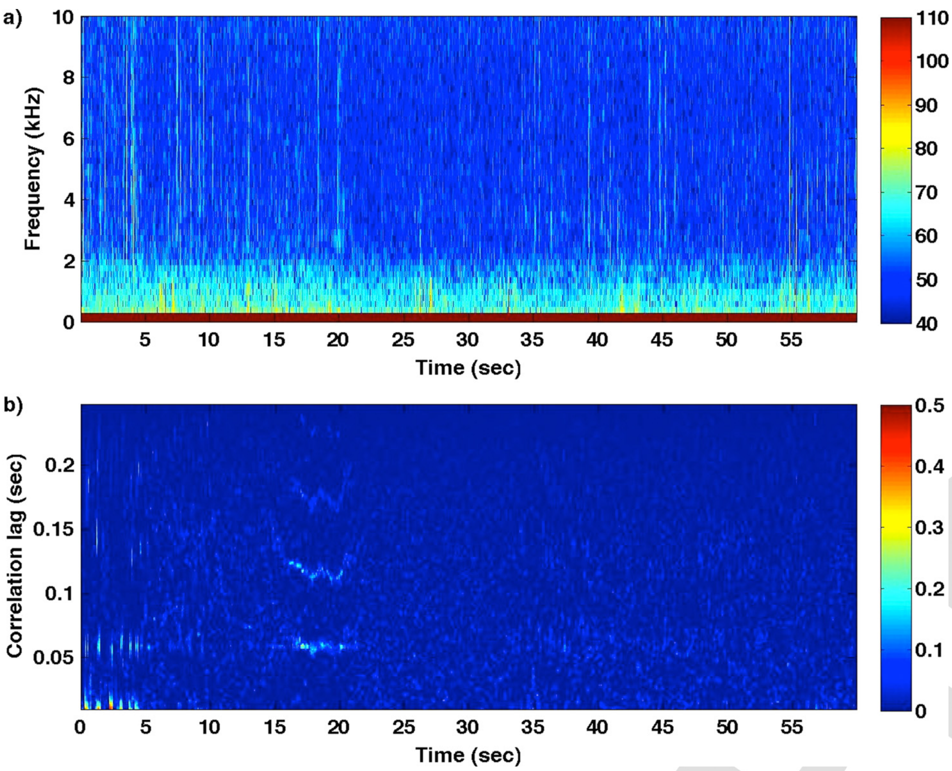

FIG. 2. (Color online) (a) 60 s from Unit 3, July 28, 2011, starting at 15:06:01. FFT length is 256, overlap is 75%. (b) Correlogram of same data shown in Fig. 2, using parameters described in text, and a frequency range of 8–9 kHz. The figure shows an echolocation creak taking place between 15 and 20 s with a pulse interval of 60 msec.

data were analyzed manually by listening to the entire data record, but in order to expedite this analysis stage, a new "correlogram" method for displaying the acoustic data was developed, and this machine-aided procedure was used to analyze the acoustic data in 2012.

A spectrogram provides a poor way for visually detecting a rapid series of pulses, even when the time window used for an individual Fourier Transform (FFT) is smaller than the timing between pulses. In order to see the individual pulses, a spectrogram would have to show only a few seconds of data at a time, a daunting procedure when reviewing days of data.

Over a time window of a second or less, the echolocation pulses in a creak are relatively evenly spaced in time, which suggests that if the spectrogram is autocorrelated along the time axis, and then integrated along a set of frequency bins where creaks are most likely to be detected, a new type of image of the acoustic data can be created, called a "correlogram." Specifically, a spectrogram is first created from 1 min of data, using an FFT length of 256 samples for 50 kHz data (2011) or 512 samples for 100 kHz data (2012), using 50% overlap between subsequent FFT samples. The FFT size was chosen to be longer than the duration of a single creak click but shorter than the minimum possible interval ($\sim 1/30$ of a second) between creaks. FFT frequency bins above 12 kHz are rejected, as creaks have a low probability of being detected above that frequency whenever sperm whales are more than a few hundred meters away from a moored autonomous recorder.

The resulting spectrogram is divided into 1 kHz frequency bands, overlapped by 500 Hz. The lowest-frequency band covers 2000–3000 Hz, the next band covers 2500–3500 Hz, etc. This frequency division was chosen because the minimum bandwidth of a weak creak is about 1 kHz.

Each bandlimited spectrogram is then split into 0.25 s time segments, an interval over which the pulse rate of a creak is expected to remain relatively constant. Each spectrogram segment thus contains 97 time bins (2.56 msec per bin) and five frequency bins (or 485 time/frequency "pixels"). For each of the five frequency bins in a given segment, the autocorrelation function of the 97 time bins is computed. Each autocorrelation is normalized to have a value of 1 for a 0 time lag. The total autocorrelation function for the spectrogram segment is then chosen to be the median value of the five autocorrelation functions (computed for the five frequency bins).

The next spectrogram time segment begins 24 time bins (62.5 msec) after the start of the previous time segment, creating 75% overlap between spectrogram segments. When the median autocorrelation values from these overlapping spectrogram segments are combined, a "correlogram" for a particular frequency band can be constructed [Fig. 2(b)], with one axis representing time (with 62.5 msec resolution), and the other representing time lag. A pixel at row i and column j represents the median correlation between a bandlimited spectrogram power spectral density at time t_j and the power spectral density i time bins earlier. Each correlogram generated by each frequency band can be sub-plotted into a single figure, allowing a complete minute of data to be viewed as a single image. As Fig. 2(b) suggests, a 1-min correlogram instantly reveals the presence of a creak that is otherwise invisible in a 1 min spectrogram, and the value of the correlation lag time yields the creak pulse interval. Even a relatively inexperienced manual analyst can review hundreds of such images in a relatively short time, and then confirm via listening that the time points in question are creaks.

In order to evaluate whether correlograms "captured" all viable creaks, three complete manual reviews that had been conducted on three fishing hauls during the 2011 season were reanalyzed with the correlogram method: both sets on July 14, and the second set from July 28. The first two sets had one or two whales present and were known to contain fairly low numbers of creaks (a fairly typical situation), while the last set had three whales present and had large numbers of creaks. The second set also had large amounts of engine noise. A second analyst then conducted a second review using the correlogram plots, without prior knowledge of the earlier manual analysis detection times. All potential creak detections and durations were noted and then the data were reviewed aurally to confirm creak presence. Thus, the analyst only needed to listen to a small fraction of the acoustic data. The full manual review was then compared with the expedited review. If two creaks lay within three seconds of each other, they were counted as a single creak, and two detection times that were within three seconds of each other were counted as the same detection.

All acoustic analyses, whether with or without correlograms, would use human analysts to confirm the presence and duration of creak events, and to classify the creak as either a "creak only" or a "creak-pause." The latter was defined as a creak event, subsequently followed by 3–4 s of silence from that particular whale. Creak-pauses could be reliably identified in the data only when three or fewer whales were present; when more whales were present and vocalizing simultaneously a human listener had difficulty recognizing whether a pause was present after a given creak, a restriction with implications for later analysis.

D. Correcting raw acoustic detection rates for background noise variations

In addition to the raw counts of creak and creak-pause events discussed in Sec. II C, measurements of the background noise levels were needed for all hauls, because the raw creak detection rate is a function not only of depredation activity but also of detection range, which is a function of background noise level. Therefore, in order to compare relative depredation rates between any two hauls, the raw creak detection rates need to be adjusted for differences between background noise levels between those two hauls. To estimate a background noise level appropriate for depredation analysis, the integrated acoustic power between 3 and 9 kHz was computed every 40 msec, for every acoustic recorder deployment. This bandwidth was chosen because the vast majority (over 95%) of sperm whale creaks are detectable over this band, and so the correlogram review is conducted over this band as well.

Instinctively, one wishes to convert this series of instantaneous measurements into a longer-term "average" background noise level, measured over 1-min intervals. However, taking a simple average of these measurements over a 1-min window yields an inappropriate measure of background noise, because the impulsive clicks of sperm whales are generally quite intense, and so the spectral properties (and received levels) of sperm whale sounds would dominate a simple average of background noise levels. This result would be inappropriate, because creak sounds can be detected during the intervals between sperm whale clicks, so the intensity and spectral characteristics of the sperm whale clicks is irrelevant to the detectability of the creaks. A more appropriate measure of this diffuse background noise is the 10th percentile of the cumulative distribution of instantaneous background levels, accumulated over 1-min intervals, since this percentile excludes all sperm whale signals (which exist in the higher percentiles) while rejecting artificially low values generated by potential acoustic dropouts caused by banging hydrophones. Using the 10th percentile is equivalent to stating that sperm whale clicks and creaks never occupy more than 90% of the monitoring period; at least 6 s of acoustic data in every minute is generally free from sperm whale sound contamination.

Figure 3 plots the values of this 10th percentile over a 3-week period during 2011, as obtained from both instruments. Since the instruments were on a fixed duty cycle, sometimes they recorded data during days when they were not deployed. Figure 3 shows examples of such spurious recordings between August 7 and August 10. The double line visible during such times arises from the fact that one recorder is being stored in a slightly noisier environment than the other.

Note that background noise levels in the water varied between 87 and roughly 96 dB re 1 uPa. In particular, between July 28 and 30 background noise levels were several dB lower (88 dB) than the other deployments (92 dB). Under such low-noise conditions, one would expect the detection range of the instruments to be greater than under the typical (92 dB) conditions. Therefore, all other circumstances being equal, the instruments would be expected to detect more creak activity under low noise conditions. Thus an essential final task in creating an acoustic depredation metric is adjusting the raw creak detection rates for variations in background noise level (and thus detection range).

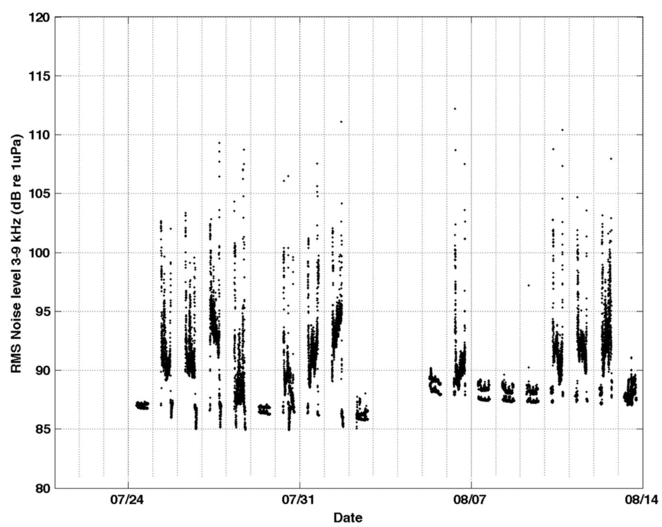

FIG. 3. The 10th percentile of RMS integrated background noise levels between 3 and 9 kHz, for all deployments between July 21 and August 14, 2011. Percentiles are computed over 1-min intervals. Data includes both autonomous recorders deployed simultaneously. Values shown between August 7 and 10 are examples of measurements that arise when the instruments are recording while being stored on the vessel and not deployed in the water.

This adjustment requires two steps: (1) defining a "reference" background noise level, and (2) selecting a model for how sound attenuates with distance in the survey environment (a propagation model).

For the first step the background reference level, integrated between 3–9 kHz, was chosen to be 93 dB re 1uPa. There were two reasons for selecting this value as a reference level. First, it was roughly the median level encountered by the deployments shown in Fig. 3. Second, this reference level corresponds to sea state 3 (Wenz, 1962). Previously supported NOAA research used a vertical array to track sperm whales in the Gulf of Alaska, and estimated that the detection range of creak sounds should be 5–6 km during this sea state (Mathias *et al.*, 2013). During 2011, the recording instruments were deployed about 1.8 km from the midpoint of a 8 km set, so a detection range of 5–6 km (background noise level of 93 dB re 1uPa) would be optimal, in the sense that the recorder would be able to detect creaks along the entire set, but would not be able to detect potential creaks generated by unrelated foraging activity at greater ranges. The 2012 deployments, which attached recorders at the end of a set instead of the middle, would still cover the last 3/4 of a haul under sea state 3.

The propagation model for the noise adjustment was chosen to be the so-called "spherical spreading model" (Richardson *et al.*, 1995), where the square-pressure of a discrete acoustic signal from a compact source is assumed to fall off with slant range r as r^{-2}. Detailed simulations of acoustic propagation in this region indicate that this simple model is valid to a 2 km range, and is roughly applicable at greater ranges, where the intensity falls off as $r^{-1.7}$ (Mathias *et al.*, 2013).

From these assumptions the following expression can be derived (Ponce *et al.*, 2012) for adjusting the raw creak count at set j, $C_{j,raw}$, to a noise-adjusted (NA) creak count $C_{j,adj}$

$$C_{j,adj} = C_{j,raw} \left(\frac{N_{j,med}}{N_{ref}} \right)^{3/2}. \quad (1)$$

Here $N_{j,med}$ is the median value of the 10th percentile noise encountered during the set (i.e., the median of the values in Fig. 3 over 3-h windows).

There are several important points regarding Eq. (1). First, it assumes that depredation behavior (and thus creak generation) is evenly distributed over a volume surrounding the sensor, when averaged over the time of the haul. In reality depredation behavior may be restricted to certain depths, which means that the factor of 3 in the exponent of Eq. (1) should be replaced by a 2, as was the case in (Ponce *et al.*, 2012). Second, note that the practical effect of the formula is to reduce the creak count on quiet days (low $N_{j,med}$) and increase the creak count on noisier days (high $N_{j,med}$). Third, note that N is expressed in *linear* units, and not dB units, so the adjusted creak count is a *sensitive* function of background noise. For example, a 3 dB change in background noise levels corresponds to a factor of 2 change in absolute noise level, which would change the raw creak count by a factor of 2.8. Clearly, this adjustment has limits; if one detects a single creak in the midst of a typhoon, one cannot conclude 1000 creaks were actually present. However, the relatively small variations in the background noise environment shown in Fig. 4 allow the multiplicative factor Eq. (1) to be restrained to reasonable limits.

Finally, note that the selection of a different noise reference level will change all the noise-adjusted creak counts for all sets by the same multiplicative factor, regardless of the actual ambient noise levels measured at each set. The relevance of this fact is that if a time series of these noise-adjusted counts is correlated with another time series, the normalized correlation coefficient will not depend on the choice of reference noise level in Eq. (1). Thus, the overall conclusions of the statistical tests discussed in the next section are not affected by the choice of the reference noise level. However, the magnitude of the depredation count on a set, as measured by acoustics, will be a function of the reference background level chosen.

E. Visual survey database analysis

During the federal sablefish survey detailed records were kept of what was captured by each 45-hook skate in a set. A subset of a database of these records, which covered 90 hauls across 45 geographic stations, was provided to SEASWAP. The database consolidated all the counts by skate and species caught; however, the time at which each individual skate was hauled was not available. Therefore, in order to facilitate a reasonable comparison between the acoustic and visual estimates of depredation, the total depredation rates per skate were combined to yield the depredation count per set.

Every set of the visual survey was assigned a unique haul number, and the analysis of the survey database began by flagging all records associated with a given haul number. Each line in the database was associated with a particular species on a particular skate. Each line also had a "nondepredated frequency" (the number of hooks per skate that had a particular species present) and "depredated frequency" (the number of hooks per skate that showed visual evidence of depredation for a given species). By adding together (across all skates) the combined catch frequencies for all species, plus "ineffective" and "baited" (untouched) hooks, the number of empty hooks could be deduced per set.

The visual records were used to generate four categories of survey depredation counts.

1. "Lips (L)"

For a given haul, database lines with a species code of "Lips or Jaws - Whale Predation" had their 'depredation frequencies' added together to yield the number of shredded lips observed per haul. These remains could be from any fish species.

2. "Sablefish (S)"

Similar to "L," but using the species code for "Sablefish." This category included visual evidence of damage to sablefish.

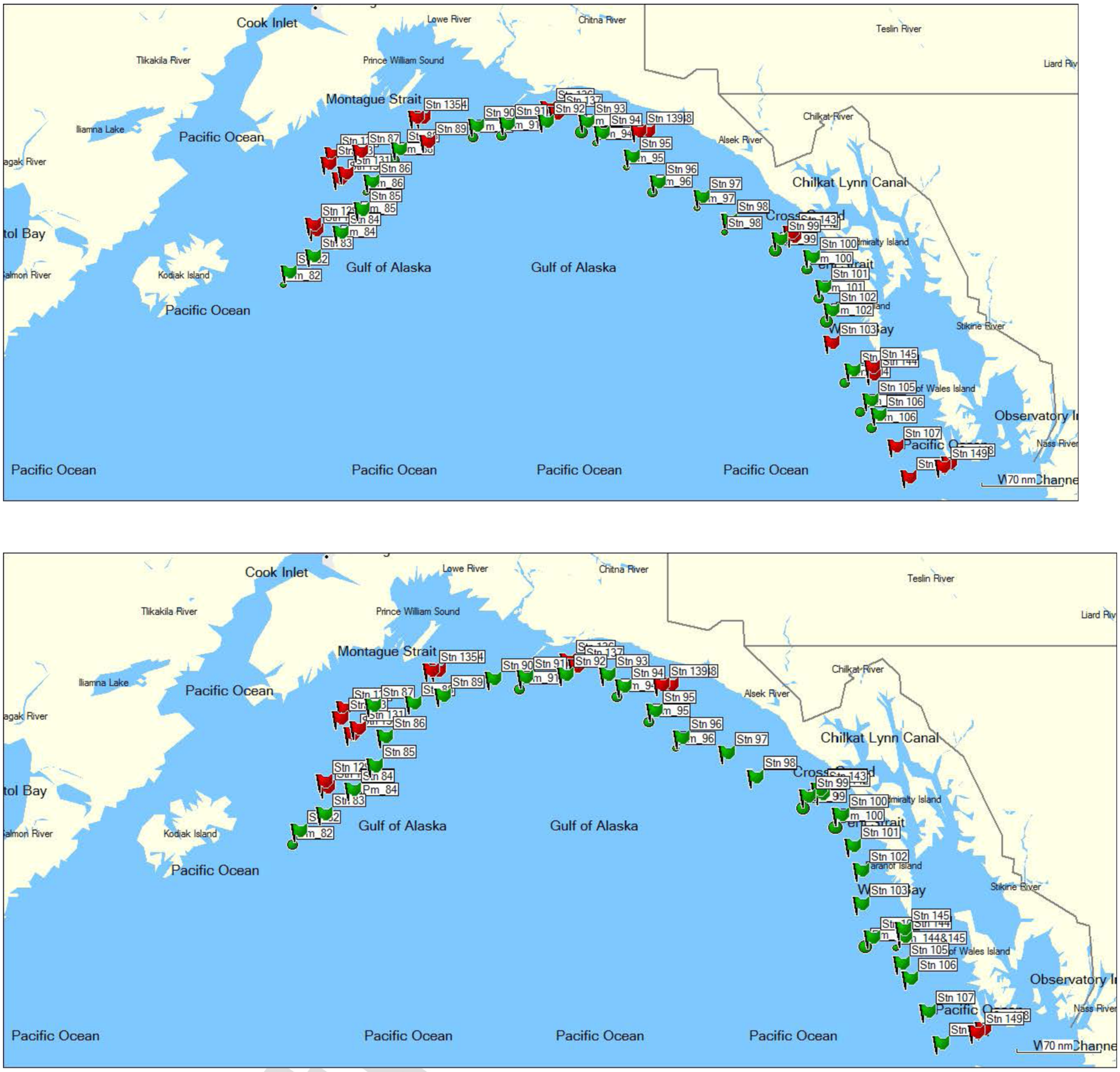

FIG. 4. (Color online) Summary of acoustic deployments along sablefish survey route in 2011 (top) and 2012 (bottom). Light flags indicate survey stations where acoustic data was collected, while dark flags are stations without acoustic data. A light dot at the base of the flag indicates that whales were present at least one of the two hauls conducted at the station; the size of the dot is proportional to the number of whales sighted. (top) The 2011 survey. (bottom) The 2012 survey.

3. *"Halibut & Grenadier (H, G)"*

Similar to "S," but using the species codes for "Giant grenadier," "Pacific grenadier," and "Pacific Halibut."

4. *"Other–Excluding L, H, S, G"*

The depredation frequencies of all species on all skates of a given haul were added together, and then the counts associated with the previous three species categories were removed.

Thus for each year of interest (2011 and 2012), four survey depredation time series could be constructed from the database.

F. Statistical comparisons between visual and acoustic depredation counts

The acoustic and survey depredation counts are compared in several ways. First, to determine the degree of correlation between the two types of depredation estimates, the normalized (Pearson's) correlation coefficient between counts is computed across a group of hauls for a given year

$$r = \sum_i \hat{C}_{i,acoustic}\hat{C}_{i,visual} \Big/ \sqrt{\left(\sum_i \hat{C}^2_{i,acoustic}\right)\left(\sum_i \hat{C}^2_{i,visual}\right)}, \quad (2)$$

where i is summed over the hauls used, and all sequences have had their means subtracted: $\hat{C} = C_i - \bar{C}$. The correlations were conducted using both the raw and noise-adjusted (labeled "NA" in subsequent tables) counts. The acoustic counts may include all creaks counted (labeled "CA" in subsequent tables) or may be restricted to creak-pause events only ("CP"). Correlations were also computed for situations where only one or two whales were sighted during a haul (labeled as "whale-restricted," or WR in the tables). The reasoning behind the WR restriction will be presented in the Results section.

An autocorrelation analysis of sperm whale click counts around the vessel found that autocorrelation of click detection rates (in terms of detected clicks per minute) was negligible at lags of more than ten minutes. The creak counts presented here are measured over 3-h intervals. These acoustic counts were arranged chronologically and autocorrelated to see whether acoustic counts conducted in adjacent geographic regions were correlated with each other. We found that the normalized autocorrelation coefficient between adjacent hauls (hauls conducted the same day) was only -0.2, and then fell to 0 for hauls conducted on different days. We thus concluded that the acoustic counts were uncorrelated between hauls.

Given this lack of correlation between acoustic samples, randomization tests became a viable approach for estimating the significance of each correlation value. 10 000 randomized trials were conducted for each correlation coefficient in Eq. (2), where the order of the acoustic counts was randomized for each trial. A p-value could then be computed by noting the number of randomized trial correlation values that exceed the measured value. Note again that Eq. (2) indicates that the choice of a reference noise level in Eq. (1) will have no impact on the correlation r.

Equation (2) is technically only valid for data that follow Gaussian statistics; as the data collected here have both non-Gaussian distributions and are expressed in terms of a rate (either visual evidence per haul or creaks per haul), a Poisson regression model may provide a more appropriate statistical comparison. Thus, a generalized linear model (GLM) in the form of a Poisson regression was conducted, where the dependent variable was one of four potential acoustic metrics (creak count, creak-pause count, noise-adjusted creak count, and noise-adjusted creak-pause count). The six predictor variables were the four survey depredation metrics discussed in Sec. II E, the number of whales present during the haul, and a categorical variable representing the year of the measurement. The Poisson regression thus had the form of

$$\log \mu_j = \sum_{i=1}^{6} \beta_i x_i. \quad (3)$$

Here μ_j is the expected value of a given acoustic metric j, and β_j is the linear regression coefficient of predictor variable x_i. The logarithms of the non-categorical predictor variables were also tested, to test whether a power-law relationship existed between the visual and acoustic metrics.

One thousand bootstrap fits to the data were conducted to determine the standard error of β. The data were overdispersed, when compared with the theoretical variance of a Poisson model, so the standard error of the coefficient estimates was scaled by the measured variance of the data to provide a conservative estimate of the standard error. A Student's t-test was then applied to determine the probability that a predictor with a true β coefficient of 0 could yield the actual estimated value. The resulting p-values for each coefficient were used to determine which of the six predictor variables (including interaction terms) could be excluded from the final model.

The final comparison between acoustic and visual depredation counts is simply the total number of depredation events observed, summed across all hauls. This value *is* dependent on the choice of a reference background noise level. The lower the reference level chosen, the higher the number of acoustic depredation events will be.

III. RESULTS

A. Summary of acoustic deployments during 2011 and 2012

Figure 4 plots where acoustic recorders were deployed along the sablefish survey stations during 2011 and 2012. Table I summarizes the number of deployments, the number of deployments with depredation present, and the number of deployments analyzed. For the rest of this paper, the terms "haul," "set," and "deployment" are used interchangeably, since one recorder was deployed for every hauled set at a station.

Table I shows that the analyzable acoustic data collected in 2012 was only 1/4 of the viable 2011 data. The reasons for this were that the number of days the survey encountered whales was lower in 2012 (18 days) vs 2011 (38). In addition, the mean number of whales present during 2012 (3) was greater than during 2011 (1.8). There were several encounters during 2012 where 4 or more whales were present. The resulting acoustic "chatter" on the hydrophone was so intense, that while identifying the presence of creaks was still viable, the manual analysts' ability to separate creak-pause echolocation events from creak-only events was compromised. Thus, 8 out of the remaining 18 sets in 2012 could not distinguish creak from creak-pause sounds, and so were excluded from the final acoustic analysis. The 2012 statistical analysis is thus implicitly restricted to cases where relatively few numbers of whales

TABLE I. Summary of acoustic deployments and analysis for 2011 and 2012.

	2011	2012
Sets (hauls)	60	60
Acoustic deployments	43	45
Acoustic deployments with whales present (median whales per set)	38 (1.8)	18 (3)
		1 killer whale
Acoustic deployments analyzed	42	18
Acoustic deployments that could distinguish between creak and creak-pause sounds	42	10

are present (an implicit WR restriction). Thus, the impact of a enforcing a similar restriction (two or fewer whales present during a haul) on the 2011 data was examined.

B. Evaluation of machine-aided creak detection software

Table II shows the results of the evaluation of the correlogram analysis, described in Sec. II C. For the two hauls with relatively typical levels of depredation activity, the correlogram method missed only one creak that the initial manual review had uncovered but found four creaks that the initial review had missed. However, the correlograms also flagged 24 "false alarms," that is, correlogram images that prompted a review that yielded no audible creaks. Thus the correlogram review required 49 reviews of the audio data across the two hauls, of which only 50% were actual creaks. However, to put this in perspective, the far right column of Table II computes what fraction of the acoustic record needed detailed aural review, assuming 20 s of review per correlogram detection. Even with the 50% false alarm rate, the correlogram method only required a review of 4%–8% of the data record, compared with the 100% required by the initial manual review.

During the one haul with substantial depredation (81 creaks over 3 hours), the initial manual review detected only 64 of these creaks, missing 17 that were later picked up by the correlogram analysis. The correlogram analysis had only ten false hits, and missed four creaks that were detected during the manual review. The large numbers of creaks detected meant that the correlogram method required that 21% of the acoustic record needed follow-on analysis.

In summary, out of all three hauls examined in detail, the correlogram method missed 5 out of 107 creaks present, caught 21 creaks that had been missed by the manual analysts, and flagged 34 detections that eventually turned out to be false alarms. Thus in principle, a correlogram review followed by an aural review to strip away false alarms could cut down the amount of time reviewed by nearly 90%. In actuality, the manual analysts reported that the correlogram analysis reduced the review times by about 50%, from 6–12 h to 3–4 h. The reason the actual review took longer than Table II would predict is that the correlogram would often flag weak creaks, which an analyst would often play aurally several times and inspect on a spectrogram to be certain it was a creak.

TABLE II. Comparisons of manual analyses of acoustic data collected from three hauls. One analysis used correlograms, the other did not. The "Total creaks" column lists the total number of unique creaks uncovered by both analyses. The "Manual detected/missed" column shows the number of creaks detected by the initial manual analysis, as well as creaks missed by the analysis but which were detected by the correlogram analysis. The "Correlogram" column shows the number of image-based detections that did not turn out to be creaks, and the number of true creaks that were missed. The "% time" column shows the percentage of the acoustic data that needed to be reviewed by a manual analyst when using correlograms.

Date, Unit	Total creaks	Manual: detected/missed	Correlogram: false/missed	%Time requiring review
July 14, Unit 1	12	12/0	16/1	8
July 14, Unit 3	14	10/4	8/0	4
July 28, Unit 3	81	64/17	10/4	21

C. Randomized correlation coefficient analyses of visual and acoustic depredation metrics

Table III displays the cumulative depredation counts for all methods, visual and acoustic, for 2011 and 2012. The table only includes survey hauls where whales were sighted, an acoustic creak analysis was conducted, and where creak-pause events could clearly be identified by a human listener during review. The four independent survey counts of depredation (Lips, Sablefish, Halibut, and Grenadier, and all other depredation) are shown. Also shown are the results of applying the whale-restricted (WR) conditions to both years.

Table IV shows the Pearson's correlation coefficient between the various survey and acoustic estimates (along with the *p*-value from the randomized permutation trials). Correlations with *p*-values under 0.05 are shown in bold, and correlations with *p*-values under 1.5e-3 (a conservative Bonferroni correction for 32 trials) are shown in bold-italics.

In 2011 both the lips and sablefish survey records significantly correlate with noise-adjusted creak-pause (CPNA) measurements, with *p* values of 0.05 or less. In 2012, only the Lips category was significantly correlated with several acoustic metrics, but the correlation was high and statistically significant, with both the noise-adjusted creak and creak-pause counts correlating with values of 0.85 and 0.89 and randomized *p*-values reaching a respective 4e-3 and 1e-3. Survey counts of halibut and grenadier (H&G) only obtained a *p* value of 0.11, and survey depredation counts that excluded lips, sablefish, halibut, and grenadier became

TABLE III. Total depredation counts from various types of visual and acoustic measurements during the 2011 and 2012 federal surveys, collected from all hauls where sperm whales were visually sighted and where creak-pause events could be discerned from the data by a human listener.[a]

Year	Depredation Count (2011)	Depredation Count (2012)	Depredation Count (2011) WR	Depredation Count (2012) WR
Total Sets	43	10	39	8
Survey Database				
L	68	25	50	17
S	27	23	24	12
H&G	158	5	136	3
Excluding L,S,H,G	62	12	61	12
Acoustic analysis				
All Creaks (CA)	466	233	291	192
Creak-Pause (CP)	290	177	184	141
All Creaks, noise-adj. (CANA)	254	86	164	68
Creak-Pause, noise-adj. (CPNA)	147	65	101	49

[a]Survey definitions: *L:* Unidentified Lips; *S:* Sablefish damage, *H&G:* Halibut and Grenadier damage; *Excluding L,S,H,G:* All other depredation on other species.
[b]Acoustic definitions: *NA:* noise-adjusted; *CA:* All creaks counted; *CP:* creak-pause counts only; *CANA:* all creak counts, noise adjusted; *CPNA:* creak-pause counts, noise adjusted.
[c]Column definitions: WR: datasets restricted to hauls with two or fewer whales present.

ID: ganeshg Time: 08:30 I Path: Q:/3b2/JAS#/Vol00000/140166/APPFile/AI-JAS#140166

TABLE IV. Pearson's correlation coefficient (*p*-value) between various combinations of survey and acoustic counts of sperm whale depredation behavior during the 2011 and 2012 federal surveys, collected from all deployments where sperm whales were visually sighted, and where creak-pause events could be discerned from the acoustic data by a human listener. **Bold** numbers indicate p-values less than 0.05; ***bold-italic*** indicates p-values less than 1.5e-3, the Bonferroni correction for 32 independent statistical tests. *N:* sets analyzed. See Table III for other definitions.

	Visual Depredation Category				
Acoustic Category	*N*	*L*	*S*	*H&G*	*Excluding S, H, G*
2011					
All Creaks (CA)	43	0.09 (0.19)	0.09 (0.26)	0.21 (0.1)	0.04 (0.35)
Creak-Pause (CP)	43	0.16 (0.11)	0.10 (0.25)	0.20 (0.1)	0.006 (0.39)
All Creaks (CANA)	43	0.26 (0.06)	0.23 (0.08)	0.13 (0.19)	0.07 (0.25)
Creak-Pause (CPNA)	43	**0.49 (0.03)**	**0.29 (0.05)**	0.18 (0.11)	0.07 (0.28)
2012					
All Creaks (CA)	10	**0.75 (0.01)**	0.004 (0.41)	0.08 (0.36)	−0.16 (0.63)
Creak-Pause (CP)	10	**0.78 (8e-3)**	0.07 (0.36)	0.15 (0.30)	−0.18 (0.65)
All Creaks (CANA)	10	**0.85 (4e-3)**	0.27 (0.22)	0.26 (0.22)	−0.10 (0.52)
Creak-Pause (CPNA)	10	***0.89 (1e-3)***	0.35 (0.17)	0.33 (0.19)	−0.11 (0.53)
2011, WR					
All Creaks (CA)	39	0.09 (0.16)	0.22 (0.10)	0.16 (0.17)	0.22 (0.12)
Creak-Pause (CP)	39	0.22 (0.06)	0.30 (0.06)	0.14 (0.19)	0.20 (0.13)
All Creaks (CANA)	39	**0.46 (0.04)**	**0.33 (0.03)**	0.11 (0.23)	0.20 (0.12)
Creak-Pause (CPNA)	39	**0.72 (4e-3)**	**0.37 (0.03)**	0.15 (0.17)	0.17 (0.16)
2012, WR					
All Creaks (CA)	8	**0.82 (0.02)**	−0.09 (0.52)	0.02 (0.40)	−0.20 (0.64)
Creak-Pause (CP)	8	**0.82 (0.02)**	−0.09 (0.51)	0.04 (0.40)	−0.19 (0.64)
All Creaks (CANA)	8	**0.85 (0.02)**	0.07 (0.39)	0.08 (0.42)	−0.09 (0.50)
Creak-Pause (CPNA)	8	**0.87 (0.02)**	0.07 (0.40)	0.09 (0.43)	−0.09 (0.50)

effectively uncorrelated from the acoustic counts (e.g., 0.07, $p = 0.28$ for the CPNA count).

To test whether the number of empty hooks was correlated with acoustic depredation measurements, the empty and ineffective (damaged) hook count from the survey data was correlated with the acoustic data, and the resulting correlations were close to zero with non-significant *p* values greater than 0.3.

While the relationship between the survey lip count and the CPNA in Table IV is highly significant in 2012 ($p = 1\text{e-}3$), the correlation drops in 2011 (0.49, $p = 0.03$), when all 43 stations are used. To understand this pattern, Figs. 5 and 6 plot the 2011 and 2012 survey lip count and acoustic counts as a function of haul number. Both figures show the haul number as the time unit, instead of the date and time, in order to avoid

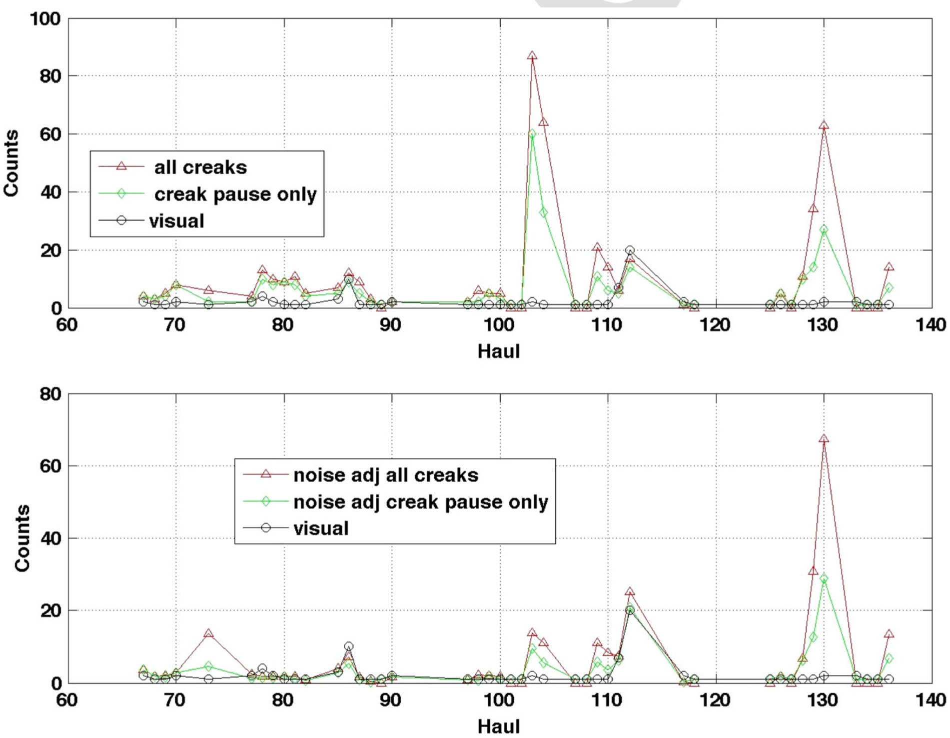

FIG. 5. (Color online) Plots of 2011 time series of "lips" depredation count from federal survey database, vs four candidature acoustic depredation measurements. (top) Raw acoustic counts of creak and creak-pause events vs lips records. (bottom) Noise-adjusted acoustic counts vs lips records. Note the substantial deviation between the acoustic and survey counts at hauls 128–130, which lowers the 2011 lips/CPNA correlation coefficient to 0.49.

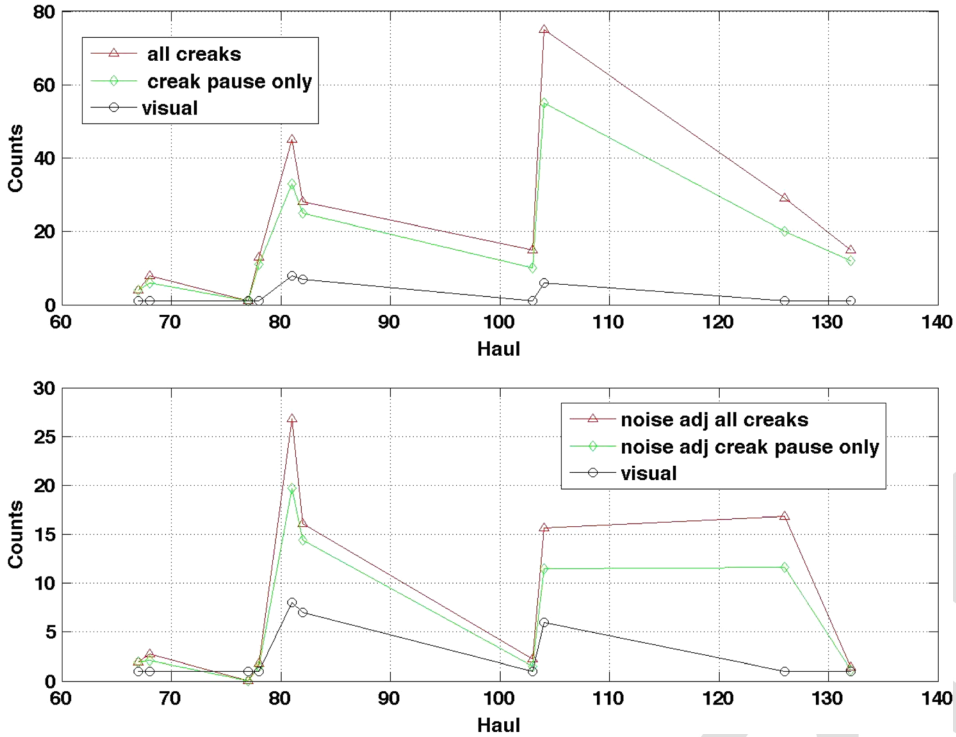

FIG. 6. (Color online) Same format as Fig. 5, but displaying the ten samples of the 2012 deployment. (top) Raw acoustic counts of creak and creak-pause events vs lips records. (bottom) Noise-adjusted acoustic counts vs lips records.

plotting the 24 h gaps between hauls. The top subplot for each figure shows the raw acoustic counts (CA and CP), and the bottom subplot displays the noise-adjusted acoustic counts (CANA and CACP). In Fig. 6 (2012 data), the correlation between the CPNA and survey lip counts is visually apparent, although the CPNA has 2.5 more counts than the survey lip count (Table III). Figure 5 seems to tell a slightly more complex story. Hauls 103 through 105 took place during relatively quiet background noise conditions on July 28 and 29 (see Fig. 3 and Sec. II D), so the impact of the noise adjustment is especially noticeable between the top and bottom subplots. The real puzzle, however, is the substantial deviation between the acoustic and survey lip counts during three particular hauls: 128, 129, and 130, which correspond to sets conducted on August 12 and 13 at Stations 84 and 85, at the far western edge of the survey, near Kodiak Island (Fig. 4).

A review of the circumstances at Stations 84 and 85 indicated no unusual conditions, other than the fact that Station 84 experienced the largest number of whales (4) present during a single haul for the entire 2011 survey (median whales per set: 1.8). Given that the 2012 surveys effectively excluded whale encounters with three or more whales present (due to the difficulties with manual analysis), a suspicion arose that the degree of correlation between the acoustic and visual depredation counts may change with the number of whales present. Thus a WR (whales restricted to less than three) analysis was performed for both years (39 hauls in 2011, 8 in 2012), with the results shown in Tables III and IV. The Lips/CPNA WR correlation increases from the unrestricted analysis, from 0.49 ($p = 0.03$) to 0.72 ($p = 4e\text{-}3$).

D. Poisson regression analysis

Of the six predictor variables tested in the Poisson regression, Table V shows that only three (the Lips metric, the

TABLE V. Results of GLM analysis using a Poisson regression [Eq. (3)]. Of the six predictor variables used (four survey metrics, number of animals sighted, and survey year), only those shown in this table were found to be significantly different from zero.[a]

	Predictor Variable		
Acoustic metric μ	*L*	*S*	*Number whales*
All Creaks (CA)	0.026 [−0.05; 0.10], *0.48*	0.085 [−0.08; 0.25], *0.30*	0.73 [0.53; 0.93], ***1e-9***
Creak-Pause (CP)	0.047 [−0.01; 0.11], *0.15*	0.11 [−0.03; 0.26], *0.12*	0.76 [0.40; 1.11], ***7.5e-5***
All Creaks (CANA)	0.063 [2e-3; 0.12], ***0.04***	0.13 [−0.01; 0.28], *0.07*	0.67 [0.45; 0.88], ***1e-7***
Creak-Pause (CPNA)	0.09 [0.05; 0.14], ***1e-4***	0.16 [0.04; 0.27], ***9e-3***	0.69 [0.36; 1.02], ***1e-4***
Predictor Variable →	*Log(L)*	*Log(S)*	*Log(Number whales)*
Creak-Pause (CPNA)	0.51 [0.20; 0.82], ***2e-3***	0.54 [0.11; 0.97], ***0.01***	1.5 [0.55; 2.5], ***3e-3***

[a]Table cell format: estimate of coefficient β [confidence intervals], *p-value of t-test for non-zero value*. *p*-values less than 0.05 are highlighted in bold italic. Predictor definitions: *L:* Lips survey metric; *S:* Sablefish survey metric; Number whales: number of individuals sighted by survey observers; See Table III for acoustic metric definitions.

ID: ganeshg Time: 08:30 I Path: Q:/3b2/JAS#/Vol00000/140166/APPFile/AI-JAS#140166

Sablefish metric, and the number of whales sighted during a haul) were found to have a significant probability of having non-zero predictor coefficients (i.e., yielding a *t*-test *p*-value under 0.05). The year of the survey, and the halibut and general depredation metrics were never significant, regardless of the particular acoustic metric tested. Also not significant were interaction terms, or whether the linear or log values of the predictors were used. In Table V, it is clear that the number of whales sighted during a haul was a significant predictor for all acoustic counts, whether noise-corrected or not. By contrast, the lips and sablefish visual metrics are only valid predictors of acoustic activity when the noise-correction factor of Eq. (1) is included; furthermore, only the noise-adjusted creak-pause metric (CPNA) yields *p*-values that remain significant after a Bonferroni correction of $0.05/12 = 5\text{e-}3$ is applied.

Figure 7 plots the performance of the best-fit GLM model (three predictors of Lips count, Sablefish count, and number of whales sighted) against the measured CPNA counts across both years. Both 2011 and 2012 are plotted on the figure, with haul numbers 1–43 indicating 2011 data, and 43–61 indicating 2012 data. "Haul number" in this figure is simply an index that is unrelated to the "Haul" axis labels in Figs. 5 and 6. For example, haul numbers 37–39 correspond to hauls 128–130 in Fig. 5.

IV. DISCUSSION

A. Statistical correlations between standard and acoustic depredation measures

Several trends emerge when the reviewing the correlations between the various acoustic and survey depredation metrics in Table IV, and the GLMs in Table V. First, both analyses indicate that halibut and grenadier damage seemed uncorrelated with sperm whale acoustic activity, as was generic depredation. Counts of empty hooks also showed no significant correlations. The lack of correlation with empty hooks was expected, because many factors besides depredation are responsible for the number of empty hooks on a set, including bottom composition, benthic predation, and currents that shake or "spin off" fish from the line during the haul to the surface.

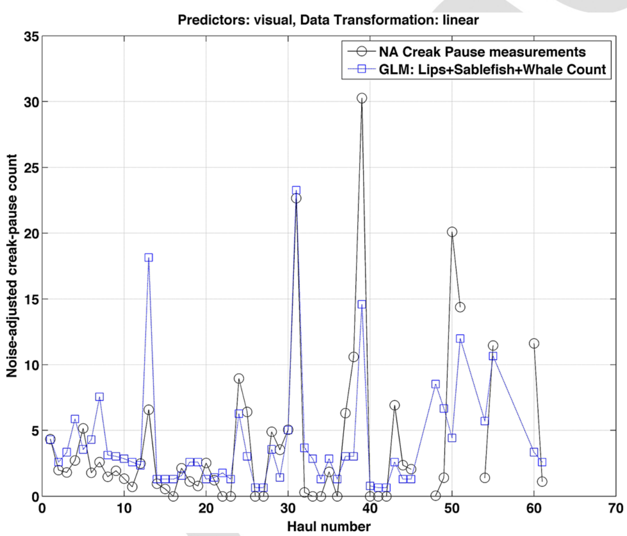

FIG. 7. (Color online) Performance of best-fit GLM (Poisson regression) in predicting CPNA acoustic metric. Circle, CPNA measurements; square, combined GLM with three predictors (shown in Table V). The "haul numbers" shown on this figure are not related to the haul indices shown in Figs. 5 and 6. The 2011 survey results are displayed in haul numbers 1–42, while the 2012 measurements are displayed in hauls 42 to 61.

By contrast, both the correlation and GLM analyses found strong links between unidentifiable fish remains stuck to the line, such as lips and jaws, and certain acoustic metrics. The GLM also found evidence that sablefish damage was related to acoustic activity, a conclusion partially supported by Table IV, which found significant correlations in the 2011 survey data but not in the 2012 data.

However, a second observation from both tables is that simple raw counts of all creaks detected are not related to survey depredation measures; instead, adjusting raw acoustic creak detection rates for variations in background noise levels is critical in obtaining associations. In almost every situation, this adjustment doubled or tripled the correlation value and reduced the *p*-value by a similar factor. For example, the 2011 correlations for (Lips/all-creaks) and (Lips/creak-pauses only) jumped from 0.09/0.16 to 0.26/0.49, respectively. A similar, but smaller, improvement was also observed in the 2012 results. The GLM only obtained significant values for the visual survey predictor coefficients when noise-adjusted metrics were used.

Finally, both analyses (particularly the GLM results in Table V) provide significant support for the idea that counting creak-pauses (CP), instead of all creaks (CA), is a better acoustic metric of depredation activity, and that noise-adjusted creak-pause counts (CPNA) are the best overall acoustic depredation metric. For example, if all hauls that took place in the presence of whales were analyzed, the number of noise-adjusted "creak-pauses," or CPNA, was found to be significantly correlated with survey counts of lips $[r(43) = 0.49 \;\; p = 0.03]$ and sablefish damage $[r(43) = 0.29,\; p = 0.05]$ in 2011, and significantly correlated with lip counts in 2012 $[r(10) = 0.89,\; p = 1\text{e-}3]$. If one assumes that a *p*-value of 0.05 is significant for a single test, and that 32 of the tests in Table IV are actually statistically independent (the WR cases are not independent datasets from the unrestricted sets), then a conservative Bonferroni-corrected *p*-value is 1.5e-3, and the CPNA/survey-lip correlations still remain significant. The CPNA metric in Table V also provides the only statistically significant result after a similar Bonferroni correction. These results provide circumstantial evidence that creak-pause events are associated with successful prey-capture attempts.

There is also evidence that the strength of the correlation between the acoustic and survey depredation counts decreases if more than two depredating whales are present during a haul. The Pearson correlation analyses in Table IV are more convincing if the analyzed hauls are restricted to circumstances when fewer than three whales are present (the WR restriction). For example, the 2011 correlation between the CPNA and survey lips count increases from $r(43) = 0.49$, $p = 0.03$, to $r(39) = 0.72$, $p = 4\text{e-}3$, when four hauls that have three or more whales are eliminated from the analysis.

The 2012 correlations are relatively unaffected by the WR restriction, a result that makes sense, since the 2012 analysis was already restricted to encounters with fewer than three whales present, given the difficulties of flagging a creak-pause event when so many other animals are acoustically active.

A review of Figs. 5 and 7 shows that implementing the WR restriction effectively removes a clustered set of hauls (128–130 at Stations 84 and 85) from the 2011 data. These rejected hauls experienced extremely high amounts of whale acoustic activity, while displaying little to no visual evidence of lips or other depredation during the hauls. A review of ship logs from those stations indicates no visual evidence of offal feeding or other unusual situations; however, up to four whales were present during these hauls, the largest number of whales encountered at one time during the 2011 survey. The removal of these data points raised the correlation between the CPNA and the survey lips count to a level comparable to what was also observed in 2012 (0.72 in 2011 vs 0.87 in 2012).

Thus, roughly speaking, when more than two animals are present the visual evidence for depredation changes relatively little, but the CPNA count grows large, destroying the linear correlation between the two depredation metrics. Figure 7 and Table V illustrate how strongly CPNA measurements are linked to the number of whales present. There are several potential non-exclusive explanations for this observation. For example, human acoustic analysts have a harder time counting creaks when more acoustically active whales are present. In addition, if enough whales were competing for the line resource, some animals may be forced to dive deeper and start feeding on "spun-off" fish or dive shallow to feed off offal thrown overboard. Either situation would not leave visual evidence but would increase the CPNA count substantially. This is an important question, because the consumption of "spin-off" fish would have as much impact on future sablefish population growth as would direct depredation from the line (both actions increase mortality in the breeding population).

B. Comparison of absolute depredation rates obtained from visual and acoustic methods

Acoustic measurements (specifically the noise-adjusted creak-pause CPNA) predict a larger total depredation count for the entire survey than what is currently shown by standard depredation measurements. The exact percentage increase depends on the year, the background noise level chosen for the noise adjustment, and whether one assumes that the CPNA is a correlated only with the Lips category, or is correlated with the Sablefish category as well.

For example, for the 2012 data year, which found strong correlations between lip counts [$r(10) = 0.89$, $p = 1\text{e-}3$] and the CPNA, but not sablefish damage [$r(10) = 0.35$, $p = 0.17$], a total of 25 lips and 23 sablefish damages were observed, and a total CPNA count of 65 obtained. If one assumes that the CPNA count should be substituted for the lips count (but not the sablefish count), then the revised depredation count (CPNA plus sablefish damage) is 1.8 times larger than the survey depredation count (lips plus sablefish damage). If one assumes that the CPNA count represents the combined lips plus sablefish count (despite the lower correlation observed with sablefish damage) then the revised depredation count is 1.4 times larger than the survey count. These numbers only cover periods when two or fewer whales are present.

The corresponding numbers for the entire 2011 survey were 68 lips, 27 sablefish damages, and 147 CPNA counts, yielding acoustic depredation rates that are 1.5 to 1.8 times higher than standard counts. However, if the three anomalous hauls at stations 84 and 85 are removed from the picture, the 2011 survey had 66 lips, 24 sablefish damages, and only 99 CPNA counts, yielding acoustic depredation rates that were 1.1 to 1.5 times higher than the survey count.

Thus, depending on the year and the assumptions relating the CPNA to sablefish damage, the acoustic depredation count is anywhere from just 10% to over 80% higher than the standard survey estimate. For haul sets where two or fewer whales are present, the CPNA places an upper bound on the bias of visual depredation counts (1.8 times the visually observed depredation rate). Unfortunately, these results heavily depend on the reference background noise level chosen for the call adjustments: An increase in the reference background noise level by 3 dB will decrease the CPNA count by a factor of 3. In this study, the reference level chosen was 93 dB re 1 uPa, integrated between 3 and 9 kHz. Previous work (Mathias *et al.*, 2013) has shown that this background level is expected to give a creak detection range of 5 km, enough to cover the entirety of a haul in 2011, and the last 3/4 of a haul in 2012.

Whenever roughly three or more whales are present, CPNA predicts a much higher depredation rate than observed rates, but at present the inability to distinguish creak-pauses from creaks during heavy whale acoustic activity limits the sample size available for analysis.

V. CONCLUSION

For most sablefish survey stations in 2011 and 2012, the noise-adjusted count of a particular echolocation sound (a "creak" followed by a few second "pause") was highly correlated with survey counts of lip remains during both years [2012: $r(10) = 0.89$, $p = 1\text{e-}3$; WR count in 2011: $r(39) = 0.72$, $p = 4\text{e-}3$], provided that less than three whales were present at a given haul (the WR case). The WR CPNA is somewhat correlated with observed sablefish damage in 2011 [$r(39) = 0.37$, $p = 0.03$], but not correlated with other species depredation or the number of empty hooks present. The noise-adjusted creak-pause (CPNA) depredation count was anywhere from 10% to 80% higher than the survey counts, depending on the survey year and assumptions employed. The observed linear correspondence between CPNA and lip remains breaks down when three or more whales are present: Under such circumstances the CPNA greatly exceeds survey counts. The application of a generalized linear model in the form of a Poisson logistic regressions support the contention that CPNA is predicted by (and thus is a good predictor of) standard Lips and Sablefish depredation counts.

Even though both survey years used different acoustic instruments, conducted different deployment strategies, and used different styles of manual analysis, the results from both years are consistent, a conclusion reflected by the fact that the GLM did not find the survey year to be a significant predictor.

This study suggests that passive acoustics can provide upper bounds on the bias of survey depredation monitoring efforts, if background noise effects are properly addressed, and if fewer than three whales are present during a set. The question of the relationship between acoustic and survey depredation counts during times when three or more whales are present remains an open question, requiring a larger sample size and additional development of automated means of new approaches for processing the acoustic data.

There are several concrete steps than could be taken in potential future work. The first is the development of a computer-assisted method for recognizing creak-pause events during circumstances when more than two whales are acoustically active. The inability of human analysts to distinguish creaks from creak-pauses during substantial whale presence substantially reduced the fraction of the 2012 dataset available for analysis. The second step is the deployment of simple vertical arrays (multi-hydrophone systems) on buoylines, instead of a single autonomous recorder, in order to add a localization capability to the system (e.g., Mathias *et al.*, 2013). There are two advantages to such a system: First, the identification of offal (shallow-surface) feeding behavior from conventional depredation behavior can be distinguished by measuring the elevation angle vs intensity of measured creaks; and second, the detection range can be doubled with a four-element system. Thus, the system could be deployed on one end of a 8 km set, while still covering the entire set, but not requiring the deployment of an extra buoyline midway down the set.

ACKNOWLEDGMENTS

This project was supported by NOAA Grant Award Number NA11NMF4370001. We would like to thank Victoria O'Connell for helping to secure funding for the project and focusing discussions about acoustic depredation measurements. We also thank the crews of the F/V Prowler and the F/V Alaskan Leader for their assistance and patience in deploying and recovering the autonomous acoustic recorders. Dan Falvey designed a quick-release clamp and bridle system that made the rapid deployment of acoustic recorders practical, and he provided useful insight into various issues involving acoustic interpretations of depredation. Natalie Sattler, Gina Horath, and Patricia Fernandez all assisted in the acoustic data analysis. The North Pacific Research Board (NPRB) (Project 918) provided the initial support to build the 2011 acoustic instruments.

Ashford, J. R., Rubilar, P. S., and Martin, A. R. (**1996**). "Interactions between cetaceans and longline fishery operations around South Georgia," Mar. Mammal Sci. **12**(3), 452–457.

Barlow, J., Kahru, M., and Mitchell, B. G. (**2008**). "Cetacean biomass, prey consumption, and primary production requirements in the California Current ecosystem," Mar. Ecol. Prog. Ser. **371**, 285–295.

Capdeville, D. (**1997**). "Interaction of marine mammals with the longline fishery around the Kerguelen Island (Division 58.5.1) during the 1995/96 cruise," Ccamlr Sci. **4**, 171–174.

DeRuiter, S. L., Bahr, A., Blanchet, M. A., Hansen, S. F., Kristensen, J. H., Madsen, P. T., Tyack, P. L., and Wahlberg, M. (**2009**). "Acoustic behaviour of echolocating porpoises during prey capture," J. Exp. Biol. **212**, 3100–3107.

Drouot, V., Gannier, A., and Goold, J. C. (**2004**). "Diving and feeding behaviour of sperm whales (*Physeter macrocephalus*) in the northwestern Mediterranean Sea," Aquat. Mammal. **30**(3), 419–426.

Goold, J. C., and Jones, S. E. (**1995**). "Time and frequency domain characteristics of sperm whale clicks," J. Acoust. Soc. Am. **98**, 1279–1291.

Gordon, J. C. D. (**1987**). "Behavior and ecology of sperm whales off Sri Lanka," PhD dissertation, University of Cambridge, Cambridge, United Kingdom, 347 pp.

Gosho, M. E. D., Rice, W., and Breiwick, J. M. (**1984**). "The sperm whale *Physeter macrocephalus*," Mar. Fish. Rev. **46**(4), 54–64.

Jaquet, N. (**1996**). "How spatial and temporal scales influence understanding of Sperm Whale distribution: A review," Mammal Rev. **26**(1), 51–65.

Jaquet, N., Dawson, S., and Douglas, L. (**2001**). "Vocal behavior of male sperm whales: Why do they click?" J. Acoust. Soc. Am. **109**, 2254–2259.

Jaquet, N., Dawson, S., and Slooten, E. (**2000**). "Seasonal distribution and diving behaviour of male sperm whales off Kaikoura: Foraging implications," Can. J. Zool. **78**, 407–419.

Johnson, M., Aguilar Soto, N., and Madsen, P. T. (**2009**). "Studying the behaviour and sensory ecology of marine mammals using acoustic recording tags: A review," Mar. Ecol. Prog. Ser. **395**, 55–73.

Madsen, P. T., Wahlberg, M., and Mohl, B. (**2002**). "Male sperm whale (*Physeter macrocephalus*) acoustics in a high latitude habitat: Implications for echolocation and communication," Behav. Ecol. Sociobiol. **53**, 31–41.

Mathias, D., Thode, A. M., Straley, J., and Andrews, R. K. (**2013**). "Depth and range tracking of sperm whales in the Gulf of Alaska using a two-element vertical array, satellite, and bioacoustic tags," J. Acoust. Soc. Am. **134**(3), 2446–2461.

Mathias, D., Thode, A. M., Straley, J., Calambokidis, J., Schorr, G. S., and Folkert, K. (**2012**). "Acoustic and diving behavior of sperm whales (*Physeter macrocephalus*) during natural and depredation foraging in the Gulf of Alaska," J. Acoust. Soc. Am. **132**, 518–532.

Mathias, D., Thode, A. M., Straley, J., and Folkert, K. (**2009**). "Relationship between sperm whale (*Physeter macrocephalus*) click structure and size derived from videocamera images of a depredating whale," J. Acoust. Soc. Am. **125**(5), 3444–3453.

Mellinger, D. K., Stafford, K. M., and Fox, C. G. (**2004**). "Seasonal occurrence of sperm whale (*Physeter macrocephalus*) sounds in the gulf of Alaska," Mar. Mammal Sci. **20**(1), 48–62.

Miller, P. J. O., Johnson, M. P., and Tyack, P. L. (**2004**). "Sperm Whale behaviour indicates the use of echolocation click buzzes 'creaks' in prey," Proc. Biol. Sci. **271**, 2239–2247.

Mullins, J., Whitehead, H., and Weilgart, L. S. (**1988**). "Behaviour and vocalizations of two single sperm whales, *Physeter macrocephalus*, off Nova Scotia," Can. J. Fish. Aquat. Sci. **45**, 1736–1743.

Nolan, C. P., and Liddle, G. M. (**2000**). "Interactions between killer whales (*Orcinus orca*) and sperm whales (*Physeter macrocephalus*) with a longline fishing vessel," Mar. Mammal Sci. **16**(3), 658–664.

Perez, M. A. (**2006**). "Analysis of marine mammal bycatch data from the trawl, longline, and pot groundfish fisheries of Alaska, 1998–2004, defined by geographic area, gear type, and target groundfish catch species," NOAA Tech. Memo. NMFS-AFSC167 (U. S. Department of Commerce, Washington, DC), 130 pp.

Ponce, D., Thode, A. M., Guerra, M., Urban, J., and Swartz, S. (**2012**). "Relationship between visual counts and call detection rates of gray whales (*Eschrichtius robustus*) in Laguna San Ignacio, Mexico," J. Acoust. Soc. Am. **131**, 2700–2713.

Purves, M. G., Agnew, D. J., Balguerias, E., Moreno, C. A., and Watkins, B. (**2004**). "Killer whale *Orcinus orca* and sperm whale *Physeter macrocephalus* interactions with longline vessels in the Patagonian toothfish fishery at South Georgia, South Atlantic," Ccamlir Sci. **11**, 111–126.

Rice, D. W. (**1989**). "Sperm Whales," in *Handbook of Marine Mammals*, edited by S. H. Ridgway and R. Harrison (Academic, London), Vol. 4, pp. 177–233.

Richardson, W. J., Greene, C. R., Malme, C. I., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic Press, San Diego, CA), Chap. 4.

Sigler, M. F., Lunsford, C. R., Straley, J. M., and Liddle, J. B. (**2008**). "Sperm whale depredation of sablefish longline gear in the northeast Pacific Ocean," Mar. Mammal Sci. **24**(1), 16–27.

Sigler, M. F., and Zenger, H. H. (**1989**). "Assessment of Gulf of Alaska sablefish and other groundfish based on the domestic longline survey, 1987," NOAA Tech. Memo. NMFS F/NWC-169 (NOAA, Washington, DC), 54 pp.

Surlykke, A., Futtrup, V., and Tougaard, J. (**2003**). "Prey-capture success revealed by echolocation signals in pipistrelle bats (*Pipistrellus pygmaeus*)," J. Exp. Biol. **206**, 93–104.

Teloni, V., Johnson, M. P., Miller, P. J. O., and Madsen, P. T. (**2008**). "Shallow food for deep divers: Dynamic foraging behavior of male sperms whales in a high latitude habitat," J. Exp. Mar. Biol. Ecol. **354**, 119–131.

Thode, A., Mathias, D., Straley, J., Calambokidis, J., Schorr, G., and Folkert, K. (**2010**). "Testing of potential alerting sound playbacks to sperm whales," 22 **06**/**01**, Project 3.7.1 (Oil and Gas Producers Association, Joint Industry Project, London, UK).

Wahlberg, W. (**2002**). "The acoustic behaviour of diving sperm whales observed with a hydrophone array," J. Exp. Mar. Biol. Ecol. **281**, 53–62.

Watkins, W. A., Daher, M. A., Fristrup, K. M., Howald, T. J., and Notarbartolo di Sciara, G. (**1993**). "Sperm whales tagged with transponders and tracked underwater by sonar," Mar. Mammal Sci. **9**(1), 55–67.

Watwood, S. L., Miller, P. J. O., Johnson, M. P., Madsen, P. T., and Tyack, P. L. (**2006**). "Deep-diving foraging behaviour of sperm whales (*Physeter macrocephalus*)," J. Anim Ecol. **75**(3), 814–825.

Wenz, G. M. (**1962**). "Acoustic ambient noise in the ocean: Spectra and sources," J. Acoust. Soc. Am. **34**, 1936–1956.

Whitehead, H. (**2003**). *Sperm Whales: Social Evolution in the Ocean* (University of Chicago Press, Chicago, IL), Chap. 3.

Whitehead, H., Brennan, S., and Grover, D. (**1992**). "Distribution and behaviour of male sperm whales on the Scotian Shelf," Can. J. Zool. **70**, 912–918.

Whitehead, H., and Weilgart, L. (**1990**). "Click rates from sperm whales," J. Acoust. Soc. Am. **87**, 1798–1806.

Worthington, L. V., and Schevill, W. E. (**1957**). "Underwater Sounds heard from Sperm Whales," Nature **180**(4580), 291.