Title

Source level and calling depth distributions of migrating bowhead whale calls in the shallow Beaufort Sea

Journal

The Journal of the Acoustical Society of America, 140(6)

ISSN

0001-4966

Authors

Thode, Aaron M
Blackwell, Susanna B
Seger, Kerri D
et al.

Publication Date

2016-12-01

DOI

10.1121/1.4968853

Peer reviewed

Source level and calling depth distributions of migrating bowhead whale calls in the shallow Beaufort Sea

Aaron M. Thode, Susanna B. Blackwell, Kerri D. Seger, Alex S. Conrad, Katherine H. Kim, and A. Michael Macrander

Citation: The Journal of the Acoustical Society of America **140**, 4288 (2016); doi: 10.1121/1.4968853
View online: http://dx.doi.org/10.1121/1.4968853
View Table of Contents: http://asa.scitation.org/toc/jas/140/6
Published by the Acoustical Society of America

Source level and calling depth distributions of migrating bowhead whale calls in the shallow Beaufort Sea

Aaron M. Thode,[1,a)] Susanna B. Blackwell,[2] Kerri D. Seger,[3] Alex S. Conrad,[2] Katherine H. Kim,[2] and A. Michael Macrander[4]

[1]*Marine Physical Laboratory, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093-0238, USA*

[2]*Greeneridge Sciences, Incorporated, 90 Arnold Place, Suite D, Santa Barbara, California 93117, USA*

[3]*University of New Hampshire, 24 Colovos Road, Durham, New Hampshire 03824, USA*

[4]*Shell Exploration and Production Company, 3601 C Street, Anchorage, Alaska 99503, USA*

(Received 6 June 2016; revised 4 November 2016; accepted 15 November 2016; published online 15 December 2016)

Automated and manual acoustic localizations of migrating bowhead whales were used to estimate source level and calling depth distributions of their frequency-modulated-modulated calls over seven years between 2008 and 2014. Whale positions were initially triangulated using directional autonomous seafloor acoustic recorders, deployed between 25 and 55 m water depth near Kaktovik, Alaska, during the fall westward migration. Calling depths were estimated by minimizing the "discrepancy" between source level estimates from at least three recorders detecting the same call. Applying a detailed waveguide propagation model to the data yielded broadband source levels of 161 ± 9 dB re 1 μPa^2 s at 1 m (SEL) for calls received between 20 and 170 Hz. Applying a simpler $15 \log_{10}(R)$ power-law propagation model yielded SEL source levels of 158 ± 10 dB. The most probable calling depths lay between 22 and 30 m: optimal depths for long-range acoustic signal transmission in this particular environment. © *2016 Acoustical Society of America*.
[http://dx.doi.org/10.1121/1.4968853]

[JAS] Pages: 4288–4297

I. INTRODUCTION

During the open-water seasons of 2007–2014, the Shell Exploration and Production Company (SEPCO) commissioned Greeneridge Sciences, Inc., to deploy directional autonomous seafloor acoustic recorders (DASARs) at five sites in the coastal Beaufort Sea (Fig. 1). The motivation behind the recording effort was to evaluate the potential impact of airgun and other industrial sounds on bowhead whale (*Balaena mysticetus*) behavior during the whales' fall migration in the relatively shallow arctic waters off Alaska.

During each of the eight recording seasons, tens to hundreds of thousands of bowhead calls were recorded. The scale of the dataset, combined with a need for timely analysis, required automatically detecting, classifying, and localizing bowhead whale sounds by exploiting the directional localization capabilities of the DASAR packages (Thode *et al.*, 2012). The results of the automated analyses were used to track seismic airgun activity around the Beaufort Sea (Thode *et al.*, 2010) and to help determine that bowhead whales change their sound production rates in response to both nearby and distant seismic surveys (Blackwell *et al.*, 2015).

Here, a portion of this automatically analyzed dataset has been combined with supplemental manual analysis in order to estimate bowhead whale call source level and calling depth distributions for frequency-modulated (FM) "moans" or "sweeps." Section II describes the equipment used, the automated detection and localization software, and the manual analysis procedures. It then details how a subset of these localizations was selected, and how the acoustic propagation environment was modeled using both simple power-law transmission loss models and more detailed waveguide propagation models. Section III displays the resulting broadband source level and source depth distribution plots obtained from the automated analyses, using three different propagation models of varying complexity. Section IV discusses the consistency of the source level distributions with previous literature and provides a physical interpretation for why whales would call at the estimated source depths.

II. METHODS

A. Equipment and deployment configuration

The acoustic data for this study were recorded on "model-C" DASARs (Greene *et al.*, 2004), autonomous acoustic recording packages equipped with one omnidirectional acoustic pressure sensor (-149 dB re V/1 μPa) and two horizontal directional sensors that measure the north-south and east-west components of acoustic particle velocity. This arrangement permits the azimuth of bowhead whale sounds to be measured from each DASAR. Each time series is sampled at 1 kHz and has a maximum usable acoustic frequency of 450 Hz. The Global Positioning System (GPS)-synchronized time is noted to within a second whenever each DASAR is started and stopped. In addition, shortly after the deployment and before the retrieval of every DASAR, a calibrated FM signal is sequentially broadcast at

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

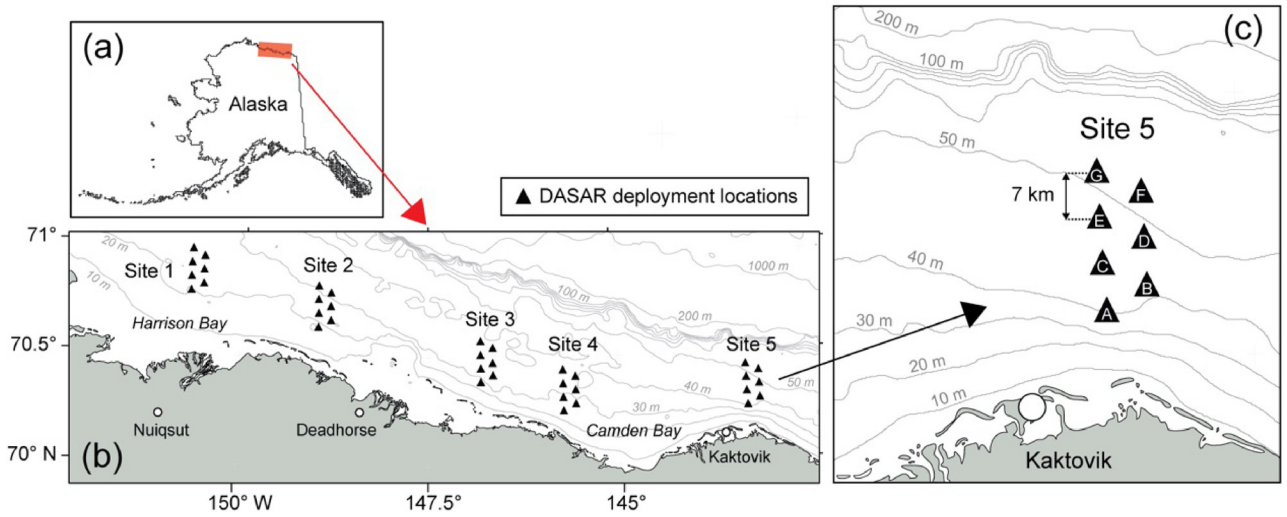

FIG. 1. (Color online) Location and bathymetry of DASAR deployments near Kaktovik, AK. (a) and (b) show the general locations of all instruments, while (c) shows the detailed deployment geometry and bathymetry of Site 5 (the focus of this paper).

roughly 4 km range from three to six positions around each recorder. These playback start times are also time-stamped with GPS data, which are used to correct each DASAR clock's drift during post-processing. Bearings to sounds of interest, derived by comparing the active acoustic intensity measured along orthogonal directions, have around 15°–20° precision for signals with signal-to-noise ratios (SNRs) less than 5 dB, and up to 1°–2° precision for signals with SNR greater than 10 dB (Greene *et al.*, 2004). Bearings to coincident calls are combined via triangulation (Lenth, 1981) to yield two-dimensional (2-D) call positions, yielding ranges from each DASAR that detected the call.

From 2008 to 2014, between 35 and 40 DASARs were deployed across a 280 km swath of continental shelf off the Alaskan North Slope in water depths of 20–53 m. The deployments were grouped into "sites," labeled 1–5 from west to east (Fig. 1). While DASAR deployments were also conducted in 2007, those instruments used different sensors than subsequent years and were thus excluded from the following analysis.

Most sites contained seven DASARs deployed in a triangular grid with 7 km separation. The southernmost DASAR of each site was labeled "A," and the northernmost was labeled "G," so the label "DASAR 5 G" refers to the deepest DASAR data collected at Site 5, location G [Fig. 1(c)]. This paper focuses only on Site 5, the easternmost and deepest site located north of Kaktovik, Alaska. As Fig. 1(c) suggests, the bathymetry of the area was gently sloping, deepening gradually from 37 to 53 m with increasing latitude. DASARs 5D through 5 G sat in 53 m depth water, 5B and 5 C sat at 46 m depth, and 5 A, the southernmost and shallowest DASAR, sat at 37 m depth. Much of the site, therefore, featured relatively uniform bathymetry, a fact that simplified acoustic propagation modeling.

B. Automated detection and localization

Every year, a six-stage automated program extracted and localized bowhead whale calls in the raw acoustic data collected from Site 5, as detailed in Thode *et al.* (2012). After some initial event detection and image preprocessing stages, two neural networks evaluated whether transient signals were whale calls. The program then linked together detections on individual DASARs, and the bearings derived from the detections were used to triangulate a robust maximum-likelihood position of the animal (Lenth, 1981; Greene *et al.*, 2004), along with a 90% confidence ellipse that yielded an estimate of localization precision. Complete details on both the algorithm and its evaluation are provided in Thode *et al.* (2012). As a by-product of this analysis, the automated processing also generated received levels of FM detections in terms of both sound exposure level (SEL) and sound pressure level (SPL), the latter being the root-mean-square (rms) measurement of the pressure time series, expressed in dB units.

C. Manual analysis

Each year of the project, between five and eight non-contiguous days of a deployment were selected for detailed manual analysis, with the intention of performing quality assurance checks on subsets of the complete season automated analysis. A team of trained analysts reviewed randomly assigned hours within the chosen dates, using custom-written MATLAB software. Identification and classification of each whale call was done by examining spectrograms of the acoustic data, one minute at a time, and by listening to recordings of each call or suspected call. The same minute at each DASAR location of a site was shown as a series of spectrograms on the analysts' screens. Using a mouse, the analysts drew a bounding box around every occurrence of a particular call for each DASAR that detected it. The bounding boxes allowed the software to calculate parameters such as the call's duration (width of the bounding box) and bandwidth (height of the bounding box), minimum frequency detected, received rms SPL and SEL, SNR, and bearing, the last of which was used to triangulate each call's position. Once all the calls included in the minute had been marked,

TABLE I. Call samples used to estimate source level distribution vs year. Columns 4–7 are restricted to locations with an equivalent 90% uncertainty radius of 150 m or less. "Percent of total" indicates what fraction each individual year contributed to the total sample size listed in the "Total" row. Note that the manual analysis was incomplete in 2013 due to very high call numbers; only a subset of the hours of the analyzed days was analyzed manually. Also, note that fewer days in 2014 were analyzed manually compared to the previous years; as a result, 2014 contributes relatively fewer samples to the supplemental manual analysis than to the automated analysis.

1	2	3	4	5	6	7
Year	Total automated positions available (percent of total)	Total manual positions available (percent of total)	Automated calls within 15 km of a DASAR (percent of total)	Manual calls within 15 km of a DASAR (percent of total)	Automated calls within 2 km of a DASAR (percent of total)	Manual calls within 2 km of a DASAR (percent of total)
2008	107 643 (14)	19 175 (13)	10 576 (17)	1414 (19)	2169 (15)	292 (18)
2009	81 235 (10)	20 830 (14)	6849 (11)	1161 (16)	1377 (10)	203 (12)
2010	99 934 (13)	13 724 (9)	10 478 (17)	677 (9)	1744 (12)	168 (10)
2011	49 254 (6)	13 008 (9)	4348 (7)	260 (4)	534 (4)	23 (1)
2012	70 835 (9)	15 967 (11)	6128 (10)	1295 (18)	1720 (12)	292 (18)
2013	253 679 (32)	46 563 (31)	11 866 (19)	1961 (27)	2613 (18)	522 (32)
2014	126 051 (16)	20 183 (13)	12 705 (20)	631 (9)	4001 (28)	139 (8)
Total	788 631	149 450	62 950	7399	14 149	1639

the analyst moved on to the next minute of data. The lead analyst, who was the same person throughout all the years of the study, performed regular checks for consistency between analysts.

D. Sample selection for source level distribution estimation

While bowhead whales make a variety of sounds, only FM moans are analyzed here. The second and third columns of Table I summarize call positions obtained by the automated and manual analyses for Site 5 between 2008 and 2014. The "percent of total" values indicate what fraction each individual year contributed to the total sample size listed in the "Total" row. For example, in column 2, the year 2013 provided 32% of the total of 788 631 total automated localizations initially available at Site 5. Nevertheless, only subsets of these calls, shown in columns 4–7 of Table I, were used to estimate source level distributions. Specifically, the source level of a given call was estimated only if the call met the following four criteria:

(a) The call had to be detected on three or more DASARs at ranges less than 50 km from every DASAR.
(b) The call's range to the closest DASAR ($R_{\min}$) had to be less than two selected values: 2 km and 15 km. The former value was chosen because previous analysis has shown that DASARs are nearly 100% effective in detecting calls generated at 2 km range or less, regardless of ambient noise conditions (Blackwell *et al.*, 2015). The 15 km value, which roughly corresponds to the distance spanned by three in-line DASARs, was selected to increase the available sample size. The 15 km choice was also used to check the accuracy of the propagation models employed, since an accurate propagation model should generate the same source level distribution regardless of the choice of $R_{\min}$.
(c) The radius of a circle with the same area as the call's 90% confidence ellipse had to be less than 150 m, (e.g., less than ~7% of the range to the nearest DASAR for the 2 km scenario). Changing this uncertainty to 300 m increased sample sizes by 50% but otherwise did not alter the final results.
(d) The fundamental frequency of the call type used for this project (an FM "moan" sweep) had to lie between 20 and 170 Hz, frequencies with acoustic wavelengths of 8 m or greater. This restriction justified the assumption that the source was effectively omnidirectional, as acoustic radiators with dimensions smaller than an acoustic wavelength tend to radiate non-directional signals.

Columns 4 and 5 of Table I reveal that 8% (62 950 calls) of the original automated Site 5 dataset and 5% (7399 calls) of the manually analyzed dataset met these criteria when $R_{\min} = 15$ km, and that 2% (14 149 calls) and 1% (1639 calls) of the automated and manual datasets, respectively, met the criteria when $R_{\min} = 2$ km (columns 6 and 7).

E. Source level estimation

Figure 2 defines the nomenclature used for estimating source levels from received level (RL) measurements from multiple DASARs. For a given DASAR i, the source level estimate SL_i can be obtained from the simplest version of the sonar equation, expressed in terms of dB units

$$\mathrm{SL}_i = \mathrm{RL}_i + \mathrm{TL}_i(R_i, \gamma_i). \quad (1)$$

The broadband received level can be expressed in terms of a variety of units, including SEL or rms SPL. The transmission loss (TL) is shown as an explicit function of the distance R_i between DASAR i and a calling whale; other parameters that influence TL are collectively represented by the symbol γ_i. Examples of such parameters for a detailed propagation model include acoustic frequency, source and receiver depths, ocean bathymetry, waterborne sound speed profile, sediment compressional and shear speeds and gradients, and sediment density and attenuation.

The source level obtained by analyzing the received level $\mathrm{RL}_{\min}$ on the closest DASAR at range $R_{\min}$ is defined as $\mathrm{SL}_{\min}$, which is presumed to be the default estimate of source level for the call, as weaker components of the call

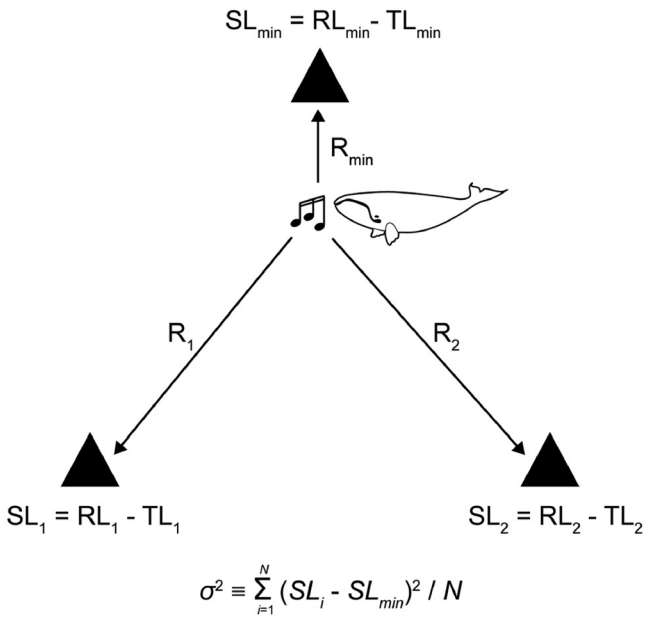

FIG. 2. Illustration of source level inversion method, including definition of various nomenclature.

may not be detected on other DASARs at larger ranges. All source level estimates SL_i should be the same as SL_{min}, provided that the following assumptions are met: the modeled propagation environment is identical to the true propagation environment, the acoustic source is omnidirectional, all components of the sweep are detected at even the most distant DASARs, and appropriate source level units are chosen. However, if the modeled environment does not match the true environment, then the source level estimate of a single call will diverge across receivers, a feature quantified here as a *discrepancy* (σ^2)

$$\sigma^2 \equiv \sum_{i=1}^{N} (\mathrm{SL}_i - \mathrm{SL}_{\min})^2 / N, \tag{2}$$

where N is the number of DASARs used to obtain a particular localization. While the discrepancy may have the form of a statistical variance, it is not one, since the differences in Eq. (2) are taken with respect to SL_{min} and not the mean value of SL_i, as would be the case for a true variance measure. By applying Eq. (2) to all calls in a sample, a distribution of discrepancy values can be created. Both the source level distribution and parameters of the transmission loss model can then be simultaneously estimated from the data by selecting transmission loss parameters that minimize the mean value of the discrepancy distribution.

1. Power-law transmission loss model

The procedure described above was first applied to a classic one-parameter "power-law" propagation model, $\mathrm{TL} = A \log_{10}(R)$, the simplest ocean propagation model. R represents the horizontal separation between the source and receiver in meters, and A is a coefficient that characterizes the propagation environment. In a "spherical spreading" case, where a source is radiating into free space, $A = 20$, whereas a source radiating into a waveguide with perfectly reflecting and lossless boundaries experiences $A = 10$. Most realistic ocean environments display values for A that lie between these two extremes.

There are two reasons for discussing this simple model, even though the following results employ more complex models: first, Sec. III uses an optimized power-law propagation model to cross-check the estimates of the more detailed models; second, this model provides physical insight into the relationship between a call's estimated source level and its discrepancy. This latter point is demonstrated by combining the power-law transmission loss formula with Eqs. (1) and (2), yielding

$$\Delta_i \equiv \mathrm{SL}_i^{\mathrm{est}} - \mathrm{SL}^{\mathrm{true}} = (A_{\mathrm{est}} - A_{\mathrm{true}}) \log_{10}(R_i). \tag{3}$$

$$\sigma^2 = \sum_{i=1}^{N} (A_{\mathrm{est}} - A_{\mathrm{true}})^2 [\log_{10}(R_i) - \log_{10}(R_{\min})]^2 / N. \tag{4}$$

The term "est" indicates the estimated value, while "true" indicates the actual underlying value. If $A_{\mathrm{est}} < A_{\mathrm{true}}$ then the SL is underestimated, and if $A_{\mathrm{est}} > A_{\mathrm{true}}$, SL is overestimated [Fig. 3(b)]. Whenever $A_{\mathrm{est}} \neq A_{\mathrm{true}}$, Eqs. (3) and (4) show that the source level estimates SL_{min}^{est} become

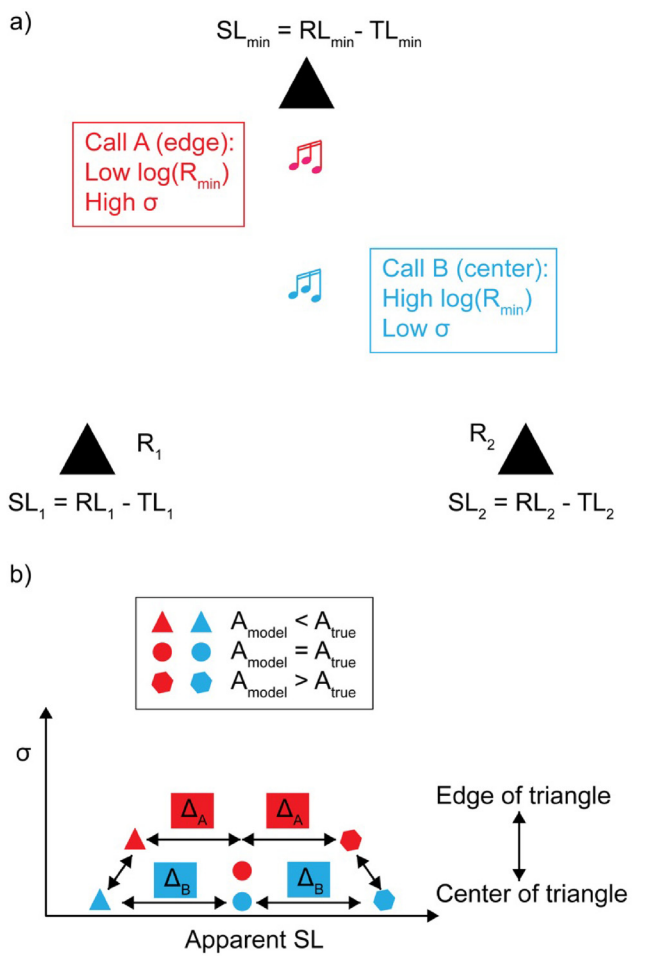

FIG. 3. (Color online) Why apparent source level is correlated with discrepancy whenever a mismatched propagation model is used to estimate source level. (a) displays two example calls inside a triangular grid of sensors and (b) illustrates how these two calls are mapped on a plot of apparent source level vs discrepancy.

correlated with their corresponding discrepancies. Figure 3 provides an intuitive explanation for why this occurs. If a whale calls very close to one DASAR with respect to the others [Call "A" in Fig. 3(a)], then $R_{\min}$ will be much smaller than the other R_i and the call's discrepancy will be large via Eq. (4); however, the shift in apparent source level for $SL_{\min}$ [Δ_A in Fig. 3(b)] will be relatively small, since $\log_{10}(R_{\min})$ will be relatively small in Eq. (3). By contrast, whenever a whale call is nearly equidistant to multiple receivers [Call "B" in Fig. 3(a)], the shift in apparent source level [Δ_B in Fig. 3(b)] will be large because of a relatively large $\log_{10}(R_{\min})$ in Eq. (3), but the discrepancy will be relatively small since $R_{\min} \sim R_i$ for all i.

Thus, plotting $SL_{\min}^{\text{est}}$ vs discrepancy for all calls, as schematically shown in Fig. 3(b), provides a convenient graphical way of deducing the direction of model mismatch; when $A_{\text{est}} < A_{\text{true}}$, the scatter plot will display a positive correlation, and vice versa. Similar properties exist for more complex propagation models, and so upcoming sections plot estimated source levels against discrepancy to provide a visual check that the source level estimates are indeed independent of their discrepancy values and, thus, the propagation model used is accurate.

2. Full waveguide propagation model

A full waveguide propagation model has advantages over a power-law model in that it permits the calling depths of animals to be estimated, provided that an accurate model of the ocean environment can be obtained (Mouy *et al.*, 2014; Delarue *et al.*, 2015). In 2010, a 15-element vertical array was deployed near DASAR 5 G and was used to obtain this detailed environmental model. A genetic algorithm global inversion scheme was combined with a normal-mode propagation model (Gerstoft, 1997) to estimate environmental parameters that yielded the best fit to the complex acoustic pressure depth profile generated by several whale calls across the vertical array (Abadi *et al.*, 2014; Bonnel *et al.*, 2014). The closest whale call used for inversion, detected 1.2 km from the array, yielded a bottom sediment compressional speed profile of 1513, 2410, and 3103 m/s at 0, 10, and 20 m depth, respectively, beneath the sediment water interface. The best-fit sediment density was 1.31 g/cc, while the sediment attenuation was relatively poorly constrained between 0.4 and 1.0 dB/wavelength, with a value of 0.5 dB/lambda being selected as being the most representative of the true attenuation. The best-fit waterborne sound speed profile was upward refracting, with a surface sound speed of 1441 m/s and a sound speed gradient of 0.2 (m/s)/m. This completed environment will be labeled as the "full" propagation environment for the rest of this paper. When the propagation model was constrained to an isovelocity sound profile and a uniform sediment halfspace (a "Pekeris" model), then the best-fit Pekeris propagation environment has a bottom speed of 1707 m/s, bottom density of 1.9 g/cc, and attenuation coefficient of 0.5 dB/wavelength. For both models, the receiver depth was set to 0.5 m above the ocean floor.

The models were applied to Eq. (1) by collecting the following information for each call in the sample subset:

(a) The minimum frequency attained by the FM call;
(b) The estimated bottom depth beneath the call, obtained by interpolating a call's 2-D position within a bathymetry map (using the actual measured DASAR depths as an accuracy check); and
(c) The range of the call from each DASAR involved in its localization.

The parameters in (a) and (b) are concrete examples of propagation parameters abstractly represented by γ in Eq. (1). The "best-fit source depth" was estimated for each call by combining its frequency range with the previously discussed environments as input to the KRAKEN normal mode program (Porter, 1991). The transmission loss to each DASAR was then computed as a function of source depth in 1 m steps, and the value that minimized the discrepancy of the call via Eq. (2) was selected as the best-fit source depth for that call.

Because the bathymetry within Site 5 got shallower at the southern end of the site, a normal mode model that assumes a constant bottom depth risks being inaccurate at those locations; either an adiabatic normal mode model or a parabolic-equation model should ideally be used, with the accompanying increase in computational time. However, for mildly range-dependent environments, where the bathymetry displays a nearly linear slope, D'Spain *et al.* (1999) found that a flat-bottomed normal mode model yielded the same spatial structure as an acoustic field propagated up a linearly sloping bathymetry, provided that the following adjustments to the source position were made before applying the flat-bathymetry model:

$$z_{\text{adj}} = z_{\text{true}} \left(\frac{D_{\text{DASAR}}}{D_{\text{whale}}} \right), \tag{5}$$

$$r_{\text{adj}} = r_{\text{true}} \left(\frac{D_{\text{DASAR}}}{D_{\text{whale}}} \right). \tag{6}$$

Here z_{true} and r_{true} are the actual depth and range, respectively, of the whale from a given DASAR, D_{DASAR} and D_{whale} are the water depths at a given DASAR's and whale's locations, respectively, and z_{adj} and r_{adj} are the adjusted whale depth and range, respectively, that are inserted into the flat-bathymetry normal mode model to best reproduce the actual acoustic field. This adiabatic depth-optimization procedure had relatively little impact on the source-depth distribution estimation, but did lead to some artifacts, as discussed in Sec. IV B.

III. RESULTS

A. Distributions of received levels

The results presented below focus first on the automated analysis using $R_{\min} = 15$ km across all years, which is the data subset with the largest sample size (Table I, column 4). Figure 4 plots a 2-D histogram of the samples' $RL_{\min}$ vs the discrepancy of the received levels [i.e., with RL replacing SL in Eq. (2)], along with the marginal distributions for both variables.

The mode of the discrepancy distribution is about 8 dB, and the mode of the RL_{min} distribution is 107 dB re 1 μPa^2 s. The received levels and discrepancies in Fig. 4(b) are clearly positively correlated; this effect arises because Fig. 4(e) can be interpreted as a source level estimate using a very mismatched power-law model with $A_{est} = 0$. Figure 3(b) illustrates how the observed positive correlation arises whenever the modeled propagation environment attenuates the acoustic field less harshly with range than the underlying true environment ($A_{est} < A_{true}$).

Figures 4(c) and 4(d) display the distribution of the minimum "robust" frequency f_{min} detected at the closest DASAR for each sample call. This variable represents the frequency that defines the lower edge of a "bounding box" that would be drawn around a call in a spectrogram of the closest DASAR's data. The term "robust" refers to the manner in which the instantaneous frequency of an FM sweep is calculated from a spectrogram; at each given time instant, the "robust" frequency is the intensity-weighted average frequency across the call bandwidth. A clear bimodal distribution in f_{min} exists, with modes at 41 Hz and 99 Hz, roughly corresponding to the fundamental and first harmonic frequencies visible in many calls. However, many of the calls with f_{min} greater than 75 Hz display no 30–50 Hz fundamental. There is a slight tendency for lower frequencies to display higher received levels [Fig. 4(d)].

B. Source level and source depth estimates using automated data set, full propagation model

Figure 5 demonstrates the results of applying the full propagation model to the dataset plotted in Fig. 4, with the source depth optimized to minimize the discrepancy for each call sample. Figure 5 also displays the source level distribution consolidated across all seasons [Fig. 5(i)], along with distributions describing the final source level discrepancy [Fig. 5(a)], call duration [Fig. 5(c)], and f_{min} [Fig. 5(e)]. 2-D histograms relating various combinations of parameters are also shown, with the intensity scale shown in terms of number of samples.

The source level distribution shows a mode at 162 dB re 1 μPa^2 s at 1 m, with a mean (*median*) and standard deviation of 161 (*162*) ± 9 dB re 1 μPa^2 s at 1 m. The typical call duration is centered around 1 s [Fig. 5(c)] and, thus, the broadband source levels of the calls, expressed in dB rms SPL values, should have nearly the same numerical values as the SEL values displayed in Fig. 5(i), which indeed turns out to be the case. Figure 5(b) shows the discrepancy and source level estimates to be relatively uncorrelated, so the detailed propagation model seems accurate.

Figures 5(g) and 5(h) display the estimated source depth distribution generated using the adiabatic approximation discussed in Eqs. (5) and (6). This distribution peaks at 26 m depth and finds a sharp drop-off in call depths below 37 m. It also displays a small but distinctive cluster of calls at less than 7 m depth with associated source levels 5–10 dB higher than the primary distribution [Fig. 5(h)]. As discussed further in Sec. IV B, both the sharp cutoff at 37 m and the surface cluster may be artifacts of the flat-bathymetry propagation model. The call frequency distributions shown in Figs. 5(e) and 5(f) suggest that calls below 50 Hz have 2–3 dB lower source levels than calls above 75 Hz.

C. Source level and source depth estimates using other models and datasets

The source level and source depth distributions displayed in Fig. 5 are similar to those produced by repeated analyses with the following variants:

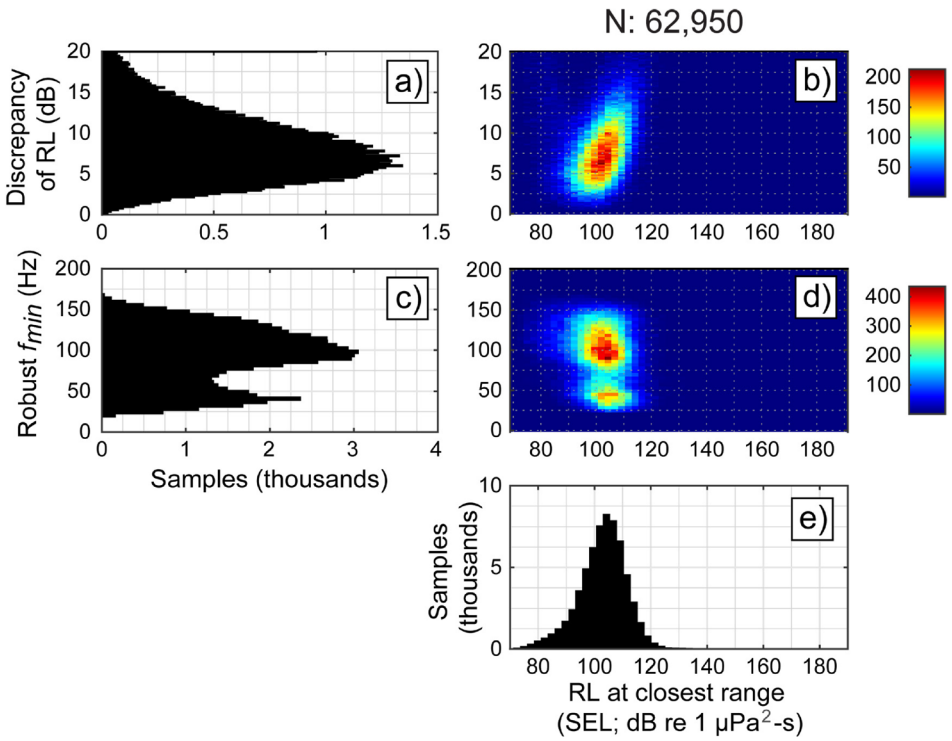

FIG. 4. (Color online) Distributions of raw measurements from 62 950 call samples, selected by the automated algorithm (Thode *et al.*, 2012), and where R_{min} has been set to 15 km (Table I, column 4). (a) Marginal distribution of received level discrepancy [Eq. (2)], bin width 0.2 dB; (b) 2-D histogram of received level at closest DASAR (*RL*) vs discrepancy; (c) marginal distribution of "robust" minimum call frequency f_{min} detected on closest DASAR (4 Hz bin width); (d) 2-D histogram of f_{min} vs *RL*; (e) marginal distribution of broadband received level in units of SEL, bin width 2 dB. Color scale indicates number of calls per 2-D cell.

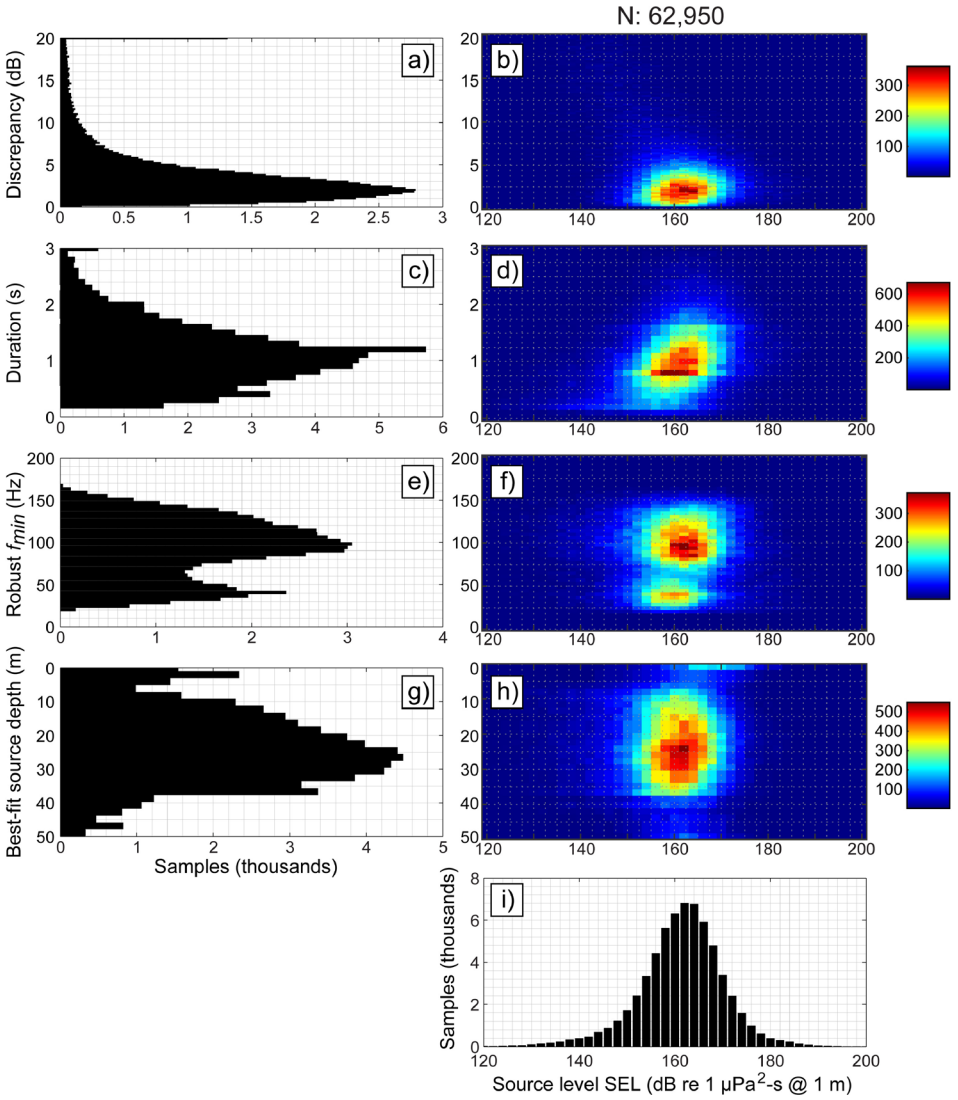

FIG. 5. (Color online) Source level estimates derived from Fig. 4 using a detailed waveguide propagation model obtained from vertical array data. Bin widths for histograms are the same as Fig. 4, unless otherwise noted. (a) Marginal distribution of source level estimate discrepancy [Eq. (2)]; (b) 2-D histogram of discrepancy vs source level estimate; (c) marginal distribution of call duration, bin width 0.1 s; (d) 2-D histogram of call duration vs source level estimate; (e) marginal distribution of $f_{\min}$, bin width 4 Hz; (f) 2-D histogram of $f_{\min}$ vs source level estimate; (g) marginal distribution of source depth estimates (depth that minimizes discrepancy), bin width 2 m; (h) 2-D histogram of best-fit source depth vs source level estimate; and (i) marginal distribution of broadband source level estimate in terms of SEL, bin width 2 dB.

(1) Manually analyzed data are used (7399 calls at $R_{\min} = 15$ km; Table I, column 5), instead of the 62 950 automatically analyzed samples.
(2) $R_{\min}$ is restricted to 2 km instead of 15 km, which shifts the mean source level 2–4 dB lower than that of the larger $R_{\min} = 15$ km subsample.
(3) A simple $15 \log_{10}(R)$ power-law transmission loss model described in Sec. II E 1 is used in lieu of the full propagation model. The coefficient $A = 15$ was obtained by minimizing the mean discrepancy of the $R_{\min} = 2$ km call subsample (Table I, column 6). The simpler Pekeris waveguide model was also tested and provided similar results.

Several of these alternative analyses are displayed in Fig. 6, where the distributions are normalized as probability density functions (PDFs). Three different propagation models are shown for the automated dataset, while only the full propagation model is shown for the manual dataset, to preserve plot legibility.

The modes of both the power-law source level [Fig. 6(d)] and discrepancy distributions [Fig. 6(f)] are similar to those obtained from the full propagation model [Figs. 5(i) and 5(a), respectively), but the mean/median discrepancy is lower (4.2/2.8 dB) for the full model, and Fig. 6(d) shows that the full model distribution (solid line) is less skewed toward lower source levels than the simple power-law model distribution. The similarity in the source level distributions arises from the fact that 50–150 Hz sources placed below 15 m depth in the full propagation environment exhibited an effective propagation loss commensurate with a $15 \log_{10}(R)$ to $18 \log_{10}(R)$ power-law formula out to source ranges of 7 km, when measured by a sensor on the ocean floor.

The Pekeris propagation model generates a larger proportion of shallow-depth (<5 m) calls than the full model [Figs. 6(b) and 6(e)], but otherwise replicates the most likely source depths produced by the full model. Figures 6(c) and 6(f), which histogram the discrepancy distributions for all the analyses, show that the automated source level estimates have modes roughly 2–3 dB less than those from the manually analyzed data. However, they generate larger discrepancy tails than the manual analysis, suggesting that the automated analysis tends to pick up fragments of calls at more distant DASARs, which increases the discrepancy of the source level measurement.

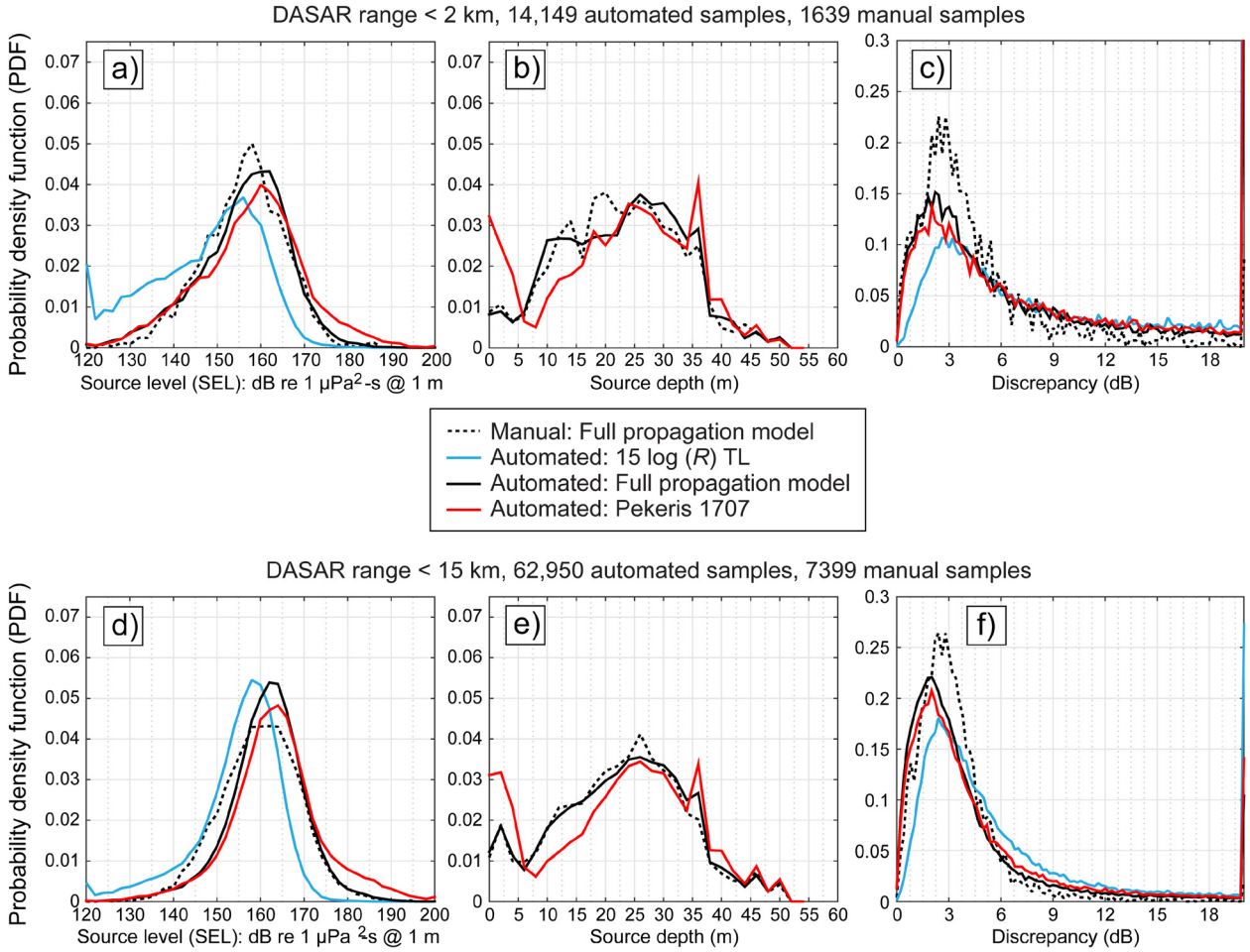

FIG. 6. (Color online) Summary of results from all years, using various analyses and propagation models, plotted in terms of probability density functions (PDFs). The top row displays data subsets where $R_{\min} = 2$ km: (a) source level estimate in terms of SEL (bin width 2 dB); (b) best-fit source depth estimate (bin width 2 m); (c) source level estimate discrepancy (bin width 0.2 dB). The bottom row displays the same results using $R_{\min} = 15$ km: (d) source level; (e) best-fit source depth; and (f) discrepancy. Legend key: cross, red solid line, full propagation model applied to manual analysis; "x," green dashed line, $15 \log_{10}(R)$ propagation model applied to automated analysis; circle, black dotted line, full propagation model applied to automated analysis (same as Fig. 5); cross, blue dashed line, Pekeris propagation model (bottom speed 1707 m/s) applied to automated analysis.

IV. DISCUSSION

A. Consistency of source level estimates between different models and previous literature

The source level distributions obtained by both the manual and automated analyses, as summarized in Fig. 6, are similar, suggesting that the automated algorithms have false alarm (precision) rates sufficiently low enough to adequately represent a manual analyst's selections in terms of call frequency range and received level. The simple power-law transmission loss model provides a peak in the source level distribution that lies within 3 dB of the peak of the full waveguide propagation model distribution. This consistency between both models, derived using different techniques, provides encouragement about the validity of the results, as does the fact that the source levels are uncorrelated with their discrepancies [e.g., Fig. 5(b)]. As detailed in Sec. II E, this lack of correlation indicates a relatively good match between the modeled and underlying propagation environments.

The peaks (modes) of the source level distributions in Figs. 6(a) and 6(d) display some evidence that the full and Pekeris propagation models are slightly mismatched with the true environment. For the automated results, the distributions peak at 160 dB for $R_{\min} = 2$ km [Fig. 6(a)] and 162 dB for $R_{\min} = 15$ km [Fig. 6(d)], with a similar shift observed in the manual analysis, whenever the full propagation model is used. This shift is interpreted as being related to Fig. 3(b); as discussed in Eq. (3) in Sec. II E, whenever the propagation environment used to estimate source levels differs from the underlying true propagation environment, calls associated with lower values of $R_{\min}$ [e.g., Fig. 6(a)] will deviate less from the true source level than calls associated with higher values of $R_{\min}$ [Fig. 6(d)]. The slightly higher source levels observed for $R_{\min} = 15$ km are consistent with a situation in Fig. 3(b) where the modeled propagation environment creates more attenuation with range than the actual environment [i.e., $A_{\text{est}} > A_{\text{true}}$ scenario in Fig. 3(b)].

Despite this small inconsistency between the two $R_{\min}$ values, the most common source levels observed are consistent with two classic papers on bowhead whale sound production: 156 dB re 1 μPa (rms) for Clark and Johnson (1984), and the distribution of source levels shown in Cummings and Holliday (1987) that peak between 156 and

160 dB re 1 μPa (rms) for moan calls. (The "spectral levels" provided in both papers are equivalent to rms source levels for narrowband signals.) Thus, over 25 years, the most common source levels for migrating bowhead whale moan calls have not changed.

B. Explanations for observed source depth profile

The source depth profiles displayed in Figs. 6(b) and 6(e) all show broad peaks between 22 and 28 m depth, regardless of the type of analysis or type of waveguide model used. Figure 7 provides some insight into why animals might choose to call at those depths at Site 5, by mapping the power-law coefficient of A that best fits the computations of the full waveguide propagation model out to 5 km range, for a source broadcasting between 20 and 200 Hz and between 0 and 55 m depth. Figure 7 assumes a flat 55 m bottom (the dominant ocean depth at Site 5) and models a receiving whale at 25 m depth (the most common bowhead whale calling depth). Figure 7 is thus simulating the efficacy of sound transmission between two whales in this environment, and not sound transmission between a whale and a bottom-mounted recorder. Shallow sources display higher propagation losses, from $18\log_{10}(R)$ to $20\log_{10}(R)$, due to destructive interference arising from inverted-phase reflections from the ocean-surface boundary.

One sees that at 40 Hz the optimum calling depth for long-range signal propagation lies between ~15 and 45 m, centered around 30 m. With an increase in source frequency, toward the right of Fig. 7, one sees that optimum calling depths (darker colors) shallow slightly. Thus, the most common source depths found for bowhead whale moan calls are consistent with optimum depths for long-range signal transmission in this particular environment. A similar optimum source depth persists when other receiver depths are modeled, including the receiver depths for DASARs.

This correspondence between estimated whale source depth and optimum propagation calling depth raises a valid issue: could the observed source depth distribution simply arise because calls at 28 m depth are relatively more likely to be detected by a DASAR due to lower propagation losses (resulting in higher SNR) for calls produced at this depth? Figure 6(b) assuages this concern by restricting the call samples to those known to be less than 2 km from the nearest DASAR. As stated previously, earlier work has shown that calls generated at less than 2 km range have a very high probability of being detected under a variety of ambient noise conditions, because estimated call spatial densities as a function of range from the DASAR are constant until 4–5 km range (Blackwell *et al.*, 2013). Thus, call detection rates for "moans" generated inside 2 km are assumed to match actual call production rates. Furthermore, one sees that Fig. 6(b) ($R_{\min} = 2$ km) still matches the smoother distributions of Fig. 6(e) ($R_{\min} = 15$ km) in terms of the distributions' peak calling depth. In other words, the derived depth distributions are not affected by the ranges of the calls used, which would have been the case had the profiles arisen from differences in relative detectability of calls generated at different source depths.

Other features observed in the source depth distributions may be artifacts: the spike at 37 m depth, for example, likely arises because the shallowest DASAR depth (5 A to the south) is 37 m. Because of this, calls generated at the northern end of the site at depths below 37 m must propagate

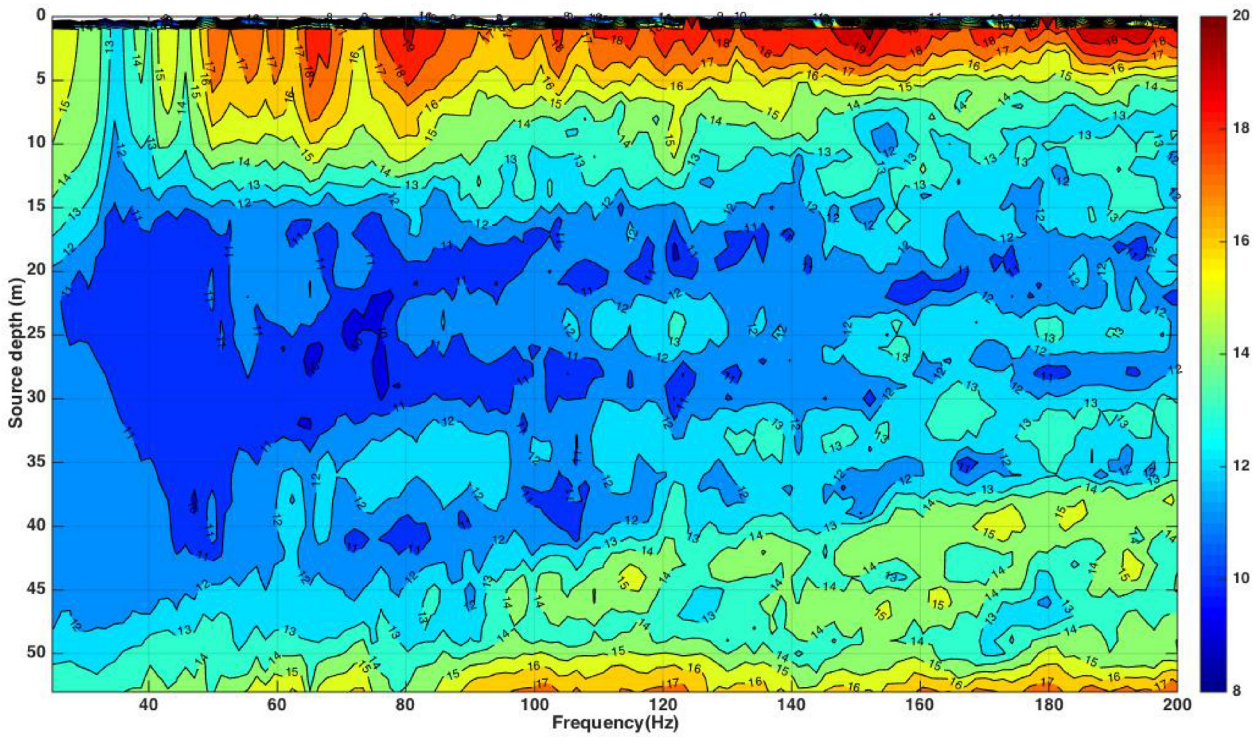

FIG. 7. (Color online) Contour map of effective power-law transmission loss coefficient A_{true} as a function of source frequency and depth, for a 25 m deep receiver in a 53 m waveguide, using the full propagation model out to 5 km range. Higher values of A_{true} indicate poorer propagation conditions.

upslope to DASAR A and rely on the validity of the adiabatic approximations of Eqs. (5) and (6) in order to be mapped correctly. If this approximation fails, then 37 m becomes the best-fit source depth by default. Indeed, one does find that many of the calls in the 37 m "spike" have discrepancies much greater than 6 dB, consistent with the interpretation that 37 m depth is a suboptimal solution for calls that are actually generated at deeper depths.

The greatest puzzle of the depth distributions visible in Figs. 6(b) and 6(e) are the small but distinctive clusters of calls mapped at 2–5 m depth, with associated higher source level estimates of 165–180 dB re 1 μPa^2 s at 1 m [Fig. 5(h)]. Although the shallow-depth source level discrepancies are similar to the discrepancies from the deeper distributions (not plotted), this depth cluster is probably an artifact. The relative fraction of calls associated with shallow depths is highly sensitive to both the type of analysis and the kind of waveguide environment used: a less complex propagation environment (e.g., the Pekeris waveguide) generates relatively more shallow-depth call estimates, as does a run of the full model without making the adiabatic corrections shown in Eqs. (5) and (6). Automated analyses generate relatively more shallow-depth estimates than manual analyses, as well [Figs. 6(b) and 6(e)]. A physical rationale for this artifact is that shallow sources generate high propagation losses with range (Fig. 7), so placing a source at shallow depth is a reliable way of reducing very large discrepancies when no other solution exists. The optimization algorithm could thus be biased toward shallow depths whenever call samples with unresolvable high discrepancies are encountered.

V. CONCLUSION

Multi-season data from an array of directional acoustic recorders have yielded source level estimates for FM "moans" from the fall migration of the Bering-Chukchi-Beaufort bowhead whale population. The source level distribution, which peaks at 162 dB re 1 μPa^2 s at 1 m for the most complex propagation models, is consistent with previous literature (e.g., Cummings and Holliday, 1987) and is robust to the distance of the calls involved, data analysis method implemented (manual or automated), and propagation model used. A power-law propagation model was found to yield good-quality results for the source level distribution, provided that the transmission loss coefficient was optimized by minimizing the mean discrepancies of the call samples. Thus, full propagation modeling of a shallow-water environment may be unnecessary for obtaining source level estimates whenever calls are detected on several relatively widely spaced recorders. Conversely, a full propagation model is still required to obtain depth estimates. Ultimately, bowhead calling depths clustered primarily between 25 and 30 m, which corresponds with optimal depths for propagating 50–100 Hz signals in a 30–50 m depth ocean waveguide.

ACKNOWLEDGMENTS

The authors wish to thank the crew of the *Norseman II*, particularly deck chief Scotty Hameister, for consistently safe and professional deployments and retrievals of the DASARs over the course of the project. Sheyna Wisdom and her colleagues at Fairweather, LLC, provided high-quality logistic support for the fieldwork. Xavier Mouy and Julien Delarue provided helpful suggestions on the source depth estimation, and Bob Norman of Greeneridge, Inc. provided helpful support on DASAR hardware and recording details. Darren Ireland provided assistance with Fig. 1. Finally, none of this work would have existed had Charles R. Greene, Jr. not had the vision to deploy DASARs off the Beaufort coast nearly two decades ago, and to recruit all of the authors to the project.

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.

Blackwell, S. B., Nations, C. S., McDonald, T. L., Greene, C. R., Jr., Thode, A. M., Guerra, M., and Macrander, A. M. (**2013**). "Effects of airgun sounds on bowhead whale calling rates in the Alaskan Beaufort Sea," Mar. Mammal Sci. **29**, E342–E365.

Blackwell, S. B., Nations, C. S., McDonald, T. L., Thode, A. M., Mathias, D., Kim, K. H., Greene, C. R., Jr., and Macrander, A. M. (**2015**). "Effects of airgun sounds on bowhead whale calling rates: Evidence for two behavioral thresholds," PLoS One **10**(6), e0125720.

Bonnel, J., Thode, A. M., and Blackwell, S. B. (**2014**). "Range estimation of bowhead whale (*Balaena mysticetus*) calls in the Arctic using time warping on a single hydrophone," J. Acoust. Soc. Am. **136**, 145–155.

Clark, C. W., and Johnson, J. H. (**1984**). "The sounds of the bowhead whale, *Balaena mysticetus*, during the spring migrations of 1979 and 1980," Can. J. Zool. **62**, 1436–1441.

Cummings, W. C., and Holliday, D. V. (**1987**). "Sounds and source levels from bowhead whales off Pt. Barrow, Alaska," J. Acoust. Soc. Am. **82**, 814–821.

Delarue, J., Mouy, X., Chadwick, J., Fischbach, A., and Hannay, D. (**2015**). "Underwater source level measurements of vocalizing walruses in the Chukchi Sea and implications of detection range and communication space in natural and human-influenced soundscapes," in *21st Biennial Society for Marine Mammalogy Conference on the Biology of Marine Mammals*, San Francisco, CA.

D'Spain, G. L., Murray, J. J., Hodgkiss, W. S., Booth, N. O., and Schey, P. W. (**1999**). "Mirages in shallow water matched field processing," J. Acoust. Soc. Am. **105**, 3245–3265.

Gerstoft, P. (**1997**). "SAGA Users Guide 2.0, an inversion software package," SM-333 (SACLANT Undersea Research Centre, La Spezia, Italy).

Greene, C. R., McLennan, M. W., Norman, R. G., McDonald, T. L., Jakubczak, R. S., and Richardson, W. J. (**2004**). "Directional frequency and recording (DIFAR) sensors in seafloor recorders to locate calling bowhead whales during their fall migration," J. Acoust. Soc. Am. **116**, 799–813.

Lenth, R. V. (**1981**). "Robust measures of location for directional data," Technometrics **23**, 77–81.

Mouy, X., Delarue, J., Martin, B., Hannay, D., Jay, C., and Fishbach, A. (**2014**). "Real-time acoustic monitoring and source level estimates of walruses in the northeastern Chukchi Sea using particle velocity sensors," J. Acoust. Soc. Am. **135**, 2361–2362.

Porter, M. B. (**1991**). "The KRAKEN normal mode program," SM-245 (SACLANT Undersea Research Centre, La Spezia, Italy).

Thode, A., Kim, K. H., Greene, C. R., Jr., and Roth, E. (**2010**). "Long range transmission loss of broadband seismic pulses in the Arctic under ice-free conditions," J. Acoust. Soc. Am. **128**, EL181–EL187.

Thode, A. M., Kim, K. H., Blackwell, S. B., Greene, C. R., Jr., and Macrander, A. M. (**2012**). "Automated detection and localization of bowhead whale sounds in the presence of seismic airgun surveys," J. Acoust. Soc. Am. **131**, 3726–3747.