Title

Using nonlinear time warping to estimate North Pacific right whale calling depths in the Bering Sea

Journal

The Journal of the Acoustical Society of America, 141(5)

ISSN

0001-4966

Authors

Thode, Aaron
Bonnel, Julien
Thieury, Margaux
et al.

Publication Date

2017-05-01

DOI

10.1121/1.4982200

Peer reviewed

Using nonlinear time warping to estimate North Pacific right whale calling depths in the Bering Sea

Aaron Thode, Julien Bonnel, Margaux Thieury, Aileen Fagan, Chris Verlinden, Dana Wright, Catherine Berchok, and Jessica Crance

Citation: The Journal of the Acoustical Society of America **141**, 3059 (2017); doi: 10.1121/1.4982200

View online: http://dx.doi.org/10.1121/1.4982200

View Table of Contents: http://asa.scitation.org/toc/jas/141/5

Published by the Acoustical Society of America

Using nonlinear time warping to estimate North Pacific right whale calling depths in the Bering Sea

Aaron Thode,[1,a)] Julien Bonnel,[2] Margaux Thieury,[2] Aileen Fagan,[3] Chris Verlinden,[3] Dana Wright,[4] Catherine Berchok,[4] and Jessica Crance[4]

[1]*Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093-0238, USA*
[2]*ENSTA Bretagne, UMR CNRS 6285 Lab-STICC, 2 rue Francois Verny, 29806 Brest Cedex 9, France*
[3]*United States Coast Guard Academy, New London, Connecticut 06320, USA*
[4]*Marine Mammal Laboratory, Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle, Washington 98115, USA*

(Received 2 December 2016; revised 6 April 2017; accepted 11 April 2017; published online 3 May 2017)

Calling depth distributions are estimated for two types of calls produced by critically endangered eastern North Pacific right whales (NPRWs) in the Bering Sea, using passive acoustic data collected with bottom-mounted hydrophone recorders. Nonlinear time resampling of 12 NPRW "upcalls" and 20 "gunshots" recorded in a critical NPRW habitat isolated individual normal mode arrivals from each call. The relative modal arrival times permitted range estimates between 1 and 40 km, while the relative modal amplitudes permitted call depth estimates, provided that environmental inversions were obtained from high signal-to-noise ratio calls. Gunshot sounds were generally only produced at a few meters depth, while upcall depths clustered between 10 and 25 m, consistent with previously published bioacoustic tagging results from North Atlantic right whales. A Wilcoxon rank sum test rejected the null hypothesis that the mean calling depths of the two call types were the same ($p = 2.9 \times 10^{-5}$); the null hypothesis was still rejected if the sample set was restricted to one call per acoustic encounter ($p = 0.02$). Propagation modeling demonstrates that deeper depths enhance acoustic propagation and that source depth estimates impact both NPRW upcall source level and detection range estimates. © *2017 Acoustical Society of America*.
[http://dx.doi.org/10.1121/1.4982200]

[JFL] Pages: 3059–3069

I. INTRODUCTION

The eastern North Pacific right whale (NPRW; *Eubalaena japonica*) is a critically endangered population that is believed to consist of just a few tens of individuals (Wade *et al.*, 2010). Passive acoustic monitoring has been used for over a decade to detect the regional and seasonal presence of these animals in the southeastern Bering Sea and Gulf of Alaska (Munger *et al.*, 2008; Clapham *et al.*, 2012; Wade *et al.*, 2006; Wright, 2015), and acoustic multipaths on similar recordings have been used to estimate NPRW ranges and source levels (Wiggins *et al.*, 2004; Munger *et al.*, 2011).

Like the congeneric North Atlantic Right Whale (NARW; *Eubalaena glacialis*), the NPRW produces several call types (McDonald and Moore 2002), including tonal calls, upsweeps, downsweeps, and down-upsweeps in addition to loud, broadband, impulsive calls called "gunshots" (McDonald and Moore, 2002; Crance and Berchok, 2016; Parks *et al.*, 2005, 2011, 2012). NPRW upsweep calls, or "upcalls," are the most common frequency-modulated (FM) call and have been recorded during the summer and fall months in the southeastern Bering Sea and south of Kodiak Island (Wade *et al.*, 2006; Clapham *et al.*, 2012; Rone *et al.*, 2012). However, similarities of call characteristics among mysticetes in the Bering Sea make accurate identification of NPRWs via passive acoustic monitoring challenging (Wright, 2015).

Here recently-developed nonlinear signal processing techniques (Bonnel *et al.*, 2010; Bonnel and Chapman, 2011; Bonnel *et al.*, 2011, 2013, 2014) have been applied to passive acoustic datasets in the Bering Sea, in order to extract range and depth estimates of NPRW calls recorded over 2 years. We believe these represent the first measurements of calling depths of NPRWs, although calling depth data exists from NARWs tagged in the Bay of Fundy, Canada (Parks *et al.*, 2011). Knowledge of calling depths has several potential applications. In the Bering Sea call depth information might aid in attributing ambiguous calls (e.g., FM sweeps, gunshots) to vocally similar species, provided that different species call at different depths. Call depth information also permits the source levels of animals to be determined accurately, which in turn permits the detection range of these animals to be calibrated as a function of sea state and other environmental conditions (Mathias *et al.*, 2013). These calibrations, in turn, are essential steps in converting raw acoustic detection rates into calling spatial densities that would provide legitimate stock management information. Additionally, calling depths might provide clues to aid in attributing ambiguous calls (e.g., FM sweeps) to vocally similar species.

The rest of this paper is organized as follows. Section II summarizes the Bering Sea dataset used for this study, and

[a)]Electronic mail: athode@ucsd.edu

then describes how the dataset was reviewed for calls, how multipath arrivals were extracted from these calls, and how the acoustic propagation environment was inverted from the data, a process which in turn led to the final range and depth estimates. Section III presents the range and depth distributions for upcalls and gunshots. Finally, Sec. IV compares the results with previously-published NARW calling depth measurements, provides a hypothesis for why upcalls tend to be generated at deeper depths than gunshots, and demonstrates why calling depth is an important consideration when attempting to invert source levels from recorded data.

II. METHODS

A. Description of acoustic dataset

The passive acoustic data analyzed here were collected as part of a Bureau of Ocean-Energy Management (BOEM) funded deployment of long-term passive acoustic recorders throughout the Bering Sea, including one mooring in Unimak Pass (BS4; 2010–2014), one mooring near the 50 m isobath (BS3; 2012–2013), and one along the 70 m isobath (M2; 2012–2013) in the southeastern Bering Sea, and several moorings in the northern Bering Sea (e.g., M8; 2009–2014; Fig. 1). Preliminary manual reviews had been conducted on these datasets (Wright, 2015), and the BS3 mooring (57.67 N, 164.72 W), for a time period between August 11, 2012 and September 13, 2013, was selected for detailed analysis, because it contained particularly clean right whale gunshot and upcall samples.

The BS3 mooring consisted of an anchor, chain, acoustic release, passive acoustic recorder (AURAL,[1] Quebec, Canada), and a 76 cm diameter steel subsurface float arranged in a linear configuration, with the recording hydrophone suspended at 44 m over the 51 m deep continental shelf. Acoustic data were sampled at 16 384 Hz, with a nominal recording bandwidth between 10 Hz and 8 kHz and a duty cycle of 85 min of recording every 5 h (yielding 6.8 h of data per day).

Spectrograms were pre-generated from the archived 10-minute WAV files and stored as PNG image files 225 s in duration. These image files (indexed to the WAV files to allow playback and zoom capability) were manually reviewed by the Alaska Fisheries Science Center's Marine Mammal Laboratory (MML, previously NMML) for the presence of various marine mammal species. Criteria for assigning sounds to a given species incorporated the seasonality and location of the detection; furthermore, FM upsweeps were identified with NPRWs whenever distinctive gunshot sounds (Crance and Berchok, 2016; Parks *et al.*, 2005, 2011, 2012) were detected contemporaneously. For every day of a given deployment, the number of images that contained sounds from a given species was tabulated, and the fraction of files containing that species' sounds for that day (files with species calls/total files per day) was recorded on a spreadsheet.

As individual calls had not previously been annotated, caution had to be taken to select files that had a high confidence that only a single species was represented in the data. This requirement quickly limited the number of days on BS3 available for analysis for the relatively rare NPRW calls: July 14 and 18, 2013 were noted as containing right whale upcalls in 24% and 28% of the files, respectively, while August 22 (37%), October 7 (88%), 19 (85%), and 24 (95%) in 2012; and September 11 in 2013 (82%) were selected as days with significant fractions of right whale gunshots. A second wave of manual analysis has been conducted, reviewing and extracting specific calls from files that had been flagged by the first-pass MML analysis as being dominated by a single species. Because the resulting 17 right whale upcalls can be confused with humpback whale sounds, this

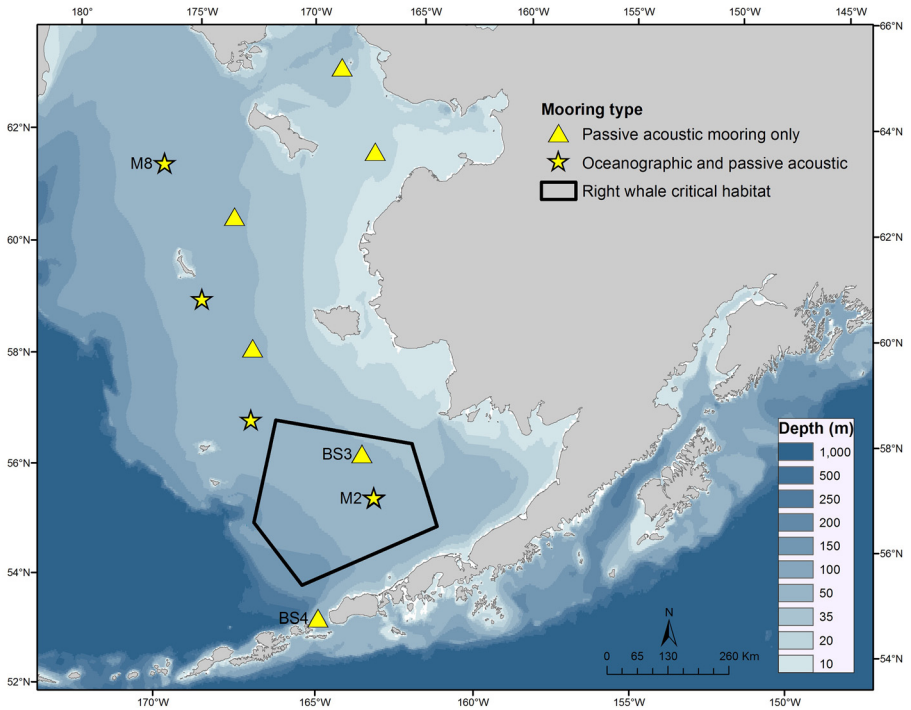

FIG. 1. (Color online) Location of "BS3," "M2," and "M8" passive acoustic recordings, along with other unlabeled BOEM deployments.

final classification was independently double-checked by MML, and five potentially ambiguous calls were removed, leaving an upcall sample size of 12.

B. Mode arrival extraction

In shallow water environments, where water depths are typically less than ten acoustic wavelengths deep, a sound propagating from a low-frequency source can be modeled as a set of "normal modes," which roughly correspond to multipath routes a sound can take to the receiver. Specifically, the pressure field p of a sound propagating through a medium of density ρ at frequency ω at time t from source depth z_s to receiver depth z_r over range r can be modeled as a sum of normal modes traveling down a waveguide (Brekhovskikh and Lysanov, 1991; Jensen *et al.*, 2011)

$$p(z_i,z_s,r,\omega)=\frac{ie^{-i\pi/4}S(\omega)}{\rho\sqrt{8\pi r}}\sum_j U_j(z_r)U_j(z_s)\frac{e^{i(k_j r-\omega t)}}{\sqrt{k_j}}, \quad (1)$$

where $S(\omega)$ is the source spectrum, while $U_j(z)$ and k_j are the mode shape and horizontal wavenumber for mode j, both of which are functions of frequency. Note how the magnitude of each modal arrival is primarily influenced by the source depth z_s, while the phase of the arrival is primarily influenced by the source range r. Mode arrival magnitudes are also influenced by the imaginary component of k_j, which arises from sound absorption and scattering losses from the ocean floor and bulk water column. Thus in principle, by measuring the relative amplitudes of several mode arrivals, one can deduce the source depth of the call, provided that range-dependent attenuation effects are accounted for. The challenge is isolating each mode from the signal, particularly at short ranges where the mode arrivals can overlap in time.

Traditionally, multiple hydrophones deployed in linear horizontal or vertical arrays have been required to spatially filter individual modes (Thode *et al.*, 2000a; Newhall *et al.*, 2009; Abadi *et al.*, 2014), but recent studies by Bonnel *et al.* have shown that if an impulsive broadband acoustic time series, such as a gunshot, is unevenly sampled in time in a particular manner, a spectrogram of the resulting "warped" signal can isolate various modes, even when a spectrogram of the original "unwarped" signal yields no obvious evidence of mode arrivals (Bonnel *et al.*, 2008, 2010). Specifically, a set of nonuniform sampling times t in the real time domain are mapped into a new evenly-sampled set of times t' in the warped domain via the formula

$$t'=\sqrt{t^2-(r/c)^2}, \quad t>r/c \quad (2)$$

with r being an estimate of source range and c being a representative estimate of the water sound speed. Note that Eq. (2) is valid only for time $t>r/c$, so that knowledge of this time origin is critical, a point that will be discussed further below. Bonnel *et al.* (2010) demonstrated that the resulting time dependent phase of the warped mode arrival becomes $\phi_j=\omega_j t'$, and thus the time series associated with the broadband mode j becomes a pure tone. The normal mode sum in Eq. (1) can thereby be transformed into a sum of tones that can be easily separated using standard filtering approaches.

Although derived for an ideal waveguide, Eq. (2) has been found to be applicable to many realistic shallow-water waveguide environments, provided that the normal mode boundary conditions of the ocean surface and bottom are dominated by reflection (Bonnel *et al.*, 2010; Bonnel and Chapman, 2011; Bonnel *et al.*, 2012), a stipulation that covers many shallow water waveguides. Furthermore, only a rough estimate of range r is required to apply the transformation under practical circumstances. The selection of an initial origin time where t is assumed equal to r/c has a much more substantial effect of the warping quality and presently must be determined by trial and error.

If the original signal is not impulsive, but is an FM sweep, then a crude deconvolution step can be used to convert the signal into a quasi-impulsive signal before applying the warping transformation. This combination of deconvolution and warping methodology has previously been used to filter modes and estimate ranges of bowhead whale FM sweeps in the Beaufort Sea (Bonnel *et al.*, 2014).

In the present work, gunshot sounds, which are true impulses, were directly converted into the warped domain using custom MATLAB software, after low-pass filtering and decimating the raw acoustic data down to an effective sampling rate of 1 kHz. A sequence of time origins was applied to determine the transformation that best separated the mode arrivals in the warped domain.

Right whale upcalls were subjected to an initial deconvolution step (Bonnel *et al.*, 2014), in order to convert the source time series into an impulsive-like signal. An example of a warping transformation applied to a right whale gunshot call is shown in Fig. 2. The highly-dispersed signal structure of a right whale gunshot call, with at least nine modal arrivals present, can be seen in Fig. 2(a); modes 1 to 4 are the four FM sweeps visible below 350 Hz. After this same signal is subjected to a nonlinear warping, with an optimized selection for the time origin, the first four modes are converted into the four horizontal bands (tones) below 60 Hz [Fig. 2(b)]. One sees that they are nicely separated, and thus can be filtered and recovered independently.

C. Environmental inversion

An accurate acoustic propagation model for the BS3 site was required to estimate the range and depths of calls. This environmental information included the water depth; receiver depth; waterborne sound speed profile; ocean floor sediment compressional speed, density, and attenuation; and potential sediment layering.

Relatively little prior environmental information was available for BS3, either in terms of bottom composition or sound speed profiles. Sediment databases only yielded core samples from tens of kilometers away, but the area was known to be "sandy." Historical sound speed profiles were derived by applying the Del Grosso sound speed equation to temperature profiles extracted from the World Ocean Atlas Database (Locarnini *et al.*, 2009). Figure 3 plots representative profiles at BS3 for the four seasons.

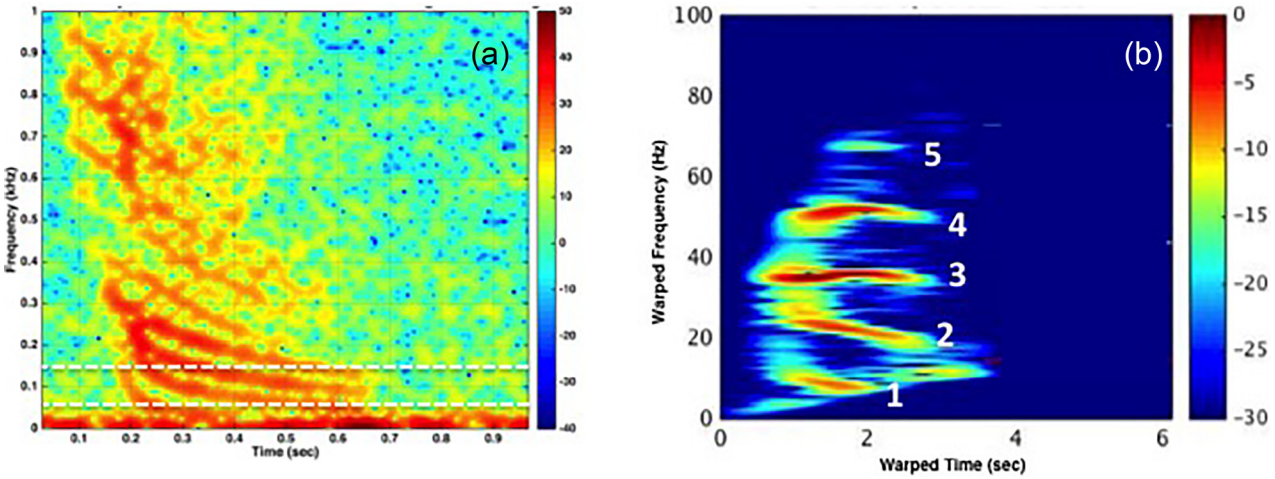

FIG. 2. (Color online) (a) Spectrogram of an impulsive NPRW gunshot call recorded at site BS3 on September 11, 2013 at 06:03:19 UMT, with dispersed modal arrivals clearly transformed into FM downsweeps. (b) Result of applying transformation in Eq. (2) to the signal, making an initial guess of $r = 9$ km and $c = 1490$ m/s.

A key feature apparent from Fig. 3 is that the local sound speed profile is heavily downward refracting during the summer and early fall, which can have a major impact on mode shapes and thus call depth estimation for frequencies above 100 Hz. Thus any environmental inversion required that the sound speed profile be taken into account.

Because of the paucity of concurrently-measured information it was necessary to use the whales' calls themselves to estimate the propagation environment, particularly the multilayer ocean floor and the waterborne sound speed profile (Collins and Kuperman, 1991; Thode *et al.*, 2000a). The bathymetry surrounding the sensor was assumed flat, a reasonable approximation for the BS3 site (Fig. 1). Separate inversions were performed for each day, since the sound speed profile could have varied from day-to-day. A given propagation model was then applied to other calls occurring the same day.

To invert for the propagation environment the following steps were taken for each whale call:

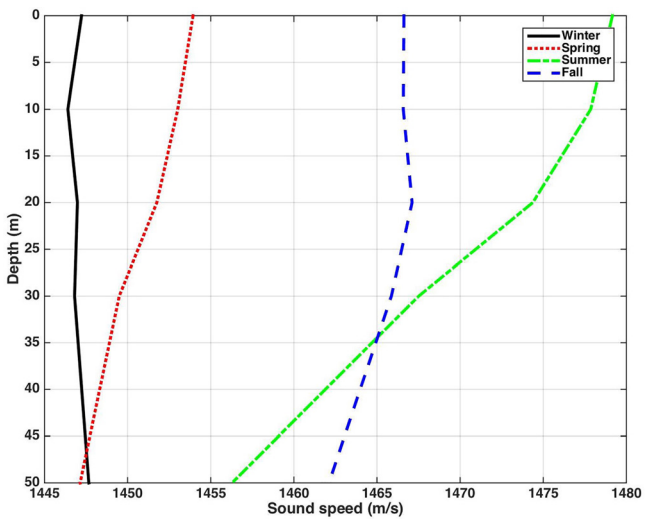

FIG. 3. (Color online) Estimated sound speed profiles at BS3 vs season.

(1) The various modal arrivals were separated as described in Sec. II B, after they had been low-pass filtered and decimated to a 1 kHz sampling rate;
(2) A Fast Fourier Transform (FFT) was applied to each modal time series, using an identical time window for all arrivals, and zero-padding the signal as needed. The mean value of the complete signal (all arrivals) was subtracted from each modal arrival before conducting the FFT, to ensure that the signal bias component was near zero. (Neglecting this step can cause spurious spectral "leakage" of the bias component at low frequencies.) For a 1-s duration signal the FFT length was typically 1024 samples.
(3) Signal bandwidths were identified where at least two modes were present. For example, between 50 and 100 Hz modes 1 and 2 were generally present, while at frequencies between 100 and 150 Hz modes 2 and 3 were often visible (while mode 1 became more difficult to extract). For each of the L frequency components lying within this bandwidth, a "data" vector $\boldsymbol{d}$ was extracted, with each element of $\mathbf{d}$ representing the complex FFT coefficient of a different modal arrival at that frequency. The relative amplitudes of these coefficients capture the relative amplitudes of the modal arrivals [as seen by the $U(z_s)$ term in Eq. (1)], while the relative phases capture the relative arrival times of each mode for that frequency, as shown by the $e^{i(k_j r - \omega t)}$ term in Eq. (1).
(4) A trial acoustic environment was selected using the parameterized model shown in Fig. 4. Five environmental parameters could be adjusted: waterborne sound speed at the ocean surface; the difference in water speed between the ocean surface and bottom; sediment attenuation; the interface (sediment) compressional speed; and the difference in sediment speed between the top and bottom of the layer. Initial geoacoustic models also adjusted the sediment layer thickness between 2 and 50 m, but the inversion was found to be insensitive to this parameter, and the final bulk inversion runs used a

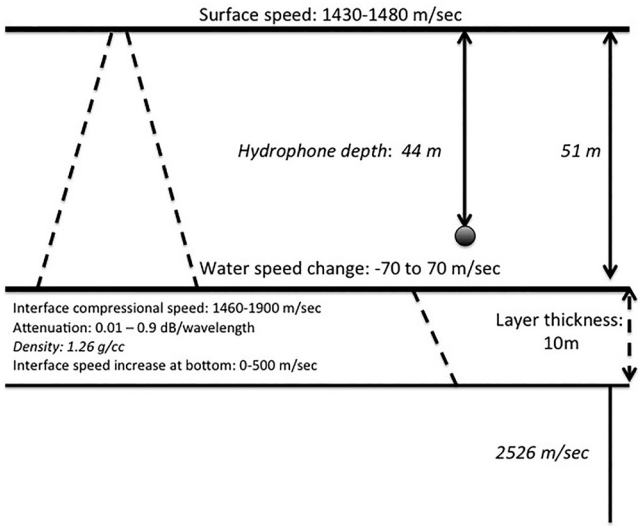

FIG. 4. Acoustic propagation parameters obtained by geoacoustic inversion, with the permitted range of the values shown. Values in italics remained fixed during global optimization runs.

fixed layer thickness of 10 m. The permissible ranges of the parameters are shown in Fig. 4. Two parameters involved the *difference* between sound speed at the top and bottom of a layer; the reason for choosing this representation is that it effectively decouples the parameters describing a linear sound speed profile (Gerstoft, 1994). The density of the sediment was taken from representative sediment cores from the region, and the water depth was held constant, as tidal variations in the region are less than a meter. From a particular instance of this trial model a series of normal modes $U_j(z)$ and associated horizontal wavenumbers $k_j(z)$ were computed using the normal mode computer program KRAKEN (Porter, 1991). A "model" vector $\mathbf{m}$ was then computed, where the jth element of $\mathbf{m}$ is

$$m_j(r,z,\omega_l) = U_j(z_r,\omega_l)U_j(z,\omega_l)\frac{e^{ik_j(\omega_l)r}}{\sqrt{k_j(\omega_l)}}. \quad (3)$$

Equation (3) explicitly shows that the vector $\mathbf{m}$ is a function of modeled source position r and z, as well as frequency component ω_l.

(5) A technique called "matched mode processing" (MMP) (Hinich and Sullivan, 1989; Yang, 1990; Baggeroer *et al.*, 1993; Lu *et al.*, 1993) was then applied to produce an "incoherent Bartlett processor" B that indicates the quality of fit between the measured data and modeled coefficients

$$B(r,z) = \sum_l B(r,z,\omega_l)/L \equiv \sum_l B_l/L = \frac{1}{L}\sum_l \frac{|\mathbf{m}(r,z,\omega_l)\cdot\mathbf{d}|}{|\mathbf{m}(r,z,\omega_l)||\mathbf{d}|}. \quad (4)$$

Here ω_l is the lth frequency of L frequency components used in the processing. The fact that the summation in Eq. (4) is incoherent and normalized ensures that the details of a whale call's source spectrum and time series structure have no impact on B. Equation (4), in effect, uses the relative arrival times and amplitudes between modes to estimate the environment concurrently with source location.

(6) Equation (4) is evaluated over a grid of candidate source ranges ($\Delta r = 20$ m) and depths ($\Delta z = 1$ m), with the source location that maximizes B selected as the best estimate for source location in that environment. An "objective function"

$$O = \max_{r,z}(B), \quad (5)$$

is the final product of a particular environmental model.

(7) A global optimization algorithm was chosen that maximizes O by selecting new environmental parameters and then repeating steps (4) through (6). In this work a MATLAB genetic algorithm (Goldberg, 1989) was used with a population of 20 individuals, a crossover fraction of 0.8, and a mutation rate of 0.01. The algorithm was run until there were no further improvements in O. Note that the resulting optimized environmental model does not precisely represent the true environment; rather, it provides a model that generates an acoustic result similar to the measured data.

(8) An important stage of the analysis then examines the quality of the data to identify if certain combinations of frequency and mode arrival give invalid results. This step takes advantage of the broadband nature of most whale calls in order to plot the optimized B_l in Eq. (4) as a function of range and frequency, using an arbitrarily-selected depth z, and selecting only two modes at a time. Figure 5 shows an example of the Bartlett function B computed using various mode pairs selected from Fig. 2(b).

According to waveguide invariant theory (Thode *et al.*, 2000b), valid modal components will generate striations in $B(r,\omega)$ that converge toward a fixed range at the limit of zero frequency [e.g., Figs. 5(a) and 5(b)]. If the nonlinear warping extracts mode estimates that are either incorrect or noise-contaminated at certain frequencies, then the striation pattern will converge toward zero or infinite range, or even vanish completely if one of the extracted modes used in a pair is simply noise. For example, modes 2 and 3 in Fig. 5(b) generate a distinctive striation pattern that converges around 7 to 8 km [similar to the mode 1–2 pattern in Fig. 5(a)], but when modes 3 and 4 are used for the calculation, the striation patterns nearly vanish, and what remains converges toward the origin [Fig. 5(c)]. Thus the extracted mode 4 for this particular call was either misidentified or heavily noise-contaminated. Mode 3 cannot be the source of the error in Fig. 5(c), as the combination of modes 2 and 3 in Fig. 5(b) yielded a good-quality result.

Once sets of valid modes have been identified, steps 2 through 7 of the inversion process can be repeated over the valid combinations of bandwidths and modes, and the final optimized value of B in Eq. (4) can be used to select the final range and depth of a call. Figure 6 shows the final optimized

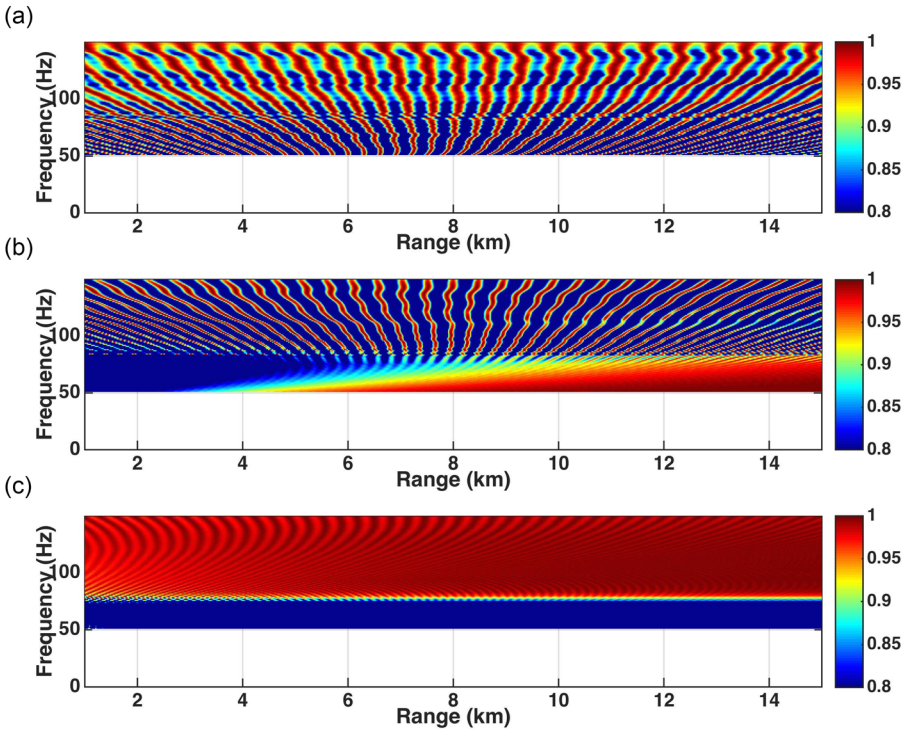

FIG. 5. (Color online) Example of how the quality of extracted modes is evaluated. Values of B_l from Eq. (4) that are created from valid pairs of modes will form striations that converge to a common range, when plotted as a range-frequency cross section (top two subplots). Values of B_l that contain invalid data form striations that converge toward zero or infinite range (bottom subplot). (a) B_l generated from modes 1 and 2, plotted using $z = 1$ m; (b) B_l generated from modes 2 and 3 only, using $z = 34$ m; (c) B_l generated from modes 3 and 4, using $z = 41$ m. The depths chosen for (a)–(c) are those that yielded the maximum value of B for that particular two-mode combination. The failure of the bottom plot to produce striations that converge between 7 and 8 km at 0 Hz indicates that mode 4 was either unsuccessfully extracted or noise-contaminated.

$B(r,z)$ for the call in Fig. 2(a), incoherently averaged over 106 frequencies using up to three modes between 50 and 150 Hz.

This "gunshot" call range was nearly 8 km, and was generated within a few meters of the ocean surface. The entire process is then repeated for another call.

III. RESULTS

Tables I and II present the estimated source ranges and depths of 20 right whale gunshots and 12 right whale upcalls, along with various information associated with the inputs and quality of each inversion. As discussed in Sec. II A, the rarity of right whales meant that multiple samples had to be estimated from the same day. Right whale upcalls were a particular problem, because it was difficult to select days where these calls were present and humpback calls were not. Since humpback whales do not produce gunshot sounds, cross-species misidentification was not an issue in Table I. As a result gunshot sounds were analyzed from five days (August 22, 2012, October 7, 2012, October 19, 2012, October 24, 2012, and September 11, 2013), but upcalls

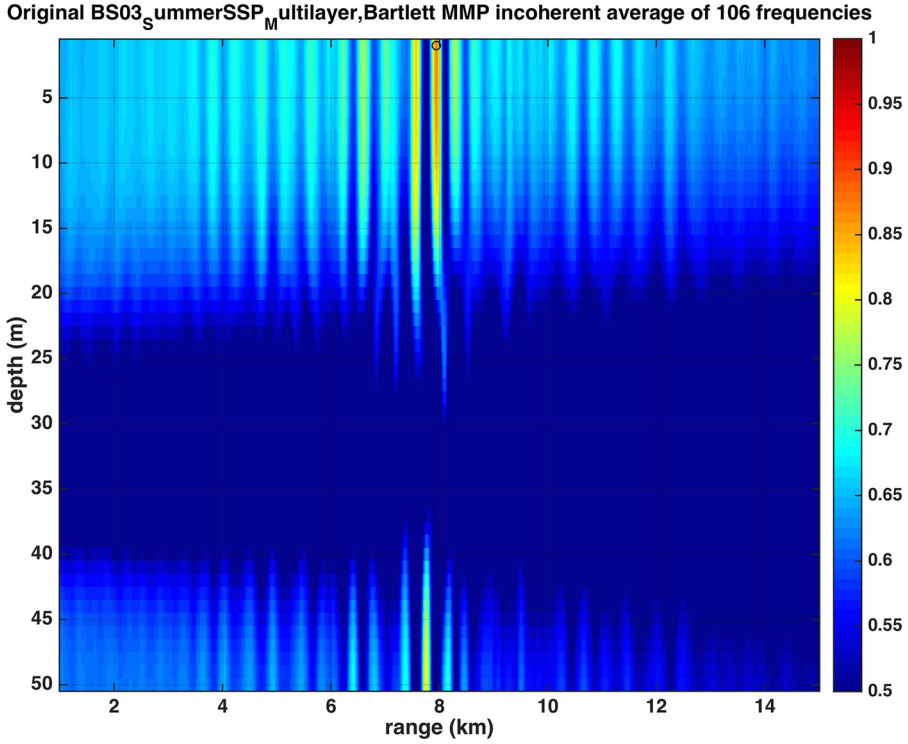

FIG. 6. (Color online) Final $B(r,z)$ used to evaluate the range and depth of Fig. 2 call, using only the first three modes that passed the quality test in Fig. 5, evaluated between 50 and 150 Hz.

TABLE I. Ranges and depths of 20 selected right whale gunshots. The order of the parameters in the "Environment" column are as follows: bottom interface sound speed (m/sec); increase in interface sound speed 10 m below ocean floor; sound speed at ocean surface; change in sound speed at ocean bottom; attenuation in dB per wavelength. The italicized row represents the signal in Fig. 2(a).

Date	Time (UTC)	# Modes	Frequency range (Hz)	Modes used for localization	Range (km)	Depth (m)	*Optimal O*	Environment
August 22, 2012	1:44:05	4	[90 160]	[1 2 3 4]	9.3	1	0.75	[1783 123 1467 -29 0.8]
August 22, 2012	6:39:46	4	[90 190]	[1 2 3 4]	8.9	1	0.69	[1838 373.1 1467 -21.11 0.6714]
October 7, 2012	7:37:35	4	[90 180]	[1 2 3 4]	5.7	7	0.87	[1780 119.8 1471 -26.31 0.46]
October 7, 2012	7:38:13	5	[100 210]	[2 3 4]	4.7	6	0.69	[1780 119.8 1471 -26.31 0.46]
October 7, 2012	7:38:20	5	[60 250]	[2 3 4]	4.6	11	0.76	[1780 119.8 1471 -26.31 0.46]
October 7, 2012	7:38:23	5	[100 270]	[2 3 4 5]	4.3	1	0.73	[1780 119.8 1471 -26.31 0.46]
October 19, 2012	4:44:49	3	[120 250]	[1 2 3]	22.7	1	0.89	[1590 199.8 1463 3.23 0.07704]
October 19, 2012	4:44:53	3	[100 300]	[1 2 3]	22.7	1	0.86	[1590 199.8 1463 3.23 0.07704]
October 19, 2012	4:45:09	3	[200 400]	[1 2 3]	22.2	1	0.85	[1590 199.8 1463 3.23 0.07704]
October 19, 2012	4:46:27	3	[200 400]	[1 2 3]	23.2	15	0.91	[1590 199.8 1463 3.23 0.07704]
October 24, 2012	4:44:02	4	[100 300]	[1 2 3 4]	23.7	1	0.81	[1590 199.8 1463 3.23 0.07704]
October 24, 2012	4:44:44	4	[100 260]	[1 2 3 4]	21.6	16	0.77	[1590 199.8 1463 3.23 0.07704]
October 24, 2012	4:45:36	3	[90 250]	[1 2 3]	14.6	1	0.86	[1590 199.8 1463 3.23 0.07704]
October 24, 2012	4:46:00	4	[100 250]	[1 2 3 4]	14.4	1	0.79	[1590 199.8 1463 3.23 0.07704]
October 24, 2012	19:40:02	5	[120 340]	[2 3 4 5]	16.6	1	0.91	[1590 199.8 1463 3.23 0.07704]
September 11, 2013	6:01:13	3	[50 140]	[1 2 3]	7.8	15	0.88	[1892 479.8 1448 35.56 0.2434]
September 11, 2013	6:01:26	4	[60 140]	[1 2 3]	8.2	14	0.87	[1892 479.8 1448 35.56 0.2434]
September 11, 2013	*6:03:18*	*5*	*[100 150]*	*[1 2 3 4 5]*	*8.2*	*1*	*0.83*	*[1780 119.8 1471 -26.31 0.46]*
September 11, 2013	6:03:27	5	[100 160]	[1 2 3 4]	8.3	6	0.87	[1780 119.8 1471 -26.31 0.46]
September 11, 2013	6:02:35	4	[100 150]	[1 2 3 4]	8.7	1	0.82	[1780 119.8 1471 -26.31 0.46]

were only analyzed from three closely-spaced days: July 5, 2013, July 14, 2013, and July 18, 2013. A glance at the call ranges shows that calls analyzed over the same day have similar ranges, suggesting that calls analyzed from the same day are likely generated from the same animal (or the same group). The analysis suggests that gunshot sounds were detected out to 22 km range, while upcalls were detected out to 40 km range.

Figure 7 displays cross-sections of $B(r,z)$ for both a gunshot (the same surface shown in Fig. 6) and an upcall (12:01:31 on July 14, 2013; Table II). Figure 7(a) shows that the relatively few modes used for the inversion limit the precision of the depth estimate to about 5 m for calls less than 10 m deep; the depth resolution for three-mode ambiguity surfaces is poorer for calls less than 8 m depth.

Figures 7(b) and 7(c) provide some insight into how $B(z)$ varies with environmental perturbations, including increasing the bottom interface speed by 100 m/s, halving the sediment speed gradient, and using an isovelocity sound profile instead of a summertime downward refracting profile. In general the perturbations do not change $B(z)$ much, in part because the depth resolution of the surface is already fairly low, and in part because modal shapes are relatively unaffected by bottom environmental perturbations. However, above 100 Hz the waterborne sound speed gradient can have a substantial impact on mode shape and thus the depth estimate; the most common impact from a mismatched sound speed profile is that the ambiguity surface mainlobe tends to jump to the "conjugate depth," or the same distance from the ocean floor as the original call's depth below the surface.

TABLE II. Ranges and depths of 12 selected right whale upcalls. The order of parameters in the Environment column is the same as in Table I. The bold typeface shows the upcall used in Fig. 7, and the environment shown in italics was used to produce Fig. 9.

Date	Time (UTC)	# Modes	Frequency range (Hz)	Modes used for localization	Range (km)	Depth (m)	*Optimal B*	Environment
July 5, 2013	8:30:33	3	[120 200]	[1 2 3]	30.5	19	0.92	[1812 349.5 1441 -8.821 0.2085]
July 14, 2013	7:55:57	3	[80 150]	[1 2 3]	15.2	18	0.89	[1872 299.1 1477 -34.81 0.2199]
July 14, 2013	7:56:01	3	[80 150]	[1 2 3]	15.1	19	0.94	[1872 299.1 1477 -34.81 0.2199]
July 14, 2013	8:17:13	3	[100 160]	[1 2 3]	17.2	17	0.94	[1684 842 1480 -25 0.8]
July 14, 2013	8:17:30	3	[100 160]	[1 2 3]	17.2	16	0.79	*[1684 842 1480 -25 0.8]*
July 14, 2013	12:00:46	4	[80 150]	[1 2 3]	5.6	14	0.84	[1872 299.1 1477 -34.81 0.2199]
July 14, 2013	**12:01:31**	**4**	**[80 140]**	**[1 2 3]**	**6.7**	**14**	**0.84**	**[1700 340.2 1476 -39.8 0.02822]**
July 14, 2013	13:14:52	4	[90 150]	[1 2 3 4]	8.5	13	0.83	[1872 299.1 1477 -34.81 0.2199]
July 18, 2013	1:01:54	3	[90 190]	[1 2 3]	14.6	30	0.77	[1872 299.1 1477 -34.81 0.2199]
July 18, 2013	1:32:44	3	[70 150]	[1 2 3]	12.8	21	0.82	[1872 299.1 1477 -34.81 0.2199]
July 18, 2013	1:47:05	3	[80 160]	[1 2 3]	13.1	20	0.89	[1872 299.1 1477 -34.81 0.2199]
July 18, 2013	1:49:26	3	[90 160]	[1 2 3]	12.4	14	0.83	[1714 361.6 1463 -24.54 0.3284]

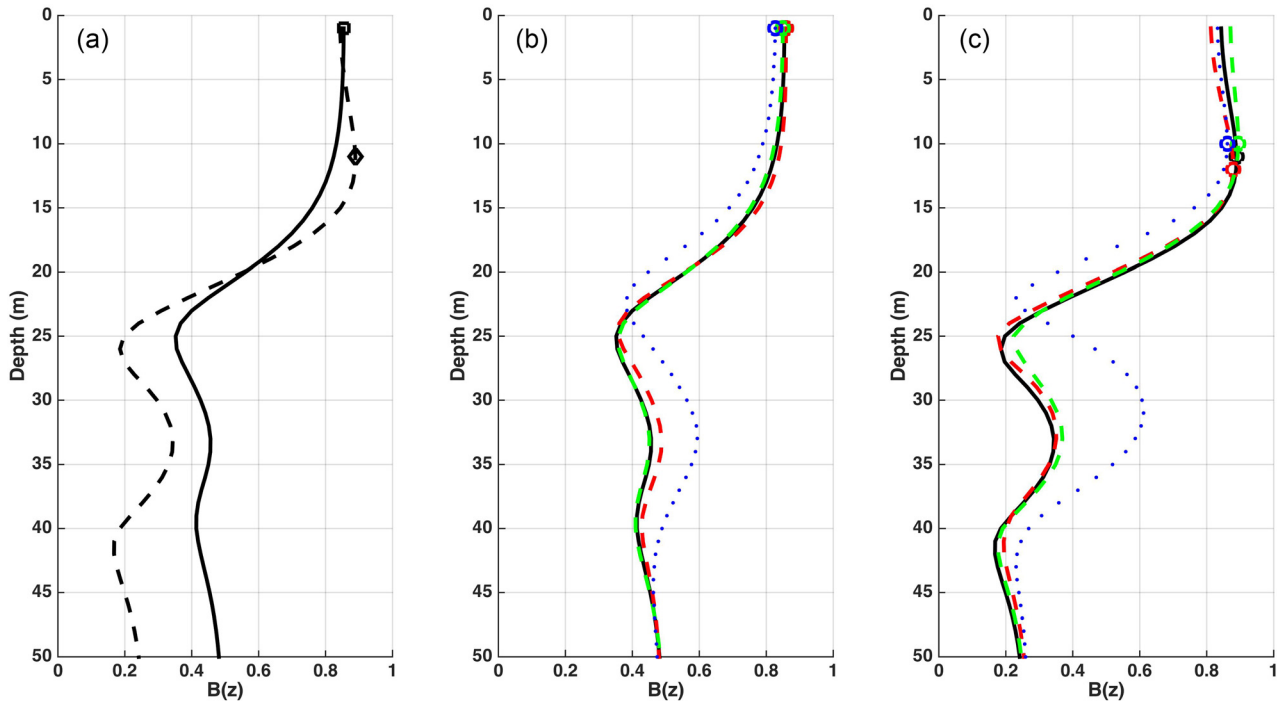

FIG. 7. (Color online) Depth resolution and environmental uncertainty of three-mode ambiguity surfaces $B(z)$. Each curve represents a vertical "slice" of the surface at the source range that yields the maximum correlation. Circles indicate depths where $B(z)$ is maximum. Same gunshot (from Fig. 6) and upcall (highlighted in bold in Table II) used in all panels. (a) Solid line: ambiguity surface from Fig. 6, for gunshot; dashed line: ambiguity surface for upcall. (b) Solid line: Ambiguity surface for gunshot, optimized environment; red dashed line: ambiguity surface with bottom interface speed increased by 100 m/s; green dashed line: ambiguity surface with bottom gradient reduced by 50%; dotted line: ambiguity surface for isovelocity waterborne sound speed profile. (c) Same as (b), but showing environmental perturbations for upcall ambiguity surface.

Figure 6 also shows this tendency, as the sidelobes of the ambiguity surface are large at the conjugate depth just above the ocean floor. This effect occurs because the relative amplitudes of modes 1 and 2 tend to be very similar at both the conjugate depth and true depth; indeed, a two-mode MMP ambiguity surface would have the same values at the two depths and would not be able to distinguish between them. The addition of a third mode in Figs. 6 and 7 provides the slight asymmetry needed to resolve the conjugate depth ambiguity.

Figure 8 summarizes depth distributions of the calls recorded in Tables I and II. A bin width of 5 m was chosen to reflect the estimated depth resolution of the matched-mode processing analysis visible in Fig. 7. There is a clear difference between the depth distributions of the right whale gunshots and upcalls. A Wilcoxon rank sum test rejects the null hypothesis that these two call distributions share the same mean ($p = 2.9 \times 10^{-5}$), using a total sample size of 32. If one limits sample sizes to eliminate the possibility of multiple calls from the same animal or group (by choosing the value that yielded the highest value of O for calls that share similar ranges over a 2-h period), six gunshot samples and three upcall samples remain. The resulting hypothesis test still rejects the null hypothesis that the call depth distributions are the same (p-value 0.02).

IV. DISCUSSION

The results in Fig. 8 suggest that nonlinear analysis of archival passive acoustic data can extract call depth estimates with sufficient resolution to distinguish depth differences between call types. The very shallow depths obtained for NPRW gunshots (median depth 0 m, interquartile range from 0 to 5 m, maximum depth 16 m) are similar to the depth distribution obtained from 189 gunshot calls recorded on tagged NARW by Parks *et al.* (2011) (median depth of 0 m, interquartile range from 0 to 0.5 m, and maximum depth of 11 m).

The depth distribution for NPRW upcalls (median depth 12 m, interquartile range from 7 to 15 m, max depth 30 m) is slightly deeper than the distribution of 264 measured NARW upcall depths (median depth 2 m, interquartile range from 0.5 to 8 m, max depth 109 m). Based on the interquartile ranges reported, both the NPRW and NARW upcall distributions are deeper than their corresponding gunshot depth distributions, but both populations remain much shallower than the water depths the data were collected in.

A natural question to ask is why the depth distributions for the two call types might be different. One could simply argue that the call types are associated with particular behavioral states, which in turn are associated with different dive depths. However, there might also be good acoustical reasons behind the different depth distributions: the lower-frequency upcalls need to be produced at deeper depths to propagate efficiently. Figure 9 illustrates this point by plotting the efficiency of sound transmission during the summer at BS3 as a function of source depth and frequency, using the "summer" sound speed profile shown in Fig. 3 and the italicized environment in Table II. The image is generated by evaluating Eq. (1) as a function of range *using* 10 m

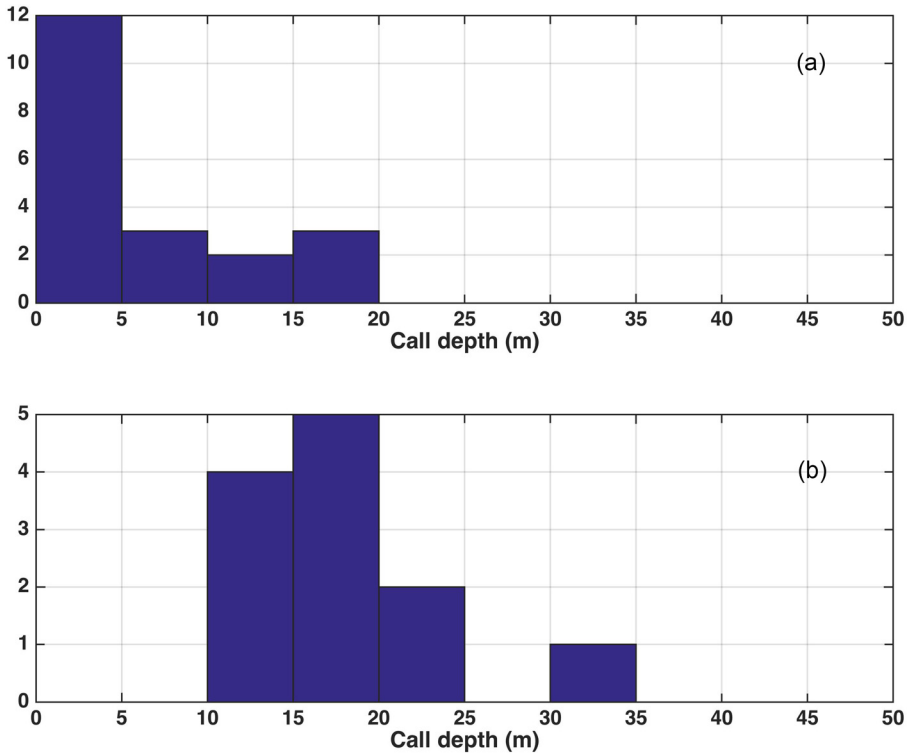

FIG. 8. (Color online) Distribution of call depths broken down by call type. (a) Right whale gunshots. (b) Right whale upcalls.

increments from 100 m to 10 km, for a particular combination of source depth and frequency. The resulting transmission loss curve is then fitted using least-squares to a power-law propagation model of the form $A\log_{10} r$, where a value of $A = 20$ would correspond to spherical spreading and $A = 10$ would correspond with cylindrical spreading. The best-fit value of A is then plotted in Fig. 9 at the particular source depth/frequency coordinate. Larger values of A correspond to less effective propagation (higher transmission loss with range). Thus Fig. 9 contours the transmission loss coefficient A, and not transmission loss.

Three receiver depths are modeled: 1 and 10 m, which correspond to hypothetical listening depths for NPRW, and 44 m, the depth of the AURAL recorder used to collect the

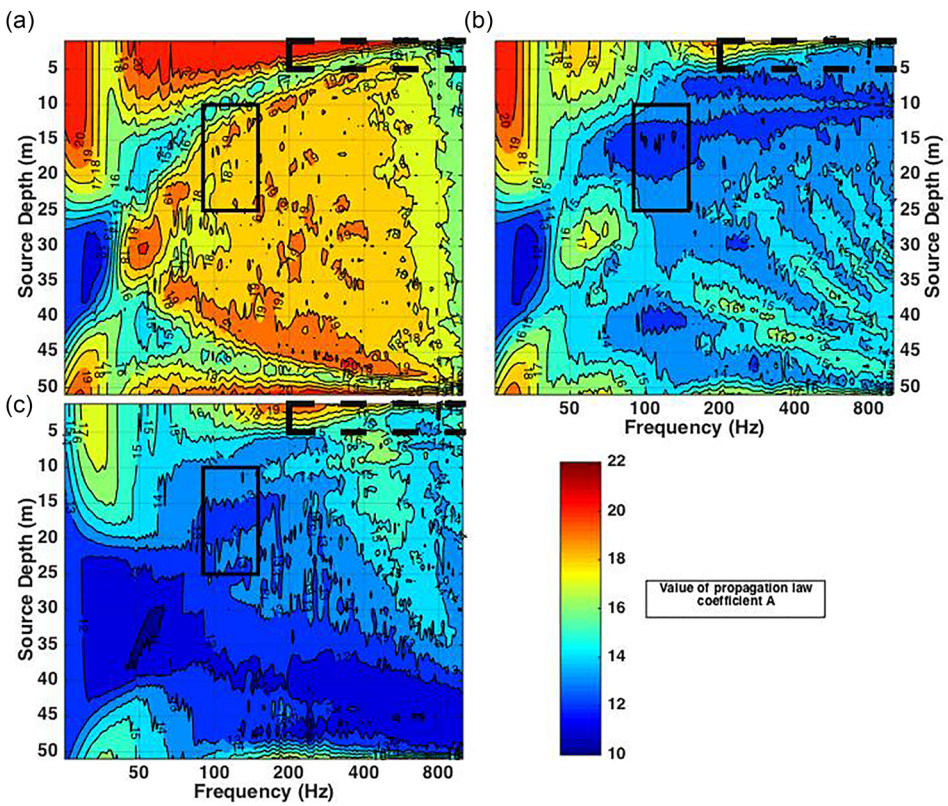

FIG. 9. (Color online) The effect of source depth and frequency on propagation loss on a (a) 1 m, (b) 10 m, and (c) 44 m deep receiver at site BS3 during summer propagation conditions. The first two receiver depths model representative whale listening depths, while the last depth represents the AURAL recording depth. For every combination and source depth and frequency, a best-fit power law coefficient A is obtained by matching the equation $A\log_{10}(r)$ to Eq. (1), computed out to a 10 km range. The resulting values of A are contoured here. Larger values of A represent less efficient acoustic propagation to a 10 km range. The solid boxes display the bandwidth and depth distribution of NPRW upcalls from Fig. 8(b), while the dashed boxes in the upper right corners show the corresponding results for gunshots (which actually extend into multiple kHz). Results are similar for fall sound speed profiles and for power-law fits to other maximum ranges.

data. The solid rectangles represent the spread of depths and frequencies exhibited by upcalls. One sees that at frequencies below 200 Hz, an acoustic source needs to descend deeper in order to achieve efficient propagation (a low value of A). Figure 9(b) shows that the depth distribution observed in Fig. 8(b) spans the most efficient propagation regime for 100–150 Hz signals received at 10 m at the BS3 site, while Fig. 9(a) shows that this upcall depth distribution (solid line box) also incorporates the narrow span of source depths that propagate most efficiently to a whale resting at the surface (1 m receiver depth).

The dashed black rectangles represent the spread of depths and frequencies covered by gunshot sounds (which actually extend into the multiple kHz range). Although the gunshots are generated at shallow depths, their higher frequencies mean that they do not suffer as high a near-surface attenuation that the lower-frequency upcall sounds would at that depth. Thus a possible factor in the shallower depth distributions of gunshots is that their high-frequency content mitigates their shallow emission depth, allowing efficient propagation.

However, Fig. 9(b) also provides a caveat about the interpretation of Fig. 8(b): as calls produced between 10 and 25 m have better propagation conditions than calls produced at the surface, there is a possibility that the depth distribution in Fig. 8(b) has become biased by the relatively better detectability of deeper calls (i.e., upcalls generated at shallow depths may be less likely to be detected and analyzed). The overall range distributions seem similar between the two call types in Tables I and II, alleviating concerns that upcalls are relatively easier to detect. However, this potential detection bias may provide an explanation for why the upcall depth distributions observed in this paper are slightly deeper than those obtained from NARW tagging, which were measured directly on the animals and thus would not suffer this potential detection bias.

Figure 9(c) illustrates how knowledge of the source depth becomes an important factor in estimating accurate source levels for baleen whale signals using bottom-mounted instruments. For example, Fig. 9(c) shows how a 100 Hz signal propagating 10 km can exhibit an effective propagation law that lies between $12\log_{10}$ R to 16 $\log_{10}$ R over a 10 km range, depending on the source depth. This difference can yield up to a $(16–12)\log_{10}(10\,000) = 16$ dB error in estimated source levels of upcalls, even if the source range is precisely known. An 8–16 dB error in source level estimates can lead to modeled detection ranges that are off by factors of 4 to 17, assuming an actual effective propagation law of $13\log_{10}$ R.

V. CONCLUSION

Depth distributions of two types of NPRW calls, upcalls, and gunshots, have been obtained from nonlinear analysis of archival passive acoustic data. The depth distributions of upcalls are found to be deeper than the generally shallower gunshots, a result consistent with the only other data available on the subject: bioacoustic tags placed on NARWs (Parks *et al.*, 2011). Acoustic propagation modeling using environments estimated from whale call data found that a potential factor behind the deeper calling depths for upcalls is that the relatively low frequency calls (100–150 Hz) experienced enhanced propagation efficiency at depths below 10 m. This enhanced propagation for deeper calls might also bias the observed depth distributions toward deeper calls that are more easily detected. The propagation analysis also illustrates how estimating a proper source depth for calling animals is an important factor in estimating their source levels and their detection ranges under various environmental and ambient noise conditions.

Recent work on depths of bowhead whale FM sweeps in the Beaufort Sea (Thode *et al.*, 2016) found that a sample of 63 000 calls were distributed over a broad swath between 15 and 35 m depth, with the peak of the distribution between 25 and 30 m depth. This distribution is slightly deeper than the NPRW FM upcall distribution in Fig. 8(b), raising the question as to whether species-specific differences exist in the calling depth distributions for superficially similar calls, such as gunshot sounds produced by both bowhead and NPRWs in the Bering Sea, and FM upsweeps produced by humpback, bowhead, and NPRWs in the same region. If such differences exist then modal multipath arrivals might provide clues for identifying the species of ambiguous calls.

ACKNOWLEDGMENTS

The authors would like to thank Cedric Arisdakessian for developing much of the time warping software used for the bulk analysis. The specific analysis conducted here was supported by the North Pacific Research Board, Grant No. 1416, and this is NPRB publication 633. Funding for data collection was provided by the Bureau of Ocean Energy Management under Inter-Agency Agreement M08PG20021; we thank J. Denton and C. Fairfield at BOEM for their project support, and the captain and crew of the F/V Aquila. Funding for the initial marine mammal analysis of the archival dataset was provided to D.W. by the International Fund for Animal Welfare. The findings and conclusions in the paper are those of the authors and do not necessarily represent the views of the National Marine Fisheries Service, NOAA.

[1]Reference to trade names does not imply endorsement by the National Marine Fisheries Service (NMFS) nor the National Oceanographic and Atmospheric Administration (NOAA).

Abadi, S. H., Thode, A. M., Blackwell, S. B., and Dowling, D. R. (**2014**). "Comparison of three methods of ranging bowhead whale calls in a shallow-water waveguide," J. Acoust. Soc. Am. **136**, 130–144.

Baggeroer, A. B., Kuperman, W. A., and Mikhalevsky, P. N. (**1993**). "An overview of matched-field methods in ocean acoustics," IEEE J. Ocean Eng. **18**, 401–424.

Bonnel, J., and Chapman, N. R. (**2011**). "Geoacoustic inversion in a dispersive waveguide using warping operators," J. Acoust. Soc. Am. **130**, EL101–EL107.

Bonnel, J., Gervaise, C., Nicolas, B., and Mars, J. I. (**2012**). "Single-receiver geoacoustic inversion using modal reversal," J. Acoust. Soc. Am. **131**, 119–128.

Bonnel, J., Gervaise, C., Roux, P., Nicolas, B., and Mars, J. I. (**2011**). "Modal depth function estimation using time-frequency analysis," J. Acoust. Soc. Am. **130**, 61–71.

Bonnel, J., Le Touze, G., Nicolas, B., and Mars, J. I. (**2013**). "Physics-based time-frequency representations for underwater acoustics," IEEE Signal Process. Mag. **30**, 120–129.
Bonnel, J., Le Touzé, G., Nicolas, B., Mars, J. I., and Gervaise, C. (**2008**). "Automatic and passive whale localization in shallow water using gunshots," in *IEEE OCEANS 2008*, pp. 1–8.
Bonnel, J., Nicolas, B., Mars, J. I., and Walker, S. C. (**2010**). "Estimation of modal group velocities with a single receiver for geoacoustic inversion in shallow water," J. Acoust. Soc. Am. **128**, 719–727.
Bonnel, J., Thode, A. M., Blackwell, S. B., Kim, K., and Macrander, A. M. (**2014**). "Range estimation of bowhead whale (Balaena mysticetus) calls in the Arctic using a single hydrophone," J. Acoust. Soc. Am. **136**, 145–155.
Brekhovskikh, L. M., and Lysanov, Y. (**1991**). *Fundamentals of Ocean Acoustics* (Springer-Verlag, Berlin), 420 pp.
Clapham, P. J., Kennedy, A. S., Rone, B. K., Berchok, C. L., Crance, J. L., and Zerbini, A. N. (**2012**). "North Pacific right whales (*Eubalaena japonica*) in the southeastern Bering Sea," Final Report, OCS Study BOEM 2012-074, National Marine Mammal Laboratory, Alaska Fisheries Science Center, NOAA, 175 pp.
Collins, M. D., and Kuperman, W. A. (**1991**). "Focalization—Environmental focusing and source localization," J. Acoust. Soc. Am. **90**, 1410–1422.
Crance, J., and Berchok, C. (**2016**). "Stereotyped, repetitive gunshot call patterns produced by the North Pacific right whale, *Eubalaena japonica*," J. Acoust. Soc. Am. **140**, 3295–3296.
Gerstoft, P. (**1994**). "Global inversion by genetic algorithms for both source position and environmental parameters," J. Comp. Acoust. **2**, 251–266.
Goldberg, D. E. (**1989**). *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison-Wesley, Reading, MA), 329 pp.
Hinich, M. J., and Sullivan, E. J. (**1989**). "Maximum-likelihood passive localization using mode filtering," J. Acoust. Soc. Am. **85**, 214–219.
Jensen, F. B., Kuperman, W. A., Porter, M. B., and Schmidt, H. (**2011**). *Computational Ocean Acoustics* (AIP, New York), 530 pp.
Locarnini, R. A., Mishonov, A. V., Antonov, J. I., Boyer, T. P., Garcia, H. E., Baranova, O. K., Zweng, M. M., and Johnson, D. R. (**2009**). *World Ocean Atlas 2009, Temperature*, Vol. 1, NOAA Atlas NESDIS 68, U.S. Government Printing Office, Washington, DC, 184 pp.
Lu, I. T., Chen, H. Y., and Voltz, P. (**1993**). "A matched-mode processing technique for localizing a transient source in the time domain," J. Acoust. Soc. Am. **93**, 1365–1373.
Mathias, D., Thode, A. M., Straley, J., and Andrews, R. D. (**2013**). "Acoustic tracking of sperm whales in the Gulf of Alaska using a two-element vertical array and tags," J. Acoust. Soc. Am. **134**, 2446–2461.
McDonald, M. A., and Moore, S. E. (**2002**). "Calls recorded from North Pacific right whales (*Eubalaena japonica*) in the eastern Bering Sea," J. Cetacean Res. Manage. **4**(3), 261–266.
Munger, L. M., Wiggins, S. M., and Hildebrand, J. A. (**2011**). "North Pacific right whale up-call source levels and propagation distance on the southeastern Bering Sea shelf," J. Acoust. Soc. Am. **129**, 4047–4054.
Munger, L. M., Wiggins, S. M., Moore, S. E., and Hildebrand, J. A. (**2008**). "North Pacific right whale (*Eubalaena japonica*) seasonal and diel calling patterns from long-term acoustic recordings in the southeastern Bering Sea, 2000-2006," Marine Mammal Sci. **24**, 795–814.
Newhall, A. E., Lin, Y. T., Lynch, J. F., and Baumgartner, M. (**2009**). "Sei whale localization and vocalization frequency sweep rate estimation during the New Jersey Shallow Water 2006 experiment (A)," J. Acoust. Soc. Am. **125**, 2738.
Parks, S. E., Hamilton, P. K., Kraus, S. D., and Tyack, P. L. (**2005**). "The gunshot sound produced by male North Atlantic right whales (*Eubalaena glacialis*) and its potential function in reproductive advertisement," Marine Mammal Sci. **21**, 458–475.
Parks, S. E., Hotchkin, C. F., Cortopassi, K. A., and Clark, C. W. (**2012**). "Characteristics of gunshot sound displays by North Atlantic right whales in the Bay of Fundy," J. Acoust. Soc. Am. **131**, 3173–3179.
Parks, S. E., Searby, A., Celerier, A., Johnson, M. P., Nowacek, D. P., and Tyack, P. L. (**2011**). "Sound production behavior of individual North Atlantic right whales: Implications for passive acoustic monitoring," Endangered Species Res. **15**, 63–76.
Porter, M. B. (**1991**). "The KRAKEN normal mode program," SM-245, SACLANT Undersea Research Centre.
Rone, B. K., Berchok, C. L., Crance, J. L., and Clapham, P. J. (**2012**). "Using air-deployed passive sonobuoys to detect and locate critically endangered North Pacific right whales," Marine Mammal Sci. **28**, E528–E538.
Thode, A. M., Blackwell, S. B., Seger, K. D., Conrad, A. S., Kim, K. H., and Macrander, A. M. (**2016**). "Source level and calling depth distributions of migrating bowhead whale calls in the shallow Beaufort Sea," J. Acoust. Soc. Am. **140**, 4288–4297.
Thode, A. M., D'Spain, G. L., and Kuperman, W. A. (**2000a**). "Matched-field processing and geoacoustic inversion of blue whale vocalizations," J. Acoust. Soc. Am. **107**, 1286–1300.
Thode, A. M., Kuperman, W. A., D'Spain, G. L., and Hodgkiss, W. S. (**2000b**). "Matched-field processing using Bartlett sidelobe interference structures," J. Acoust. Soc. Am. **107**, 278–286.
Wade, P., Heide-Jørgensen, M. P., Shelden, K., Barlow, J., Carretta, J., Durban, J., DeDuc, R., Munger, L., Rankin, S., Sauter, A., and Stinchcomb, C. (**2006**). "Acoustic detection and satellite-tracking leads to discovery of rate concentration of endangered North Pacific right whales," Biol. Lett. **2**, 417–419.
Wade, P. R., Kennedy, A., LeDuc, R., Barlow, J., Carretta, J., Shelden, K., Perryman, W., Pitman, R., Robertson, K., Rone, B., and Salinas, J. C. (**2010**). "The world's smallest whale population?," Biol. Lett. **7**, 83–85.
Wiggins, S., McDonald, M., Munger, L., Moore, S., and Hildebrand, J. (**2004**). "Waveguide propagation allows range estimates for North Pacific right whales in the Bering Sea," Canadian Acoust. **32**, 146–154.
Wright, D. L. (**2015**). "Simultaneous identification of four mysticete species in the Bering Sea using passive acoustic monitoring increases confidence in acoustic identification of the critically endangered North Pacific right whale (*Eubalaena japonica*)," Final Report for the International Fund for Animal Welfare, Alaska Fisheries Science Center Marine Mammal Laboratory, NOAA, 63 pp.
Yang, T. (**1990**). "Effectiveness of mode filtering: A comparison of matched-field and matched-mode processing," J. Acoust. Soc. Am. **87**, 2072–2084.