Supplement 1

A. Precautionary analysis of real time pitch tracks

Numerous pitch tracks resembled right whale upcalls but were not completely convincing, so they were scored as "possibly detected" (n = 79 for the buoy; n = 54 for the glider). The main text of this article presents results acquired using a conservative protocol in which the pitch tracks scored as "possibly detected" were treated as "not detected". This protocol is designed to minimize false detections at the expense of increased missed detections. We opted to focus on this protocol in the main text as it has been extensively employed on all previous deployments in the NW Atlantic (*e.g.*, Baumgartner *et al*. 2019). An alternative approach is to use a precautionary protocol that treats calls scored as "possibly detected" as "detected". This is intended for a science or mitigation application that seeks to minimize missed calls at the expense of false detections.

In our study, employing a precautionary protocol caused the probability of detecting localized calls to increase for both the buoy and the glider across all ranges, but especially at close range (Figures 8, S1.7). For the buoy in average noise conditions (100 dB re 1 $\mu Pa^2\ Hz^{-1}$), the fitted regression suggested that a probability of detection of 0.5 (95% CI: 0.427-0.571) occurred at 5.8 km. For the glider in average noise (100 dB re 1 $\mu Pa^2\ Hz^{-1}$) and depth (15 m) conditions, the fitted regression suggested that a probability of detection of 0.5 (95% CI: 0.400-0.602) occurred at 12.3 km (Figures 8, S1.7).

Pitch tracks were labeled as “possibly detected” for several reasons. The primary cause was poorly formed pitch tracks; these were responsible for 58% and 44% of pitch tracks being scored as “possibly detected” for the buoy and glider, respectively (Table S1.1). For the glider, 35% of “possibly detected” pitch tracks were missed because of humpback whale song, many of which occurred at very close ranges. In contrast, song was only implicated in 15% of the pitch tracks that were labeled as “possibly detected” for the buoy. Approximately 19% of “possibly detected” pitch tracks were caused by human error in the buoy analysis, compared to only 6% in the glider analysis (Table S1.1; Figure S1.6).

B. Analysis of archival audio

The goal of our study was to quantify the range-dependent accuracy of the DMON/LFDCS on board an ocean glider and moored buoy. The DMON/LFDCS provides near real-time acoustic presence estimates by producing an abstraction of the raw acoustic signal in the form of pitch tracks that can then be transmitted back to shore for review. The article outlines a number of analyses to evaluate the performance of the near real-time DMON/LFDCS. Here we repeat those analyses for the archival audio data recorded on the DMON. The difference in performance between the pitch track and archival audio results provides an indication of the cost, or the reduction in system performance, associated with the review of pitch tracks alone (*i.e.*, without the aid of spectrograms or audio) in near real time.

The 541 localized calls were used to determine the probability of detection for each platform based on the manual review of archival audio data. The spatial distribution of localized calls for

this analysis was non-uniform (Figure S1.1). Because the DMON on the moored buoy recorded on a 50% duty cycle, audio was not available for approximately half (301/541) of all localized calls; analysis of the archival audio from the buoy was conducted using the 240 localized calls for which audio was available. The proportion of localized calls that were detected decreased with range (Figure S1.2); 87% of localized calls within 5 km (67/77) were detected while 43% of localized calls between 15 and 40 km (3/7) were detected. Calls were missed for a variety of reasons: 9.1% were missed because they were faint or absent from the spectrogram, 1.7% were missed due to interfering biological sounds (*i.e.*, humpback whale song), 0.8% were missed due to interfering non-biological sounds (*e.g.*, other platform noise, ship noise), and 10% were missed due to human error in reviewing the audio data (Table S1.1; Figure S1.6). For the buoy in average noise conditions (100 dB re 1 $\mu Pa^2\ Hz^{-1}$), the fitted regression suggested that a probability of detection of 0.5 (95% CI: 0.328-0.673) occurred at 13.2 km.

For the glider, the 114 localized calls that were excluded from the near real-time analysis due to platform noise were also excluded from the archival audio analysis. The proportion of the remaining 427 localized calls that were detected decreased with range (Figure S1.2); 79.9% of localized calls within 5 km (147/184) were detected while 37.5% of localized calls between 15 and 40 km (6/16) were detected. Calls were missed for a variety of reasons: 15.2% were missed because they were faint or absent from the spectrogram, 1.4% were missed due to interfering biological sounds (*i.e.*, humpback whale song), 5.4% were missed due to interfering non-biological sounds (*e.g.*, other platform noise, ship noise), and 9.6% were missed due to human error in reviewing the audio data (Table S1.1; Figure S1.6). For the glider in average noise (100

dB re 1 μPa^2 Hz^{-1}) and depth (15 m) conditions, the fitted regression suggested that a probability of detection of 0.5 (95% CI: 0.365-0.635) occurred at 15.8 km (Figure 8A).

Tables

Table S1.1. Diagnostic scores associated with calls labeled as "possibly detected" during manual review of glider and buoy pitch track records of calls localized by the array. Here **n** refers to the number of calls, while **%** is the percentage of total calls labeled as "possibly detected" on each platform (54 for the glider, 79 for the buoy).

Score	**Definition**	**Glider**		**Buoy**	
		n	***%***	***n***	***%***
Poor	Calls were not pitch tracked accurately/completely because of low amplitude or poor shape	24	44.4	46	58.2
Song	Uncertainty due to interfering species calls	19	35.2	12	15.2
Noise	Calls were not pitch tracked accurately/completely because of interfering sound	8	14.8	6	7.6
Missed	Human error (analyst chose wrong score erroneously)	3	5.6	15	19

Table S1.2. Results from manual scoring of glider and buoy archival audio records of calls localized by the array (total number of calls = 541). Here **n** refers to the number of calls, while **%** is the percentage of total localized calls available for detection (*i.e.*, does not consider excluded calls).

Score	**Definition**	**Glider**		**Buoy**	
		n	***%***	***n***	***%***
Absent	Calls were not detected because of low amplitude	52	12.2	3	1.2
Poor	Calls were detected accurately/completely because of low amplitude or poor shape	13	3.0	19	7.9
Song	Uncertainty due to interfering species calls	6	1.4	4	1.7
Noise	Calls were not detected accurately/completely because of interfering sound	23	5.4	2	0.8
Missed	Human error (analyst chose wrong score erroneously)	41	9.6	25	10.4
Detected	Calls were detected by analyst	292	68.4	187	77.9
Exclude	Calls were not available for detection because the platform was not monitoring	114	N/A	301	N/A

Figures

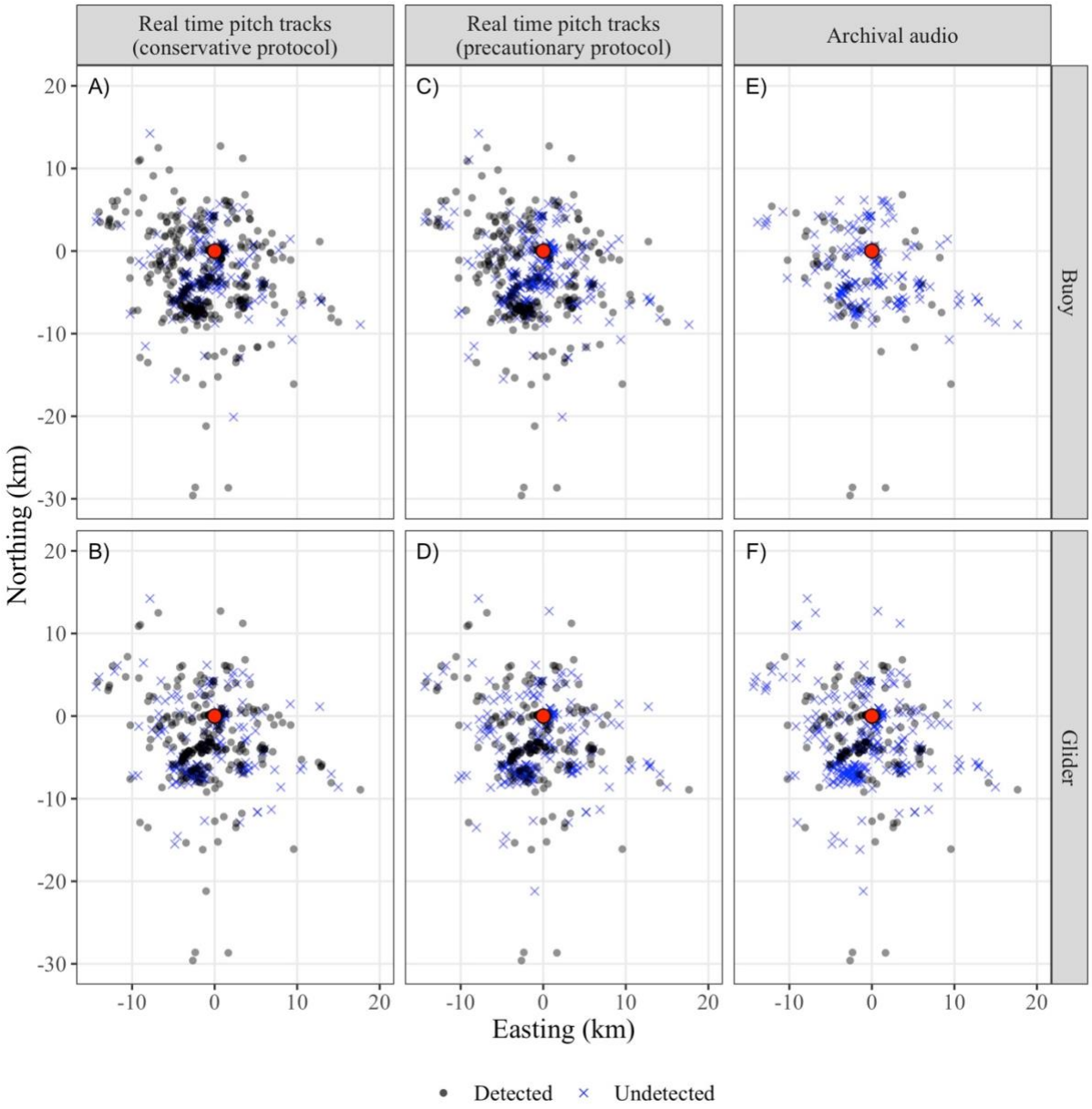

Figure S1.1. The spatial distribution of localized right whale upcalls that were either detected (grey circles) or not detected (blue crosses) by the buoy (top row; n = 541) or the glider (bottom row; n = 426) in each analysis (shown in columns). The red circle at the origin indicates the location of the array. Panels A and B are shown in Figure 5 in the main text.

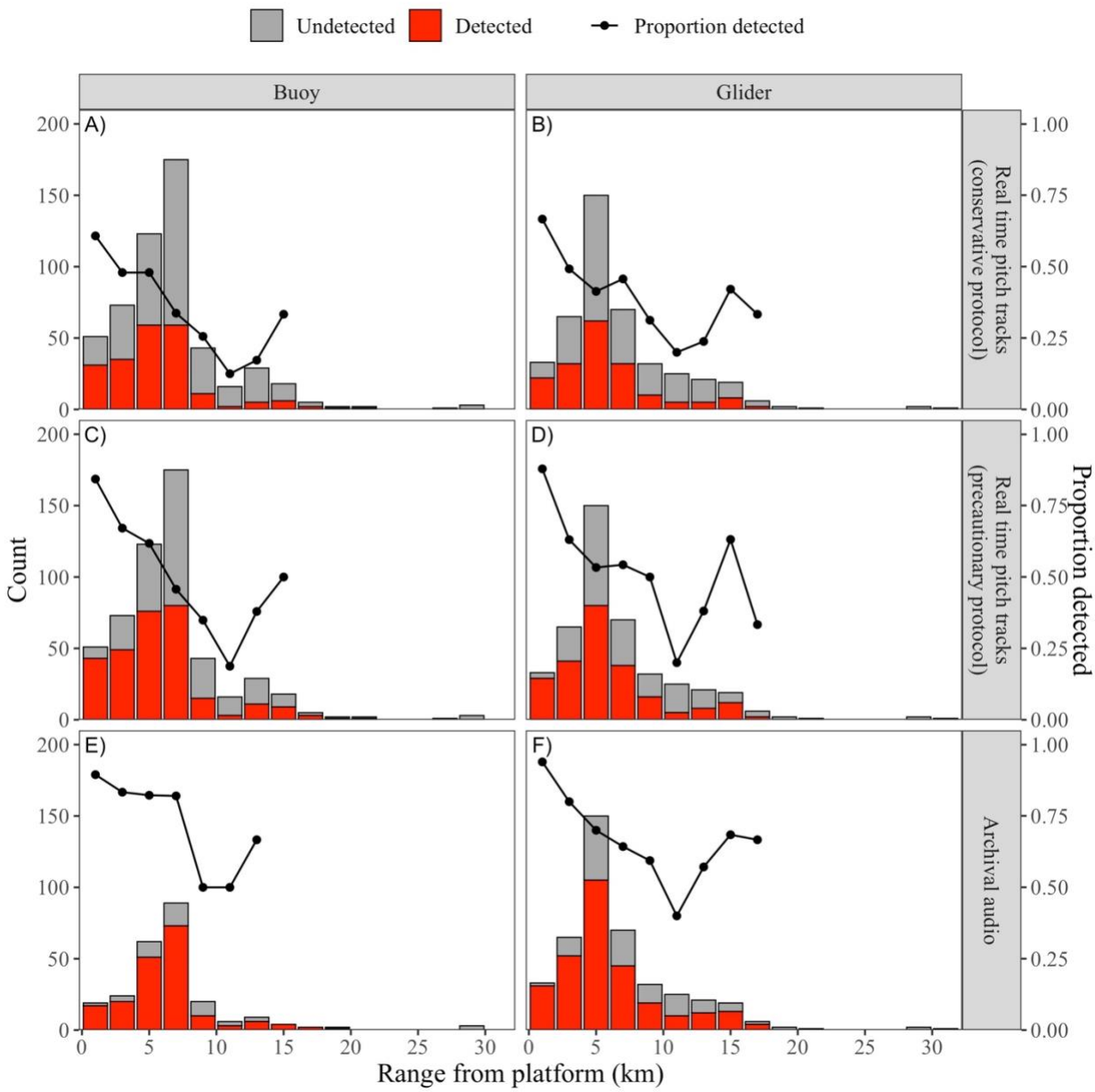

Figure S1.2. Distribution of ranges from each platform (shown in columns) to right whale upcalls localized by the array and detected in a given analysis (shown in rows). Total numbers of localized calls in 2-km bins are shown in gray and localized calls detected in near real-time are shown in red. The black line shows proportions of localized calls detected in 2-km bins with more than 5 calls. Panels A and B are shown in Figure 6 in the main text.

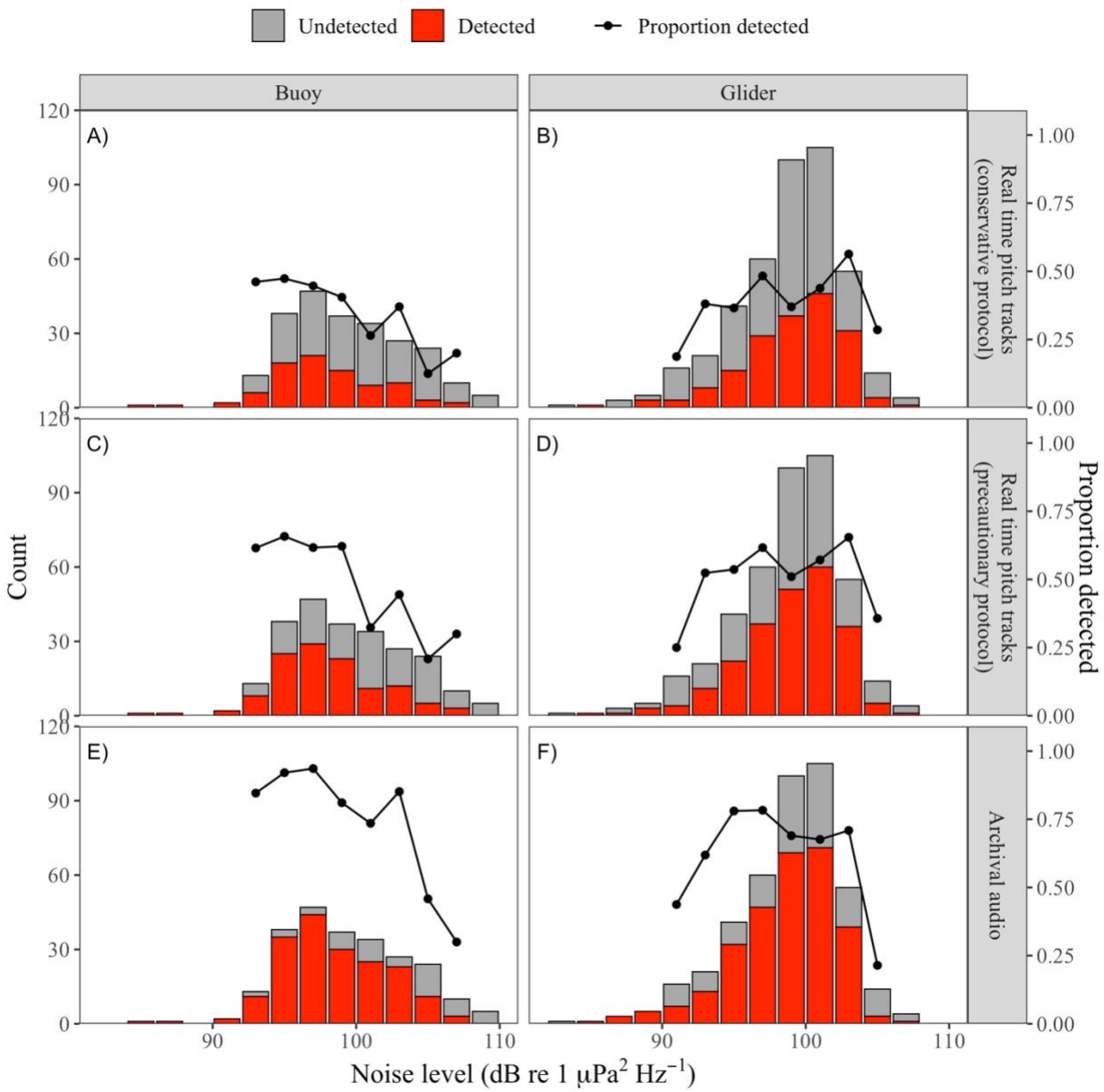

Figure S1.3. Distribution of noise levels for each platform (shown in columns) associated with right whale upcalls localized by the array and detected in a given analysis (shown in rows). Total numbers of localized calls in 2-dB bins are shown in gray and localized calls detected in near real-time are shown in red. The black line shows proportions of localized calls detected in 2-dB bins with more than 5 calls. Panels A and B are shown in Figure 6 in the main text.

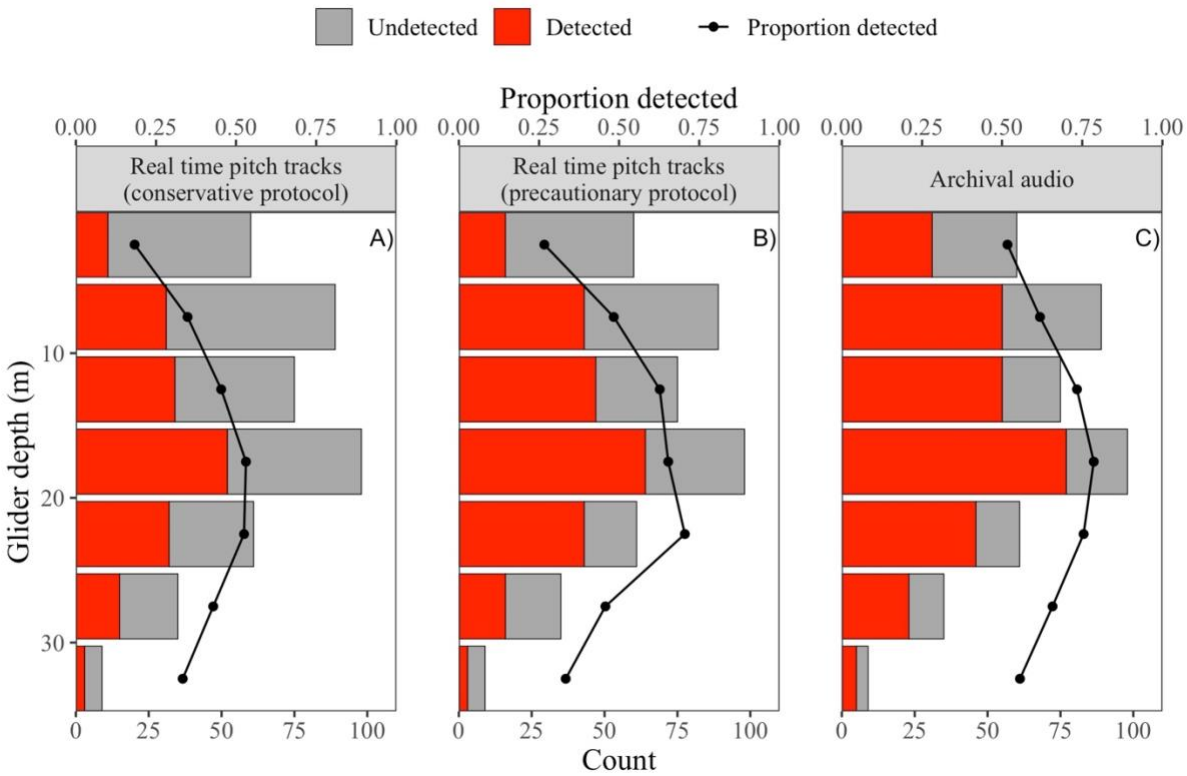

Figure S1.4. Distribution of glider depths associated with right whale upcalls localized by the array and detected in a given analysis (shown in columns). Total numbers of localized calls in 5-m bins are shown in gray and localized calls detected in near real-time are shown in red. The black line shows proportions of localized calls detected in 5-m bins with more than 5 calls. Data in panel A are shown in Figure 6 in the main text.

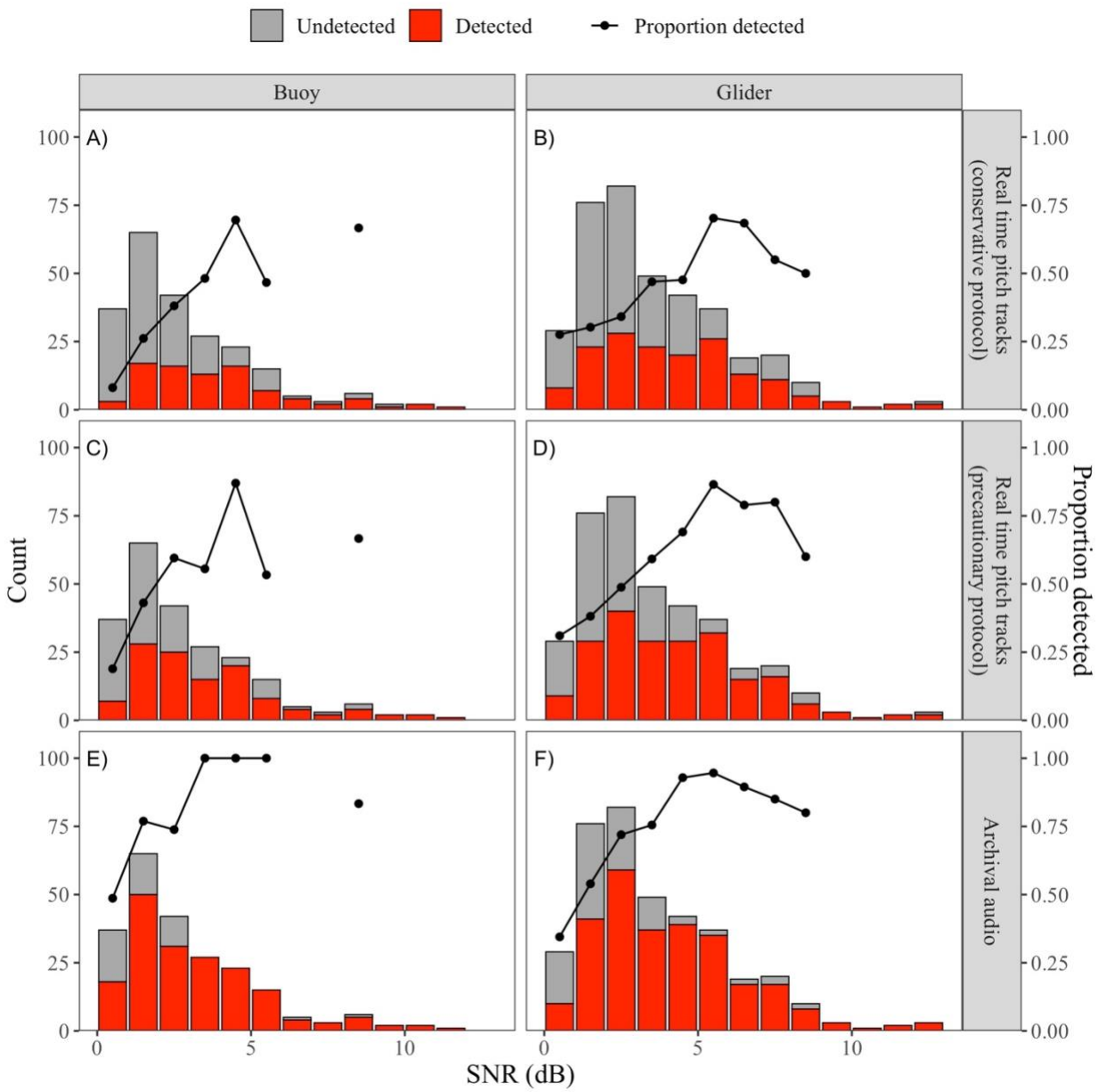

Figure S1.5. Distribution of signal-to-noise ratios (SNR) for each platform (shown in columns) associated with right whale upcalls localized by the array and detected in a given analysis (shown in rows). Total numbers of localized calls in 1-dB bins are shown in gray and localized calls detected in near real-time are shown in red. The black line shows proportions of localized calls detected in 1-dB bins with more than 5 calls. Panels A and B are shown in Figure 6 in the main text.

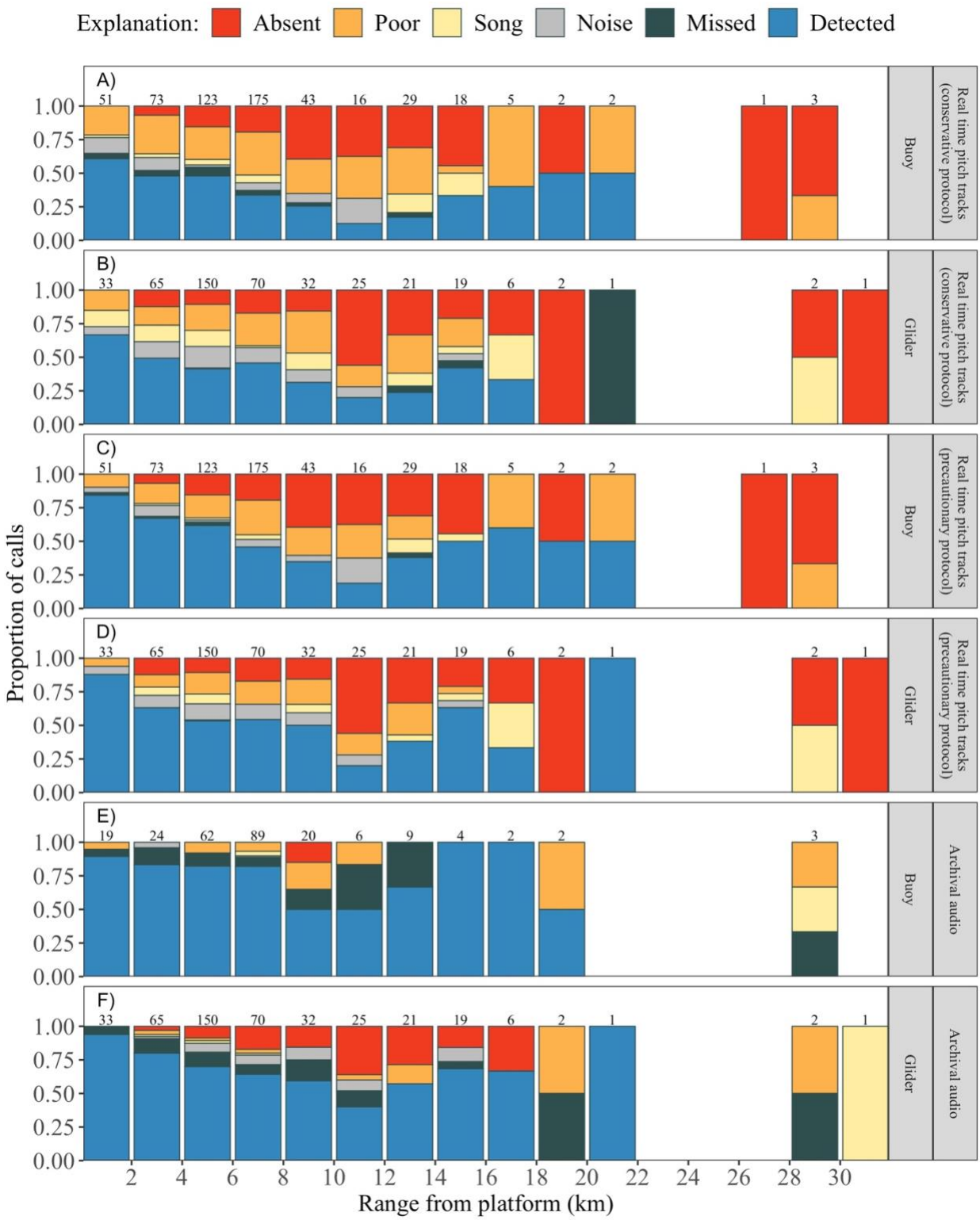

Figure S1.6. Proportion of localized calls assigned to each score category based on analyses of the near real-time pitch track data or archival audio as a function of range from the platform. The

platform and analysis are shown at the right of each plot. Colors indicate the proportion of calls of a given score in 2-km range bins, while the number of calls in each bin is shown above each bar. Definitions of each category are provided in Tables I, S1, and S2. Panels A and B are shown in Figure 7 in the main text.

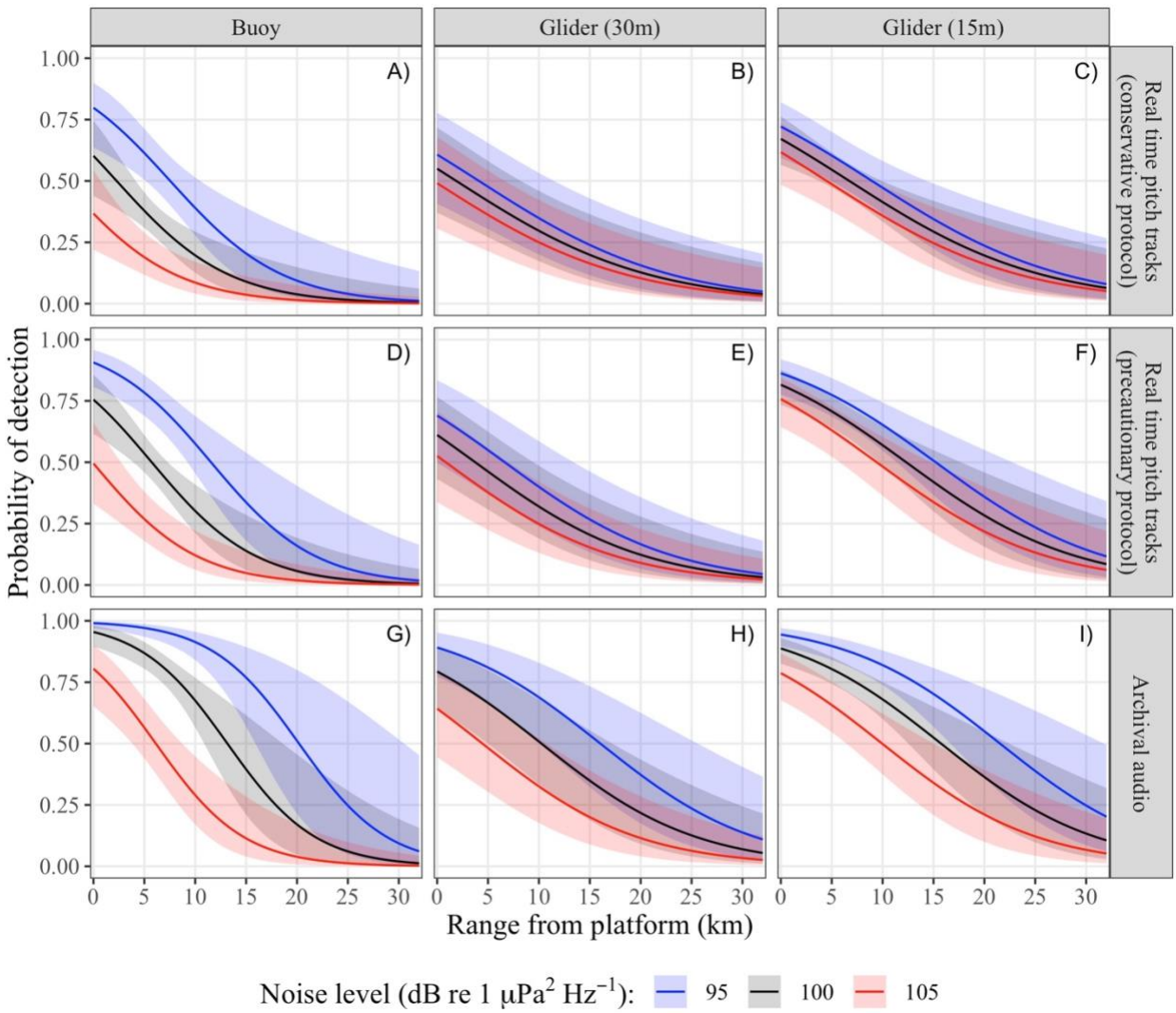

Figure S1.7. Estimated probability of detection of localized right whale upcalls as a function of range to the buoy (left column) and glider at a fixed depth of 30 m (center column) or 15 m (right column) at low (blue lines), medium (black lines), and high (red lines) noise levels based on in a given analysis (rows). The fitted regression models are shown as solid lines, while the 95% confidence intervals are shown as shaded regions. An alternate representation of these data directly comparing each platform is shown in Figure 8 in the main text.

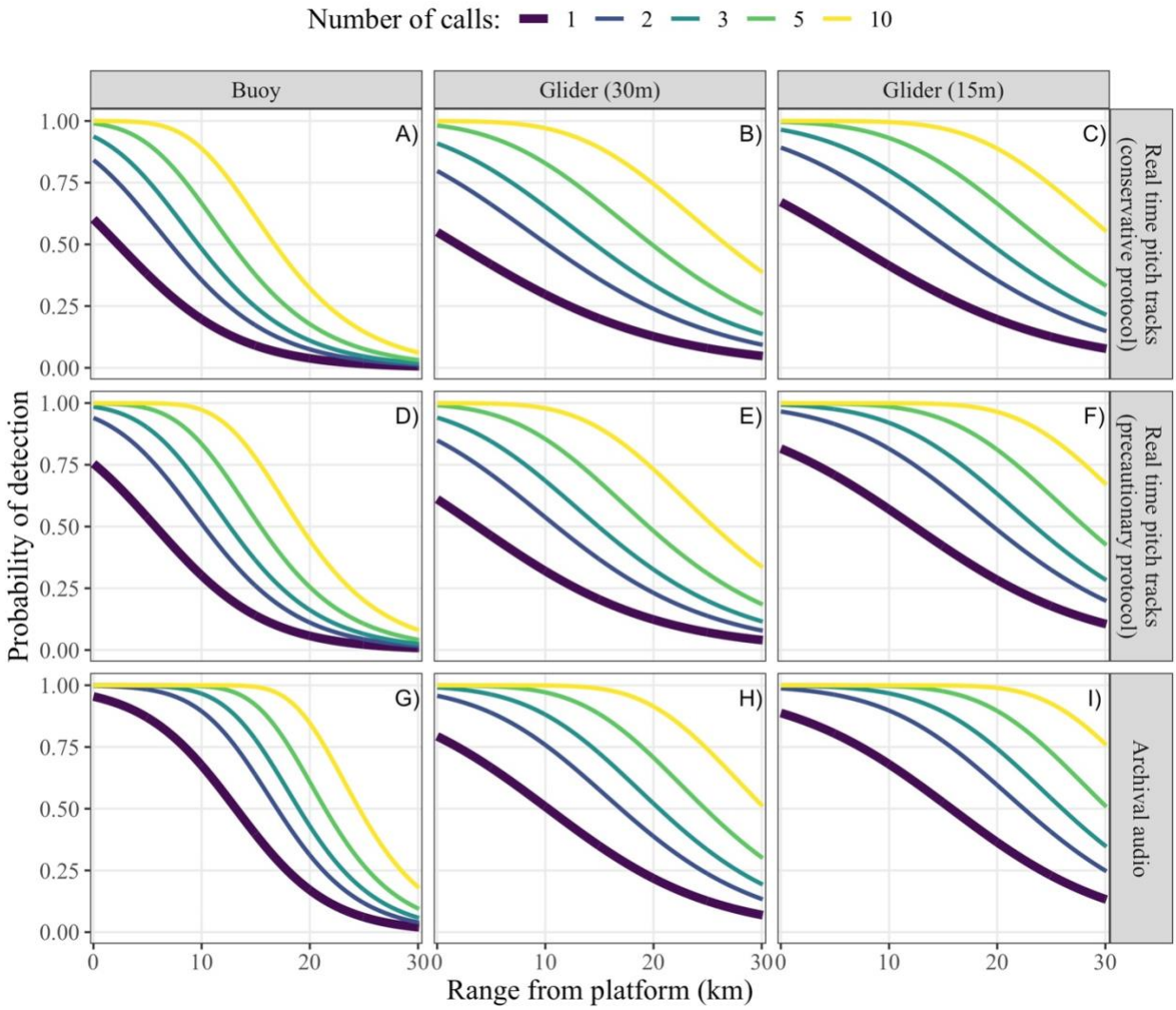

Figure S1.8. Results of a thought experiment showing probability of detecting one of multiple available right whale upcalls as a function of range to range to the buoy (left column) and glider at a fixed depth of 30 m (center column) or 15 m (right column) based on a given analysis (shown in rows) using a fixed noise level of 100 dB re 1 μPa^2 Hz^{-1}. Each colored line shows the probability of detecting one call out of 1, 2, 3, 5, or 10 available calls during some fixed time period. This analysis relies upon the unlikely assumption that calls are detected independently, so the probabilities of detection are likely overestimated (see main text). Panel A is shown in Figure 10 in the main text.

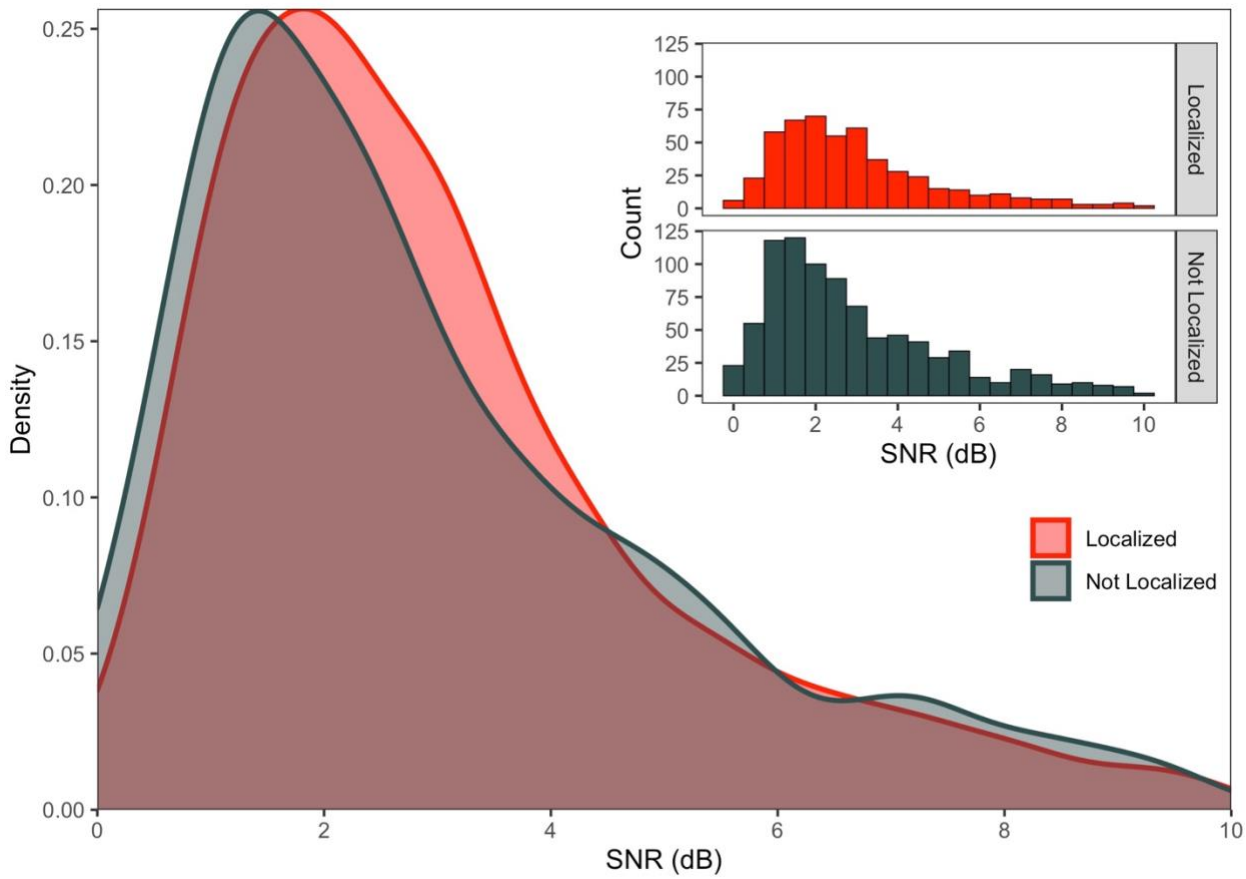

Figure S1.9. The signal-to-noise (SNR) level (in dB) for calls that were detected and not localized (n = 863; black) versus those that were detected and successfully localized (n = 513; red) on the array. SNR was calculated on the array channel in which the call was detected using the method described in the main text. Calls with SNR above 10 dB (n = 101) were assumed contaminated by impulsive noise and not included. The inset shows the histogram of the raw data in 0.25 dB bins, while the main plot shows the smoothed (kernel density) estimates overlayed to facilitate comparison. The median SNR of detected calls was 2.3 dB (IQR: 2.7 dB), compared to a median of 2.6 dB (IQR: 2.3 dB) for localized calls. Results of a Mann-Whitney U test failed to reject the null hypothesis that both distributions are equal (p = 0.112).