Near real-time marine mammal monitoring from gliders: Practical challenges, system development, and management implications

Katie A. Kowarski,[1,a)] Briand J. Gaudet,[1] Arthur J. Cole,[1] Emily E Maxner,[1] Stephen P Turner,[1] S. Bruce Martin,[1,b)] Hansen D. Johnson,[2] and John E. Moloney[1]

[1]*JASCO Applied Sciences (Canada) Ltd., 202-32 Troop Avenue, Dartmouth, Nova Scotia, B3B 1Z1, Canada*

[2]*Department of Oceanography, Dalhousie University, 1355 Oxford Street, Halifax, Nova Scotia, B3H 4R2, Canada*

ABSTRACT:

In 2017, an endangered North Atlantic right whale mortality event in the Gulf of St. Lawrence, Canada, triggered the implementation of dynamic mitigation measures that required real-time information on whale distribution. Underwater glider-based acoustic monitoring offers a possible solution for collecting near real-time information but has many practical challenges including self-noise, energy restrictions, and computing capacity, as well as limited glider-to-shore data transfer bandwidth. This paper describes the development of a near real-time baleen whale acoustic monitoring glider system and its evaluation in the Gulf of St. Lawrence in 2018. Development focused on identifying and prioritizing important acoustic events and on sending contextual information to shore for human validation. The system performance was evaluated post-retrieval, then the trial was simulated using optimized parameters. Trial simulation evaluation revealed that the validated detections of right, fin, and blue whales produced by the system were all correct; the proportion of species occurrence missed varied depending on the timeframe considered. Glider-based near real-time monitoring can be an effective and reliable technique to inform dynamic mitigation strategies for species such as the North Atlantic right whale.

I. INTRODUCTION

Negative interactions between anthropogenic activities and marine mammals in the Gulf of St. Lawrence, Canada, were brought to the forefront in summer 2017 when 12 endangered North Atlantic right whales (*Eubalaena glacialis*) were found dead. Necropsies on six right whales confirmed that two individuals perished from entanglement, while the remaining four were confirmed or suspected to have died due to vessel strikes (Daoust *et al.*, 2017). In addition to the right whale mortalities, 13 other baleen whales (one blue whale *Balaenoptera musculus*, seven minke whales *Balaenoptera acutorostrata*, and five fin whales *Balaenoptera physalus*) were reported dead in the southern Gulf of St. Lawrence to the Marine Animal Response Society reporting hotline in summer 2017. Apart from one highly decomposed specimen, none of these animals could be necropsied to determine cause of death, though four showed signs of entanglement and vessel strikes could not be ruled out (Wimmer, 2020). The North Atlantic right whale mortalities triggered stakeholders and the

Government of Canada to undertake a series of mitigation measures including fisheries closures, enforced static and dynamic vessel slow-down zones, and voluntary vessel slow-down periods to protect North Atlantic right whales (DFO, 2019a; Transport Canada, 2019b). A key source of information required for the mitigation measures to be effective was the whale distribution in the gulf. Specifically, management bodies required real-time data on right whale locations to effectively implement their dynamic mitigation strategies. The present paper proposes a method to address this requirement using passive acoustic monitoring (PAM).

PAM can provide valuable insight into the occurrence and distribution of acoustically active marine mammals efficiently, and at low cost when compared to traditional visual survey methods. PAM also has few limitations in terms of weather, season, and time of day when compared to visual survey techniques, which are most commonly used for identifying marine mammal presence (Mellinger *et al.*, 2007). Analyzing data after retrieving autonomous archival acoustic recorders is an effective method for determining long-term trends in baleen whale distribution (e.g., Širović *et al.*, 2009; van Parijs *et al.*, 2009; Thomisch *et al.*, 2016). In recent years, methods have been developed to transmit information ashore from acoustic systems without the requirement of equipment recovery. The systems include

a)Also at: Department of Biology, Dalhousie University, 1355 Oxford Street, Halifax, Nova Scotia, B3H 4R2, Canada. Electronic mail: katie.kowarski@jasco.com, ORCID: 0000-0002-1325-8321.

b)ORCID: 0000-0002-6681-9129.

cabled observatories that transfer acoustic data to shore via a direct connection (Hannay *et al.*, 2016), surface buoys that transmit messages via cellular or satellite networks (Spaulding *et al.*, 2009; Baumgartner *et al.*, 2019), and oceanic gliders that detect vocalizations during dives and periodically surface to transmit messages via satellite (Baumgartner *et al.*, 2013). Cabled observatories can be costly and are limited to monitoring one area that is sufficiently close to shore. Surface buoys provide persistent information, but they are similarly limited to monitoring one location, are difficult and expensive to deploy in deep water, are prone to strum noise, and they require mooring lines that pose entanglement hazards for marine life. Though gliders have more limited surface time (transmission time) than buoys, they do not require lines, are small and easy to deploy and recover, and have the benefit of monitoring different areas as they move through the water. Another advantage of gliders is that they can be equipped with additional oceanographic sensors that provide high-resolution information about various aspects of the water column.

In the last decade, there has been an emergence of successful reports of marine mammal monitoring using buoyancy-driven profiling autonomous oceanic gliders, both post-retrieval and in near real-time. Near real-time refers to results reported pre-retrieval, often within 24 h. Post-retrieval, the acoustic signals of blue, humpback (*Megaptera novaeangliae*), sperm (*Physeter macrocephalus*), killer (*Orcinus orca*), and sei (*Balaenoptera borealis*) whales as well as dolphins have been observed in acoustic data recorded onboard oceanic gliders (Moore *et al.*, 2007; Baumgartner and Fratantoni, 2008; Klinck *et al.*, 2012; Küsel *et al.*, 2017; Silva *et al.*, 2019). Near real-time detection of marine mammals from a glider was first reported in 2012 when beaked whale clicks were detected from a glider deployed off Hawaii (Klinck *et al.*, 2012). Baumgartner *et al.* (2013) used two gliders to report on the near real-time acoustic occurrence of fin, humpback, sei, and right whales in the Gulf of Maine and off Nova Scotia, Canada. Davis *et al.* (2016) successfully detected humpback, fin, right, and sei whale signals in near real-time. Finally, Baumgartner *et al.* (2014) reported bowhead whale (*Balaena mysticetus*) acoustic occurrence in near real-time. Glider technology was expanded beyond cetaceans in 2014 when bearded seal (*Erignathus barbatus*) and walrus (*Odobenus rosmarus*) vocalizations were detected in near real-time in the Arctic (Baumgartner *et al.*, 2014).

While there has been some success in near real-time monitoring from gliders, there are many practical challenges relative to other PAM methodologies. The first challenge is the limited amount of information a glider can send. When at the surface between dives, communication between a glider and the shore typically occurs over the Iridium satellite constellation. The transmission budget of the glider while at the surface is restricted due to the limitations of Iridium-based communication, and messages can be lost if the satellite connection is dropped (Baumgartner *et al.*, 2013). Acoustic data of any meaningful size to capture

marine mammal vocalizations cannot be transferred; instead, metadata in the form of "pitch tracks" are transferred (Baumgartner *et al.*, 2013), where pitch tracks are lines that follow the fundamental frequency of sounds automatically detected on board the glider. The amount of metadata sent is limited by the length of time the glider is drifting at the surface and the supported transmission rate. During transmission, there is typically no acoustic monitoring due to the hydrophone being exposed to surface-related noise.

Glider self-noise and movement produce sounds that can mask acoustic signals of interest and reduce the time when acoustic monitoring is effective. Sounds associated with flow noise, fin steering, battery movement, the volume piston, the air pump, non-acoustic sensors, and other devices functioning onboard the glider can be problematic (Küsel *et al.*, 2017). Many of these sounds can be reduced by configuring and operating the glider in a way that maximizes quiet periods. To optimize missions for PAM, glider settings and operations need to be chosen carefully to ensure both acoustic data and supplemental oceanographic data can be collected successfully.

Archival PAM practical challenges are exacerbated for near real-time PAM from gliders. Considerations include the amount of acoustic data that can be stored onboard the glider, the extent of the computational capabilities, and how often data is transmitted. These factors influence overall power consumption on a system that must also power other mission sensors and communicate with the controller onshore.

All challenges faced by analysts that determine marine mammal occurrence in acoustic data after equipment recovery are similarly faced by those interpreting near real-time data from gliders. This includes differentiating between species with overlapping acoustic repertoires, differentiating signals of interest from anthropogenic sounds, interpreting faint signals, understanding spatial and temporal contextual information, and developing effective automated systems to support analysis (Wimmer *et al.*, 2010). Analysts interpreting data in near real-time from gliders have the additional challenge of not having access to the acoustic data. Indeed, PAM analysts typically determine marine mammal acoustic occurrence through the aural and visual review of spectrograms, often with the assistance of automated detectors (Wimmer *et al.*, 2010). To date, analysts interpreting near real-time glider data for mammal occurrence have largely been limited to visually inspecting "pitch tracks" (Baumgartner *et al.*, 2013; Baumgartner *et al.*, 2014; Baumgartner *et al.*, 2020). As the aim of near real-time monitoring can be to trigger costly mitigation measures such as vessels avoiding an area or slowing down, it is critical that a positive marine mammal identification is correct and timely. Such accurate and rapid reporting is typically not required of post-retrieval PAM analysis.

This paper presents a system for monitoring marine mammal acoustic occurrence in near real-time using an oceanic glider and describes the approaches developed to

1216   J. Acoust. Soc. Am. **148** (3), September 2020

Kowarski *et al.*

manage the practical challenges of using PAM from gliders to mitigate interactions between vessels and marine mammals. The system was trialed in the Gulf of St. Lawrence, Canada, fall 2018. The strategies implemented and technologies developed combined techniques previously applied effectively to bottom-moored systems (Delarue *et al.*, 2014; Martin *et al.*, 2014; Frouin-Mouy *et al.*, 2017; JASCO Applied Sciences, 2018; Kowarski *et al.*, 2018) with novel approaches to PAM on gliders. We describe the performance of the system for monitoring baleen whales and compare baleen whale occurrence between near real-time analysis and post-retrieval analysis.

II. METHODS

The methodology for the present paper occurred in three phases. Before glider deployment, a near real-time baleen whale acoustic monitoring glider system was developed (see Secs. II A and II B). The second phase was a glider trial in the Gulf of St. Lawrence described in Sec. II C. Finally, post-retrieval, the system was evaluated and optimized before the trial was simulated as described in Sec. II D. The performance results described in Sec. III are based on how the optimized system performed during simulation.

A. PAM and recording system

Acoustic data were collected via an OceanObserver[TM] (JASCO Applied Sciences; Moloney *et al.*, 2018) installed aboard Teledyne Webb Research's generation-3 Slocum glider. The OceanObserver[TM] hardware was fixed within the glider's science bay and was connected to a single hydrophone (GeoSpectrum Technologies Inc. M36-V35–100; nominal sensitivity of -165 dBV/μPa) mounted on the dorsal exterior of the glider (Fig. 1). The OceanObserver ran within a Java virtual machine (VM) that operated within a Linux operating system running on a dual-core Zynq XC7Z020 chip. This environment allowed hardware-independent algorithms and software to run directly on the embedded platform. The Zynq chip was interfaced with a mid-speed 24-bit analog-to-digital converter. A gain of 14 dB was used. The spectral noise floor and maximum received sound pressure level (SPL) of the recording system were limited by the hydrophone to approximately 30 dB re 1 μPa^2/Hz and 165 dB re 1 μPa^2, respectively. The OceanObserver power draw was 2–3 W and recorded continuously at sampling rates of 512 and 16 kHz, though only the 16 kHz data was used in the present research. While the present research focused on vocalizations of baleen whales under 1000 Hz, higher sampling rates were recorded for future odontocete whistle and click analysis. The OceanObserver ran on its own clock that was not synced with the glider GPS during surfacing. Acoustic data were stored on four 512 GB SD cards for post-retrieval analysis. The OceanObserver was integrated with the glider's mission computer, allowing it to send messages to the mission computer that could subsequently be transmitted to shore via the glider's Iridium telemetry hardware.

FIG. 1. (Color online) Teledyne Webb Research's generation-3 Slocum glider with a top-mounted hydrophone (circled in red) before deployment in the Gulf of St. Lawrence.

B. Near real-time detection of baleen whales

The near real-time acoustic detection of baleen whales occurred in three stages: (1) acoustic signals were identified in real-time on the OceanObserver; (2) metadata of these signals were sent ashore and subsequently emailed to human analysts; and (3) human analysts confirmed the occurrence of marine mammal acoustic signals. The time-lag between an acoustic signal being recorded by the OceanObserver and marine mammal acoustic metadata being transmitted to shore and emailed to a human analyst ranged from 15 min to 3 h 15 min, depending on when the signal occurred within the glider's dive cycle, the weather conditions, and the reporting schedule programmed into the glider. Processes were implemented to address the challenges inherent in near real-time PAM within each stage of the detection are described below.

1. Note on terminology

Traditionally, algorithms designed to automatically identify signals of interest are referred to as "automated detectors" or "autodetectors" and they produce "detections." We propose that for the emerging field of near real-time monitoring of marine mammals, such terminology can be misleading to interested stakeholders, particularly if the outputs will be used to implement mitigations such as vessel speed reductions. The language implies that the signal automatically identified is indeed that of a specific whale, dolphin, or seal. In the present system (where "system" refers to a culmination of every step of the process including both automated and manual stages), the step of classifying an acoustic signal to the species level is completed by a human.

J. Acoust. Soc. Am. **148** (3), September 2020

Kowarski *et al.* 1217

Therefore, we use the term "candidate detection" to describe the output of automated detectors at stage 1 of the process, and "validated detection" to describe detections completed manually by a human analyst at the final stage. It is the validated detections that are recommended for use in mitigation decisions.

2. Stage 1: Automated detectors

In real-time, automated detectors were run on the 16 kHz data using PAMlab software (JASCO Applied Sciences) integrated into the OceanObserver. The automated detectors used the same software previously employed to determine marine mammal occurrence in acoustic recordings post-retrieval as described in Delarue *et al.* (2014), Martin *et al.* (2014), Frouin-Mouy *et al.* (2017), and Kowarski *et al.* (2018). The algorithm applies pre-set Fast Fourier Transform (FFT) settings to create a magnitude spectrogram of length "N" seconds; each frequency in the spectrogram is then median-normalized (Table I). A binary spectrogram is then created with a "1" in each time-frequency bin where the median-normalized value exceeds an empirical threshold (in the range of 1.7–5). The bins assigned 1 are joined to neighbouring 1's using a contour-following algorithm. For each contour, the minimum and maximum frequency, duration, sweep-rate (maximum-minimum frequency/duration), and spectral occupancy of each time bin are computed (the frequency bandwidth within each time bin which is represented by 1's). The contours are then classified as specific candidate detection types if they fall within pre-defined bounds (e.g., minimum frequency, maximum frequency, minimum duration, maximum duration, sweep rates, percent occupancy of the spectra). The FFT parameters, spectral candidate detection threshold, and contour parameters were determined by a trained analyst (K.K.) during the tuning of automated detectors for real-time monitoring of baleen whales in the Gulf of St Lawrence (Table I).

Automated detectors were created both for specific vocalization-types known to be produced by baleen whales of interest and to capture more general acoustic signals

expected to occur in the acoustic data, including those produced by the glider itself (Moore *et al.*, 2007; Küsel *et al.*, 2017). The two types of detectors are referred to as vocalization-specific automated detectors and general automated detectors, respectively. The vocalization-specific automated detectors combined with the general automated detectors provided the context necessary to determine the occurrence of marine mammals in the acoustic data.

Vocalization-specific automated detectors were developed to target North Atlantic right whale upcalls (Mohammad and McHugh, 2011), blue whale infrasonic and audible moans (Marotte and Moors-Murphy, 2015) and fin whale 20 Hz pulses (Delarue *et al.*, 2009; Table II). Emphasis was placed on these species due to their at-risk status (right and blue whale), prevalence in the region, and well-described and relatively species-unique vocalizations. Due to the critical state of the North Atlantic right whale population, and the importance of reliably identifying their vocalizations in the data, three upcall automated detectors with varying levels of performance were developed. The right whale 1 automated detector was developed to identify very high quality, clear signals, and be consistently correct, but allow a high number of false negatives (FNs) (missed candidate detections). The right whale 2 automated detector was developed to identify medium and high quality upcalls but could be falsely triggered by some humpback songs and glider noise. The right whale 3 automated detector was developed to identify poor, medium, and high quality upcalls, but allowed frequent false positives (FPs).

In addition to six vocalization-specific automated detectors (Table II), five general automated detectors were developed to provide contextual information on all acoustic signals expected in the data. The general automated detectors captured the acoustic signals of all remaining baleen whale species in the Gulf of St. Lawrence region whose vocalizations either overlap with those of other oceanic sounds or are too variable in nature to create an effective vocalization-specific automated detector. This includes minke, sei, and humpback whale vocalizations (Table II). Additionally, general automated detectors captured right,

TABLE I. Summary of parameters applied to automated detectors used during the post-retrieval simulation. All automated detectors used a hanning window.

Automated detector	Frequency (Hz)	Duration (s)	Sweep rate (Hz/s)	Peak intensity (Hz)	Bandwidth (Hz)	Frequency resolution (Hz)	Time window (s)	Time step (s)	Advance (s)	Threshold
Right whale 1	65–260	0.60–1.2	30–290	NA	70–195	4.000	0.128	0.032	8	2.5
Right whale 2	65–260	0.50–1.2	30–290	NA	NA	4.000	0.128	0.032	8	3.0
Right whale 3	30–400	0.50–10	10–500	NA	25–NA	4.000	0.170	0.025	10	3.0
Fin whale	10–40	0.40–3.0	−100–0	20–22	6–NA	1.000	0.200	0.050	5	1.7
Blue whale infrasonic	14–22	8.0–30	−500–0	16.5–17.5	1–5	0.125	2.000	0.500	40	4.0
Blue whale audible	40–100	2.0–10	NA	NA	15–50	2.000	0.250	0.050	10	3.0
General 1	10–100	0.30–10	NA	NA	10–NA	2.000	0.200	0.050	15	4.0
General 2	30–400	0.08–0.60	NA	NA	25–NA	4.000	0.170	0.025	10	3.0
General 3	40–300	0.50–10	NA	NA	15–50	2.000	0.250	0.050	10	3.0
General 4	100–700	0.50–8.0	NA	NA	50–200	4.000	0.200	0.050	5	3.0
General 5	300–2500	0.50–8.0	NA	NA	150–300	8.000	0.125	0.050	5	3.0

1218   J. Acoust. Soc. Am. **148** (3), September 2020

Kowarski *et al.*

TABLE II. Baleen whale vocalizations expected in the Gulf of St. Lawrence including the vocalization type, the producing sex, other acoustic signals that overlap in characteristics with the vocalization (conflicting signals), and the automated detectors designed to capture the vocalizations. "General" refers to a group of five automated detectors that captured all signals and were not designed for specific vocalizations. The same general automated detector was capable of capturing multiple vocalization-types.

Species	Vocalization-type	Sex	Conflicting signals	Automated detector(s)	Reference(s)
Right whale	Upcall	Both	Humpback whale and glider sounds	Right whale 1, Right whale 2, Right whale 3, General 2 and 3	Mohammad and McHugh (2011); Parks *et al.* (2011); Baumgartner *et al.* (2019)
	Gunshot	Male	Seismic and glider sounds	General 2, 4, and 5	
	Other	Both	Humpback whale and glider sounds	General 2, 3, 4, and 5	
Blue whale	Infrasonic	Male	Vessel sounds	Blue whale infrasonic	(Berchok *et al.*, 2006); Clark and Altman (2006); Marotte and Moors-Murphy (2015)
	Audible	Both	Fin, sei, and humpback whale and glider sounds	Blue whale audible, General 1 and 3	
Fin whale	20 Hz pulse	Male	Glider sounds	Fin whale, General 1	Delarue (2008); Delarue *et al.* (2009)
	Audible	Both	Blue, sei, and humpback whale and glider sounds	General 3	
Minke whale	Pulse train	Unclear	Glider sounds	General 2	Risch *et al.* (2013)
Sei whale	Downsweep	Unclear	Blue, fin, and humpback whale and glider sounds	General 3	Tremblay *et al.* (2019)
Humpback whale	Song	Male	Right, minke, blue, sei, fin whale, and glider sounds	General 3, 4, and 5	Kowarski *et al.* (2019)
	Non-song	Both	Right and minke whale and glider sounds	General 2, 3, 4, and 5	

blue, and fin whale vocalizations, including those not identified by any vocalization-specific automated detector. General automated detectors also triggered on glider noises, further providing the context critical to the human detection stage. While there were only five general automated detectors, the same general automated detector was capable of capturing multiple vocalization-types (Table II). In total, 11 automatic contour detectors were developed.

Vocalization-specific automated detectors were developed and optimized using a subset of training data. Ideally, training would have used acoustic data containing the marine mammal vocalizations of interest that were collected on a Slocum glider, but no such recordings were available at the time of automated detector development. Available training data were collected from the western North Atlantic Ocean using Autonomous Marine Acoustic Recorders (JASCO Applied Sciences) that were moored at or near the seafloor for up to 1 year (see Delarue *et al.*, 2018; Kowarski *et al.*, 2019). All training data had previously been analyzed for marine mammal acoustic occurrence and provided a plethora of acoustic files (each 10.5–11.2 min in duration) for automated detector development. For each species/vocalization of interest, data from 20 to 40 acoustic files were used. Approximately one-quarter of the files contained high quality signals of interest (high signal-to-noise ratio, SNR, and no competing signals), one-quarter contained low quality signals of interest (low SNR, no competing signals), one-quarter contained the signal of interest and competing signals, and one-quarter contained only competing signals. Signals were qualitatively considered high or low SNR based on how clearly they could be visually and aurally observed in the spectrogram relative to other signals. Additionally, six 30 min acoustic files recorded on an

oceanic glider that contained sounds associated with operating and moving the glider were used to learn how the automated detectors reacted to these sounds.

The selected training files were used to determine the optimum automated detector parameters for near real-time monitoring of each vocalization type. Parameters that were optimized included FFT settings, time-frequency restrictions, and the amplitude of the signal compared to the median sound level. These parameters were optimized for real-time monitoring by reducing the FPs, which unavoidably caused the automated detectors to have higher FNs, or missed detections. For each vocalization of interest, the selected acoustic files were opened in PAMlab, the automated detector was run, and then the automated detector was edited within PAMlab until the desired results were obtained (high TP and low FP). Once the automated detectors performed well (had low FP) on the training files, they were evaluated on three large acoustic data sets (3–12 months of recordings independent of training files) collected off Nova Scotia, Canada, and the automated detector performance was checked in terms of precision (P) and recall (R) when evaluated against the known presence of vocalizations in the larger data set. This was an iterative process that continued until each vocalization-specific automated detector performed satisfactorily (e.g., right whale 1 automated detector had a P of 1.00). Here P and R are defined as

$$P = \frac{TP}{TP + FP}; \quad R = \frac{TP}{TP + FN}, \quad (1)$$

where P is the proportion of TPs correctly identified and R is the proportion of TPs identified out of the actual number of acoustic signals in the data.

3. Stage 2: Candidate detection prioritization and data transfer

The limited Iridium bandwidth greatly restricted the number of candidate detections that could be sent to shore when the glider surfaced; therefore, a method to strategically decide which to send was developed. Each automated detection was assigned a priority ranking: right whale upcall candidate detections had the highest priority, followed by blue whale candidate detections and fin whale candidate detections. The general automated detections were all given the lowest priority. The OceanObserver software accumulated candidate detections until either it had collected 2000 candidate detections, or an hour had passed since candidate detections were last transferred to the glider's communication computer. The accumulated candidate detections were then scanned using a sliding 5 min window to locate the group (or ensemble) of candidate detections with the highest-ranked score. Each ensemble had a four-number score, which was the number of candidate detections at each priority level. Ensembles were ranked by comparing candidate detection counts in priority order (highest to lowest). For example, if ensemble E1 had more high priority candidate detections than ensemble E2, then E1 received a higher score than E2. If the highest priority candidate detection counts were equal, then the same comparison was performed for the second highest priority level, and so on. The highest-ranked ensemble in the entire buffer was then sent to the glider mission computer. Candidate detections within the ensembles were sent in the form of up to seven points that best followed the centroid of the time-frequency pitch track associated with the candidate detection. Where transmission to the glider computer was triggered by the passing of one hour, PAMlab continued selecting and sending ensembles until the maximum hourly transmission budget (8 KB; ex. 24 KB in 3 h dive) was exhausted. Where transmission was triggered by the buffer containing 2000 candidate detections, ensembles ceased being sent once the candidate detection buffer size dropped below 2000 candidate detections or the hourly transmission budget was reached.

Ensembles added to the glider's mission computer throughout the dive were stored along with their ranking. When the glider surfaced, the ensembles were sent in rank order to shore via the glider's Iridium telemetry system. Ensemble data were received by Teledyne Webb's server and accessed by JASCO over the internet. The data were parsed and archived in a relational database. Each ensemble was then automatically distributed by email to human analysts. Emails included figures that presented the pitch track of every candidate detection that occurred within the ensemble, plotted across frequency and time. Figures had a frequency display bandwidth of from 0 to 1000 Hz and a 5 min duration to match the entire duration of the ensemble. Additionally, the emails included consecutive 30 s "zoomed-in" sections of the 5 min ensemble.

4. Stage 3: Manual validation

The final stage in determining the acoustic occurrence of baleen whales in near real-time was the human manual analysis. Experienced analysts received the glider emails and used the information within the figures to validate whether vocalizations of baleen whale species were present, absent, or possibly present within each email (where each email included one 5 min ensemble and multiple emails were sent per dive). Emails were monitored from approximately 8 am to 8 pm, seven days a week, for the duration of the trial. Each email was reviewed by two analysts, ensuring that someone was always available to deliver as near to real-time service as possible.

During automated detector development, a decision protocol to guide the manual validation decision process was created.[1] The protocol was comprised of multiple decision trees for each baleen whale species with a cascade of yes/no questions that resulted in the final decision by the analyst. The protocol sought to encompass all contextual aspects typically applied during manual analysis of recorded acoustic data including whether the species had been detected recently, the number of pitch tracks, the shape of pitch tracks, and any pattern or repetition of pitch tracks. The protocol was designed to be extremely conservative, with the goal of avoiding all false positive detections. To successfully interpret the protocol instructions, analysts using the protocol were expected to have experience analyzing acoustic data, be familiar with baleen whale vocalizations, and be familiar with how these acoustic signals look in pitch track form. The two analysts in the present study gained this experience during automated detector and protocol development. Using the protocol, each email was reviewed and categorized for each species as being acoustically present (definite validated detection), absent (no validated detection), or possibly present (possible validated detection). A definite validated detection could not be made unless both analysts categorized an email as such. On the rare occasion where analysts differed, the more conservative outcome was considered correct.

Validated detection results were stored in a database that could be readily shared with online public resources such as Dalhousie University's WhaleMap (Johnson, 2018) for distribution to interested stakeholders.

C. Gulf of St. Lawrence trial

One Slocum glider was deployed by Dalhousie University's Ocean Tracking Network near the Orpheline Trough in the Gulf of St. Lawrence on 15 September 2018 as part of a larger program monitoring the habitat use of the Gulf by North Atlantic right whales (DFO, 2019b). The Orpheline Trough is an area known to be frequented by right whales in the summer months (Johnson, 2018). The glider monitored the region, relaying messages to shore, until it was retrieved on 30 October 2018. Though the glider monitored for 45 days, the non-volatile storage onboard the OceanObserver was filled after 16 days; these data were
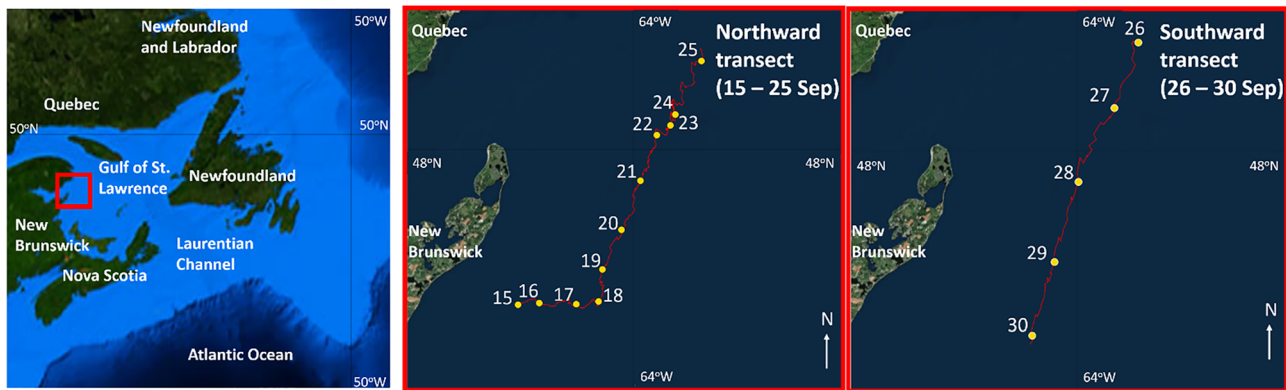
FIG. 2. (Color online) Slocum glider northward (middle) and southward (right) transects in the Gulf of St. Lawrence, Canada from 15 to 30 September, 2018, where the location at the start of each day is indicated by a labelled yellow dot.

analyzed for this paper (15–30 September 2018).[2] During the analyzed period, the glider transited to the Orpheline Trough (15–16 September), followed a northward transect against the predominant current (16–25 September), and then a southward transect moving with the current (26–30 September; Fig. 2).

D. Post-retrieval system evaluation, optimization, and performance analysis

Once the glider trial concluded, the near real-time results were compared to the 16 days of continuously recorded 16 kHz audio data to evaluate the system performance. Methods to optimize system performance were developed. Finally, the entire trial was simulated using the optimized system configuration and the performance of the final version was measured.

Evaluation began with a detailed manual review of all acoustic recordings using PAMlab. A single experienced acoustic analyst reviewed every file for the occurrence of marine mammal vocalizations. Files were opened in PAMlab with the following FFT settings: a 2 Hz frequency resolution, 0.128 s time window, 0.032 s time step, and Hamming window. Data were reviewed from 0 to 1000 Hz, in 30 s windows which corresponded to the view sent in the emails in near real-time. Data were visually and aurally reviewed, and every marine mammal vocalization was annotated to the vocalization-type level. Where the analyst was uncertain in assignment of an acoustic signal, the signal was annotated as possibly being produced by a suspected species. Annotations were made conservatively: if there was any doubt as to the source of a signal, it was considered "possible." To investigate the occurrence of blue whale infrasonic moans, files were re-analyzed using spectrogram parameters that allowed for easier visualization of such long, tonal signals (0.4 Hz frequency resolution, 2 s time window, 0.5 s time step, Hamming window, from 0 to 100 Hz, 5 min at a time), which were similar FFT settings employed by the near real-time system. The annotations created during manual review were considered truth data.

Throughout the manual review, the analyst identified areas where the system did not perform as expected. This was accomplished by viewing the spectrograms as automated detector pitch tracks (a view option of PAMlab) and comparing them to the ensembles sent via email throughout the trial. Where emails or validated detections were different than would be expected based on the truth data, the cause was investigated. Weaknesses, and, in some instances, software bugs, were identified in the candidate detection prioritization algorithm, the ensemble creation software, the email protocol, and the automated detectors.

There were three significant improvements made during system optimization. The first was the use of automated detector contours (drawn using 30–50 points) rather than automated detector pitch tracks (up to seven points). Contours trace the outline of the energy of a candidate detection and more accurately capture the spectral shape of both tonal and broadband signals than is possible using pitch tacks. The second improvement was introducing flexibility in ensemble duration that could be created in 30 s, 1, 2, 5, or 10 min durations rather than being restricted to 5 min. This was accomplished via a sliding 10 min window that located the highest score ensemble. When the ensemble was too big to be sent to shore, it was reduced in time around the highest score portion of the 10 min ensemble until the size of the shortened ensemble was such that it could be sent to shore.

The third improvement was minor edits to automated detectors and the email protocol to better manage glider-related sounds, a process that could not be completed optimally pre-deployment as we did not previously have access to recordings where both glider sounds and marine mammal vocalizations simultaneously occurred. Two acoustic files that encompassed the duration of two dive cycles or 4.75 h (on 17 and 24 September) were used to optimize the automated detectors and the protocol. Automated detector parameter edits were only deemed necessary for the right whale 3 detector where a sweep rate parameter was added to avoid capturing glider noise. Final parameters for automated detectors are included in Table I. The email protocol was updated to be more specific in its instructions to avoid inter-analyst variability and more restrictive in its assignment of definite validated detections to minimize the misclassification of glider noise as marine mammal vocalizations.

J. Acoust. Soc. Am. **148** (3), September 2020

Kowarski *et al.*   1221

The protocol was also altered to allow the manual analyst to assign a possible validated detection rather than the protocol's suggested negative validated detection if the analyst deemed it appropriate (i.e., the analyst's experience is such that some information not captured in the current protocol iteration leads them to believe a species may be present).

The acoustic data were processed and re-analyzed using the previously stated improvements. The 16 days of acoustic data were fed into the system as if the glider was indeed on mission. Considerable efforts were made to accurately simulate the glider mission. The simulated trial was run on a workstation using the same Java VM configuration as on the glider. The updated automated detector configuration and glider processing software was tested on the OceanObserver hardware to confirm that the central processing unit (CPU) demands did not exceed its capabilities for real-time execution. No significant change in CPU load was observed, so power consumption was expected to remain the same. The simulation was run using glider system components and configured exactly as when running on a glider in terms of acoustic signal processing, automated database import of glider messages (with a maximum of 8 kB per hour), and automated generation of notification emails. The only system component which differed from an actual glider mission was that all glider messages were delivered directly to the database, rather than being sent via Iridium, which may drop some messages. Emails were received and analyzed by an experienced analyst who had not previously viewed the recorded data or the findings of the detailed manual review.

The performance of the optimized system was calculated and presented as P and R for both candidate detections and validated detections. Data used to edit automated detectors and the analyst protocol before conducting the simulation were excluded from the calculations of optimized system performance. To understand the reliability of the system in different contexts, the human validated detector performance was calculated and presented by email, hour, glider dive, and day. Emails, where possible validated detections occurred, were considered negative validated detections in calculating human detector performance. To understand how performance was impacted by signal SNR, the by email performance metrics were calculated for all emails as well as separated into emails considered low and high SNR. SNR was calculated from the truth annotations as vocalization SPL minus ambient SPL computed over the same duration as the vocalization. Vocalization SPL was calculated from the middle 90% of energy in the annotation. Ambient SPL was challenging to compute due to the regular occurrence of loud glider sounds that could skew the SNR results and misrepresent vocalization SNRs as negative. To minimize the chance of including glider noise in the ambient levels, SPLs were calculated for three periods before and three periods after the annotation offset by $1\times$, $2\times$, and $3\times$ the annotation duration. The SPL of each of the six periods was calculated (using the same frequency range and duration of the annotation). The period with the lowest SPL was used as the ambient SPL for that vocalization. The average SNR per vocalization-type was calculated for each email and the entire

truth data set. For every vocalization-type in an email, the email was labelled either low or high SNR. Emails were considered low SNR when the average SNR of the vocalizations in the email were lower than the average SNR of that vocalization-type for all truth data. Emails were considered high SNR when the average email SNR was greater than or equal to the average truth data SNR for that vocalization-type. Describing how vocalization SNR influenced performance metrics when investigating timeframes greater than one email (e.g., by hour, glider dive, or day) was inappropriate as such extended timeframes could not reliably be classified as containing "low" or "high" SNR signals as varying acoustic conditions and would be encountered.

III. RESULTS

A. Glider mission

The glider monitored the Orpheline Trough in the Gulf of St. Lawrence from 15 September to 30 October 2018 for a total monitoring period of 45 days. The daily energy demands of the glider (including the OceanObserver and a CTD) averaged 8 coulomb amp-h. During the period with acoustic data recordings (15 to 30 September 2018), the glider undertook 116 dives that averaged a duration of 2.8 h (of which approximately 7.8 min was surface time), a distance travelled of 1783 m, and a dive depth of 60 m. Over the 15 recording days, the OceanObserver stored 1.82 TB of continuously recorded acoustic data.

B. Truth data

The manual analysis of all 16 kHz data post-retrieval revealed the acoustic occurrence of right, fin, minke, and blue whales (Figs. 3 and 4). Only ten dives did not contain marine mammal vocalizations. Most of these dives occurred when the glider was in transit on 15 and 16 September. Acoustic signals resembling those of grey seals and sei whales were observed, but their occurrence was never considered definite. In addition to baleen whales, acoustic signals of delphinids were also observed but were not investigated as part of the present research. The SNR of acoustic signals ranged greatly from as low as -12 dB for minke whale pulse trains to over 50 dB for right whale gunshots and fin whale 20 Hz pulses with an average across vocalization types from 6 to 16 dB (Fig. 5). The SNR of minke whale pulse trains was skewed low due to the challenge of calculating ambient SNR for such long, sometimes broadband, vocalizations where entire pulse trains were annotated, not individual pulses. Right whale upcall SNRs were at times negative, indicating the SNR calculation algorithm could not successfully find a period before or after the annotation that did not contain signals louder than the upcall in question.

Except for transiting days, North Atlantic right whales were present (definite) or thought to be present (possible) on every recording day based on the occurrence of upcalls, gunshots, or both. Many baleen whale moans overlapped in characteristics with both right and humpback whales. These moans were commonly associated with right whale gunshots and upcalls, and were, therefore, likely produced by right

1222   J. Acoust. Soc. Am. **148** (3), September 2020

Kowarski *et al.*

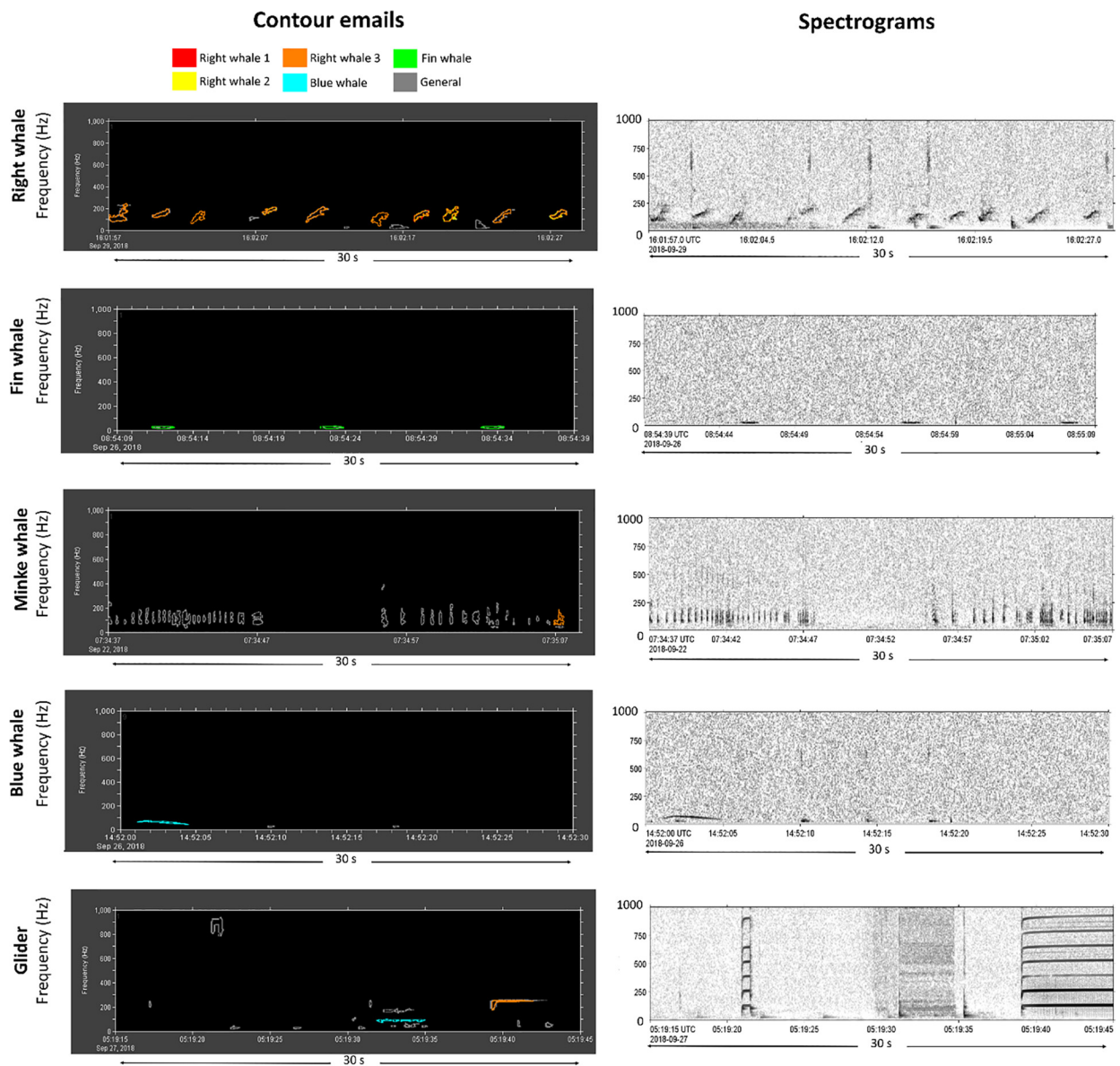**Contour emails**   **Spectrograms**

FIG. 3. (Color online) Contour emails (left) sent during a simulation of a Gulf of St. Lawrence glider mission from 15 to 30 September 2018 and the associated spectrogram (right; 2 Hz frequency resolution, 0.128 s time window, 0.032 s time step, Hamming window) created post retrieval for blue, fin, and minke whale vocalizations as well as glider noise. Contours are colored according to the automated detector that created them.

whales. However, it was impossible to be certain that humpback whales were not also present. This ambiguity was captured as possible right whale moans (Fig. 4). Acoustic signals of fin whales (20 Hz pulses) and minke whales (pulse trains) were prevalent throughout the recordings, confirmed on 12 and 14 out of the 16 days, respectively (Fig. 4). Blue whale vocalizations were rarer, with audible vocalizations confirmed on only two dives on 25 and 26 September and no infrasonic moans observed (Fig. 4).

C. Simulated near real-time system performance

During the simulated glider mission, 651 contour emails were created, representing 14.9 h of contour data. The vocalizations of right, fin, minke, and blue whales were

accurately represented within the emails (Fig. 3). Unsurprisingly, with an average of only 56 min of contour data delivered for each recorded day (averaging 41 ensembles per day), the near real-time occurrence results were limited when compared to truth data where 24 h of acoustic data was reviewed each day, with a bias towards right whales that were given highest priority (Fig. 4). The system produced definite validated detections of right, fin, and blue whales on eight, five, and one day(s), respectively (Fig. 4). Possible validated detections were created for right, fin, and minke whales. When compared with truth acoustic data, approximately 50% of possible validated detections were found to be accurate (12 of 24 possible validated right whale detections were true, and 8 of 14 possible validated fin whale detections were true). The protocol was such that a
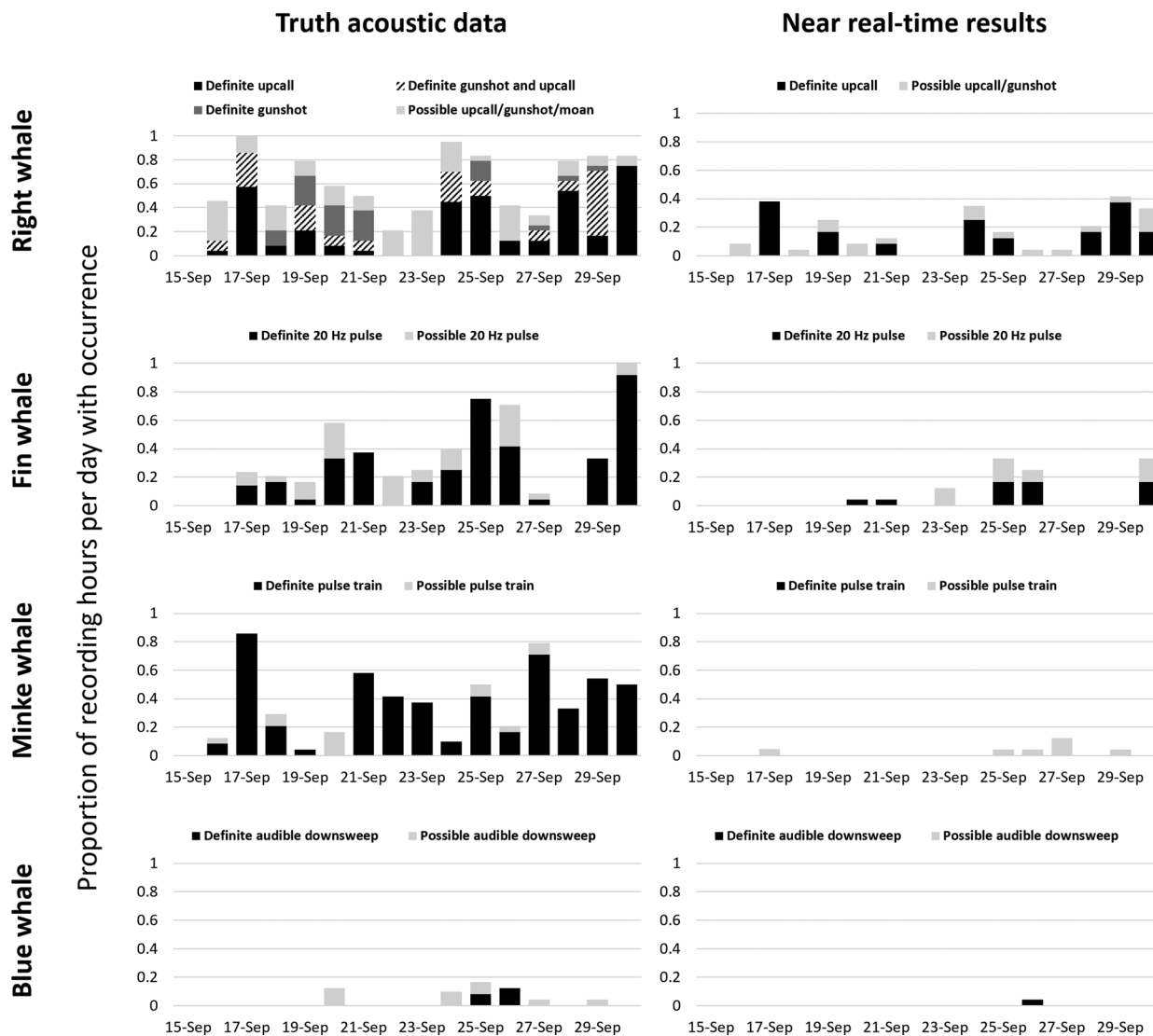
FIG. 4. The proportion of recording hours per day (of a glider that continuously monitored the Gulf of St. Lawrence from 15 to 30 September 2018) that contained the vocalizations of different marine mammals as determined from manual review of data post-retrieval (truth acoustic data) and in a simulated near real-time glider mission (near real-time results). Recording hours from acoustic files used to edit automated detectors pre simulation were excluded.

definite validated minke whale detection could not be made in near real-time due to the high overlap in characteristics with glider noise and humpback whale grunt sequences (Kowarski *et al.*, 2019); however, 70% of possible validated
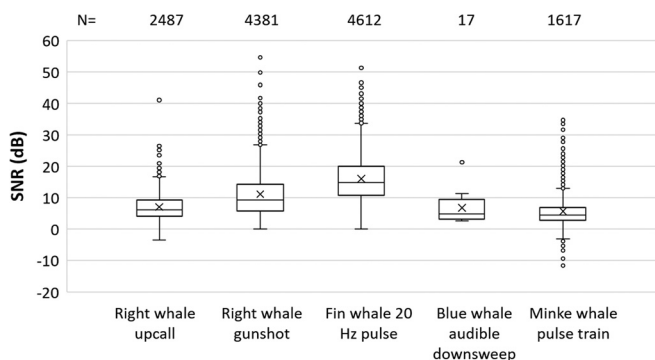

FIG. 5. Boxplots of the range of SNRs observed for each vocalization-type in the entire Gulf of St. Lawrence glider mission from 15 to 30 September 2018 as calculated from truth annotations.

minke whale detections were determined to be accurate based on comparison with recorded audio.

Vocalization-specific automated detectors performed as expected based on automated detector development (Table III). The "right whale 1" upcall automated detector had the highest precision of the right whale automated detectors, with 71% of candidate detections being true right whale upcall events, though it missed 89% of upcall events. The more inclusive right whale upcall automated detectors (2 and 3) had higher recalls (0.47 and 0.94, respectively) but were increasingly less precise (0.62 and 0.19, respectively). The fin whale 20 Hz automated detector captured 90% of fin whale vocalization events, but it regularly triggered on glider noise. With few blue whale vocalizations in the data (Fig. 4), the automated detectors could not be thoroughly assessed; the audible automated detector identified 50% of the blue whale vocalization events but was frequently triggered by glider noise (Fig. 3). One general automated

TABLE III. Performance by email (restricted to timeframes sent to shore) for all vocalization-specific automated detectors as well as one general automated detectors that captured minke whale pulse trains employed during the Gulf of St. Lawrence glider trial from 15 to 30 September 2018. Recordings used to edit automated detectors pre simulation were excluded from performance calculations.

Automated detector	Precision	Recall
Right whale 1	0.71	0.11
Right whale 2	0.62	0.47
Right whale 3	0.19	0.94
Fin whale	0.26	0.90
Blue whale audible	0.33	0.50
Minke whale (general automated detector)	0.10	0.98

detector triggered on 98% of minke whale pulse train events, but also triggered regularly on glider sounds.

Validated detections produced by human manual analysis in simulated near real-time produced highly precise

validated detections with zero false positive detections for right, fin, or blue whales ($P = 1.00$) when evaluating at a per email, dive, hourly, or daily basis (Table IV). Recall was more variable, depending on species and timeframe of performance evaluation. On a per-email basis (validated detections compared to truth data over the timeframe of the emails) R varied from 14% to 76% (Table IV). Acoustic presence of species was missed because there was too little information for the analyst to make a definite validated detection or signals were too faint (low SNR) for an automated detector to identify. SNR was found to impact recall on a per email basis with the recall higher for high than low SNR emails for right and fin whale vocalizations. Such a pattern was not apparent in either blue whale audible downsweeps where the sample size was extremely low or minke whale pulse trains where challenges were found in accurately calculating SNR as described previously (Table IV).

TABLE IV. Human validated near real-time detector performance including precision (P) and recall (R) for each species recorded during the simulation of the Gulf of St. Lawrence glider trial from 15 to 30 September 2018. Performance is given for the human detector (by email) and the entire system (by dive, hour, and day). The by email performance is restricted to timeframes associated with emails while the by dive, hour, and day performance incorporates all recordings, including periods never sent to shore. By email performance metrics are included for all emails and with the emails separated into those containing either high or low SNR vocalizations. Recordings (and their associated emails) used to edit automated detectors pre simulation were excluded from performance calculations. Minke whale metrics are for possible detections, while definite detection performance is presented for the remaining species.

	By email ($n = 641$)			Hourly ($n = 350$)	By dive ($n = 114$)	Daily ($n = 16$)
	All	**High SNR**	**Low SNR**			
Right whale (upcall)						
TP	72	57	15	37	23	8
FP	0	0	0	0	0	0
FN	23	14	9	81	37	5
P	1.00	1.00	1.00	1.00	1.00	1.00
R	0.76	0.80	0.63	0.31	0.38	0.62
Right whale (upcall and gunshot)						
TP	72	36	36	37	23	8
FP	0	0	0	0	0	0
FN	47	11	36	109	42	5
P	1.00	1.00	1.00	1.00	1.00	1.00
R	0.61	0.77	0.50	0.25	0.36	0.62
Fin whale (20 Hz pulse)						
TP	15	5	10	12	10	5
FP	0	0	0	0	0	0
FN	44	8	36	70	29	7
P	1.00	1.00	1.00	1.00	1.00	1.00
R	0.25	0.38	0.22	0.15	0.26	0.42
Blue whale (audible downsweep)						
TP	1	0	1	1	1	1
FP	0	0	0	0	0	0
FN	1	1	0	4	4	1
P	1.00	NA	1.00	1.00	1.00	1.00
R	0.50	0.00	1.00	0.20	0.20	0.50
Possible minke whale (pulse train)						
TP	7	2	5	7	7	5
FP	3	2	1	0	0	0
FN	42	23	19	112	53	9
P	0.70	0.50	0.83	1.00	1.00	1.00
R	0.14	0.08	0.21	0.06	0.12	0.36

The system was optimized to automatically detect only one vocalization-type of the right whale repertoire: the upcall. However, the truth data revealed the regular occurrence of gunshots (Fig. 4). The recall of right whale vocalizations differed depending on which vocalization types and timeframes were considered. Per email, recall decreased by 16% when gunshots were included. In contrast, on a daily basis, recall did not vary between when gunshots were included versus when only considering upcalls (Table IV).

IV. DISCUSSION

A. Addressing practical challenges

The glider successfully transmitted metadata to shore from the Gulf of St. Lawrence for six weeks, of which two weeks of acoustic data were recorded onboard the glider and used to optimize the system. The glider mission was simulated with the optimized system and evaluated for monitoring of baleen whale acoustic signals in near real-time. The practical challenges of PAM in near real-time from a glider platform were addressed throughout every stage of the optimized system.

Balancing the restriction of limited bandwidth with the information requirements for effective species validated detections was a consideration throughout the process. At the onboard candidate detection stage, automated detectors were used that focused on both identifying signals of interest and on capturing useful context around those signals. The embedded computer was sufficiently powerful to process eight different FFT settings that spanned across the 11 automated detectors. Timeframes likely to contain vocalizations of target species were prioritized for transmission to shore and subsequent manual review, rather than sending data in chronological order.

Data were sent from the glider as ensembles of candidate detections (candidate detections considered high priority as well as all surrounding candidate detections), maintaining contextual information through every step of the process. Context is an extremely important factor when differentiating baleen whale vocalizations from those of other species or other oceanic sounds (e.g., Baumgartner et al., 2019; Wright et al., 2019). By capturing context (as many sounds as possible) at the automated detector stage and presenting these as complete ensembles, the human validators were able to effectively identify vocalizations of marine mammals. The length of ensembles was flexible such that when a longer ensemble could not be sent, a shorter ensemble, restricted to the highest score within the large ensemble, would be created.

Candidate detector information was sent as contours. While being more costly in terms of bandwidth, this allowed the shape of the acoustical signal to more accurately mimic what would be seen on a spectrogram. The successful application of contours is a great option for sounds with more complex or specific shapes such as minke whale pulse trains but may be unnecessarily costly to the limited bandwidth budget for other vocalizations such as repeated blue whale infrasonic moans. In future glider missions, the present

system can be configured to produce contours, pitch-tracks, or both depending on the goals of the project and the balance the researchers seek to find between receiving more, longer emails with less information (pitch-track based), fewer, shorter emails with more information (contour-based), or a balance between the two (pitch-tracks for some automated detectors and contours for others). Each of the aforementioned steps was aimed to alleviate the challenge imposed by limited bandwidth and provide the manual analyst with sufficient information.

Another practical challenge addressed was ensuring the validated detections avoided all FPs, a requirement when validated detections can influence stakeholders in near real-time. The contextual information captured in emails, along with the rigorous protocol, allowed analysts to differentiate acoustic signals between species and from those produced by the glider. The protocol, combined with using analysts with previous experience with both baleen whale acoustic signals and those of gliders, contributed to the reliability of the validated detections.

Gliders have considerable energy constraints when compared to other PAM methodologies, which result in the practical challenge of balancing the energy available with the power and storage capabilities of the scientific payload. Work is currently underway to increase efficiency and reduce energy demands from the 2–3 W experienced in the present trial. While the current mission was only 45 days in duration, with similar energy demands, the use of an extended glider (with energy bay), and the inclusion of sufficient memory cards, a glider mission with an OceanObserver could operate and record acoustic data for approximately 100 days. The maximum feasible deployment length would depend on all energy demands on the glider (e.g., other sensors, dive depth, surface time). The OceanObserver successfully demonstrated high processing power and stored multiple terabytes of data at two sampling rates and processed 11 automated detectors with eight FFT settings simultaneously. Preliminary tests found that 21 automated detectors with ten different FFT settings (including delphinid whistle automated detectors) applied to data sampled at a rate of 32 kHz could successfully run on an OceanObserver simultaneously.

B. Outcomes and future improvements

The manual validation analysis to interpret metadata sent from gliders in near real-time has been applied to glider monitoring programs elsewhere (Klinck et al., 2012; Baumgartner et al., 2013; Baumgartner et al., 2014; Baumgartner et al., 2020) and was similarly found to effectively eliminate FPs in the present study. Indeed, the performance of the candidate detections alone was insufficient considering the high level of certainty required when informing potentially costly mitigation measures.

The system performed as designed during the simulation, with every definite baleen whale validated detection being accurate, though the proportion of missed validated

1226   J. Acoust. Soc. Am. **148** (3), September 2020

Kowarski et al.

detections varied depending on the timeframe and species considered. These findings should be considered during the planning of near real-time monitoring programs. The closest to real-time possible is the time between signal production and the glider surfacing plus 5–15 min for data transfer and analysis (assuming no delay in manual validation onshore). For practical purposes, dive time is the smallest unit that should be used when reporting real-time validated detection performance. In the present trial, glider dives lasted for approximately three hours, though two hours is commonly used (Baumgartner *et al.*, 2013; Baumgartner *et al.*, 2014; Baumgartner *et al.*, 2020). Differences in dive time will result in varied distance travelled, which then impacts the size of the area that can be acoustically monitored. If reporting acoustic occurrence on a daily basis (e.g., Baumgartner *et al.*, 2019), the recall will be higher (Table IV) and per-dive performance is less important. By allowing a greater timeframe (e.g., per day) for determining whale acoustic presence, the effects of variability in vocalization rate are reduced (e.g., a whale may be present and not vocalizing during one glider dive but vocalizing in the following dive).

As we have successfully created a system that produced reliable results on a simulated trial, the next step is to undertake additional trials to confirm the simulated system performs as expected. With every effort made to ensure the simulation was representative of the ocean trial, we expect the system to perform similarly in a subsequent fall trial in the Gulf of St. Lawrence. However, it would not be expected to perform exactly the same in different acoustic conditions (e.g., different season, location, or glider activity), a conclusion made by many researchers that utilize automation for identifying baleen whale vocalizations in acoustic recordings (Hodge *et al.*, 2015; Širović *et al.*, 2015; Erbs *et al.*, 2017). A high volume of glider missions repeatedly capturing the same species would give a more representative average of the system's performance per-species.

Future work should focus on improving the system's recall. Vocalizations of interest were missed either because the signals were too faint to trigger the automated detectors onboard the glider or the transmission budget was such that all ensembles could not be sent from the glider to shore. Faint signals will always be challenging for PAM data analysis, whether it is done in real-time or post-retrieval, but the limitations of transmission budget may be alleviated in the future as communications technology continues to improve.

To achieve a high recall, future missions can plan accordingly to reduce glider noise as any bandwidth wasted sending glider noise to shore could have been allocated to whale vocalizations. Glider noise can result in self-induced masking of baleen whale vocalizations and falsely triggering vocalization-specific automated detectors (Baumgartner *et al.*, 2013). The dive patterns of a glider mission should be altered based on the reason for data collection and in the case of acoustics, the pilot should create an energy efficient mission with slow dives, little thrust, and reduced need to ballast (Fregosi *et al.*, 2020). The position of the hydrophone on the glider should also be taken into consideration. For Slocum gliders, better noise reduction

has been associated with mounting the hydrophone on the aft of the glider (Lorenzo-Lopez, 2019). If a glider's mission is such that perfect precision is not required, recall can be increased by creating a more inclusive protocol or accepting possible validated detections as definite. Relevant glider information such as dive depth, heading, speed, and times of noisy operations should be sent along with acoustic information to analysts onshore and incorporated into protocols for determining marine mammal occurrence. Many of the aforementioned noise-reducing steps were not taken in the present trial, therefore, future trials with such improvements applied have the potential for successfully achieving increased recall.

The human detector performance presented here considers possible validated detections as negative to maintain a conservative outcome; however, more than 50% of possible validated detections were true for right and fin whales. If all possible validated detections were considered definite the system would have a perfect daily performance (P, 1.00; R, 1.00; Fig. 4). In general, by including possible validated detections as definite, R was increased but P was lowered. Given the importance of producing only accurate results for near real-time monitoring, the present approach that favored P to the detriment of R was necessary. However, in the future, management bodies and stakeholders should consider the inevitable pitfalls of favoring a highly precise system that misses many instances of whale presence. An incorrectly validated detection may result in unnecessary costs associated with changing vessel speed, but a missed validated detection may result in an injured or deceased North Atlantic right whale due to vessel strike that was avoidable. By shifting the methods to an approach that balances the importance of P and R, management can reach an arguably more appropriate compromise between minimizing impact to industry and still effectively protecting an endangered species.

The optimum surface time should also be reconsidered. If the glider spends more time at the surface between dives, more metadata can be sent through the limited iridium network, reducing the chance of an acoustic signal of interest not making it to shore. The present study sent 8 kB of ensemble data per hour to the glider's computer. In previous studies, up to 12 kB of metadata were sent per hour (Baumgartner *et al.*, 2014). The optimum dive and surface durations should be determined. This represents a trade-off between the amount of information a glider can send (surface time) and the amount of time a glider effectively monitors marine mammals (dive time). Another consideration is that long surface time can negate any forward progress of the previous dive cycle due to drift. Wave gliders that remain at the surface while towing a hydrophone present an opportunity for balance between surface transmission time and recording time, and they should be investigated in the future, though wave gliders come with a separate set of challenges including power restrictions, flow noise from the propulsion, and platform noise (Darling *et al.*, 2019). Other surface vehicles such as Datamaran (Autonomous Marine Systems Inc.), SailBuoy (Offshore Sensing), DriX (iXblue), and Saildrone could also be considered for integrating acoustics for near real-time marine mammal monitoring.

The present work revealed that by limiting efforts to a portion of the repertoire of each species, true acoustic occurrence was underestimated. For example, upcalls are the most common vocalization used to identify the occurrence of right whales in PAM data because they are produced by all age and sex classes; they make up a high percentage of right whale vocal repertoire, and they have little overlap in characteristics with other oceanic sounds (Table II; Parks *et al.*, 2011; Baumgartner *et al.*, 2019). However, because right whale gunshots were not identified and prioritized, the right whale recall was reduced by 16%, when considering short timescales. Gunshots were captured by general automated detectors that could not be given priority because they were also triggered by glider noise. Such was not an issue when considering daily timescales where recall was high regardless of gunshot inclusion, but for instances where information is important on a shorter timescale, reduced glider noise combined with more effective gunshot automated detectors in future missions would mitigate the problem.

Furthermore, the repertoires of most cetacean species are incompletely described. For example, many vocalizations cannot be confidently attributed to a specific species, resulting in "truth" datasets that include a large portion of unknowns or "possible" vocalizations (Fig. 4). When interpreting results from PAM one must consider that PAM methods can only determine the acoustic occurrence of animals that are acoustically active (often creating a bias towards detecting males, ex. Table II), and producing species-unique, previously described signals that are within detection range of the acoustic recorder.

Detection range of the species observed in the present study would have been impacted by species, vocalization type, the movements of the glider when the signal was detected, and other sounds contributing to the soundscape of the area such as currents and vessels. For example, right whale upcalls have a lower source level than gunshots, resulting in a smaller detection range (Parks and Tyack, 2005; Munger *et al.*, 2011). Laurinolli *et al.* (2003) investigated the detectability of North Atlantic right whale vocalizations near a shipping lane in the Bay of Fundy using static acoustic recorders and concluded that right whales could not be heard from more than 30 km away. In contrast, Munger *et al.* (2011) found that North Pacific right whales could be detected at a distance of 100 km in the shallow waters of the Bering Sea. Given the high vessel traffic in the Gulf of St. Lawrence, which reduces listening range of the animals (Pine *et al.*, 2018), and the interference from sounds caused by the constantly moving platform, the right whale detection range in the present study is likely closer to, if not less than, that described by Laurinolli *et al.* (2003).

C. Implications for North Atlantic right whale management

Since the North Atlantic right whale mortality event in the Gulf of St. Lawrence in the summer of 2017, the Government of Canada has dedicated an unprecedented amount of resources to protecting this species. In 2018, no right whales were reported dead in the Gulf of St. Lawrence, though eight perished in 2019 (Fisheries, 2019). To date, the implementation of dynamic mitigation zones has been based solely on visual survey data (Transport Canada, 2019a). However, gliders have been reporting whale occurrence in the region for years (Johnson, 2018) and PAM methods have demonstrated reliable results at lower cost and effort than sighting surveys (Soldevilla *et al.*, 2014; Baumgartner *et al.*, 2013; Baumgartner *et al.*, 2019).

Considering the present demonstration of an effective system, we propose that supplementing current aerial survey efforts with near real-time PAM on gliders or utilizing gliders to direct aerial surveys can reduce risk to whales, in the Gulf of St. Lawrence, and elsewhere. For example, currently in the Gulf of St. Lawrence, if a right whale is sighted within a slowdown buffer of a shipping lane (from 2.5 to 5 nm), the shipping lane speed restrictions are triggered (Transport Canada, 2019a). Gliders could effectively monitor these buffer regions, reporting when whales are present. Considerations should be made for possible validated detections. We found that possible validated detections were correct 50% of the time or more. Such validated detections should, therefore, be reported to stakeholders so that precautions can be taken. For example, vessel captains may watch for whales more vigilantly or voluntarily reduce speed if they are notified that there is a 50% chance whales are in or near the shipping lane.

V. CONCLUSIONS

A system to accurately report the acoustic occurrence of baleen whales in near real-time from a sub-surface glider platform was presented. Existing and novel methods were implemented to address the practical challenges of PAM on a glider, many of which contributed to ensuring that what little data could be sent from the glider was of high importance (prioritization) and contained enough contextual information (ensembles of candidate detection contours) for analysts onshore to create validated detections. The human analyst was key to the system, which resulted in perfect precision for all detected species. Recall was variable, depending on vocalization-type and timeframe considered. We propose that such systems should be implemented to inform dynamic management decisions. The methods can be applied not only to protect North Atlantic right whales in the Gulf of St. Lawrence, but to inform research, industry, and governments around the world of the occurrence of acoustically active aquatic species in near real-time.

1228   J. Acoust. Soc. Am. **148** (3), September 2020

Kowarski *et al.*

[1]See supplementary material at https://doi.org/10.1121/10.0001811 for manual analyst protocol.

[2]The OceanObserver can hold up to twelve 512 GB SD cards that would have allowed recording for the full duration of the trial (assuming no power constraints), but equipment availability at the time of deployment limited the system to four SD cards. Worth noting, the trial recorded data sampled at a rate of 512 kHz. By recording a single channel at a reduced sampling rate, future trials could greatly increase the recording duration.

Baumgartner, M. F., Bonnell, J., Corkeron, P. J., van Parijs, S. M., Hotchkin, C., Hodges, B. A., Bort Thornton, J., Mensi, B. L., and Bruner, S. M. (**2020**). "Slocum gliders provide accurate near real-time estimates of baleen whale presence from human-reviewed passive acoustic detection information," Front. Mar. Sci. **7**, 100.

Baumgartner, M. F., Bonnell, J., van Parijs, S. M., Corkeron, P. J., Hotchkin, C., Ball, K., Pelletier, L. P., Partan, J., Peters, D., and Kemp, J. (**2019**). "Persistent near real-time passive acoustic monitoring for baleen whales from a moored buoy: System description and evaluation," Methods Ecol. Evol. **10**, 1476–1489.

Baumgartner, M. F., and Fratantoni, D. M. (**2008**). "Diel periodicity in both sei whale vocalization rates and the vertical migration of their copepod prey observed from ocean gliders," Limnol. Oceanogr. Methods **53**, 2197–2209.

Baumgartner, M. F., Fratantoni, D. M., Hurst, T. P., Brown, M. W., Cole, T. V. N., van Parijs, S. M., and Johnson, M. (**2013**). "Real-time reporting of baleen whale passive acoustic detections from ocean gliders," J. Acoust. Soc. Am. **134**, 1814–1823.

Baumgartner, M. F., Stafford, K. M., Winsor, P., Statscewich, H., and Fratantoni, D. M. (**2014**). "Glider-based passive acoustic monitoring in the Arctic," Mar. Technol. Soc. J. **48**, 40–51.

Berchok, C. L., Bradley, D. L., and Gabrielson, T. B. (**2006**). "St. Lawrence blue whale vocalizations revisited: Characterization of calls detected from 1998 to 2001," J. Acoust. Soc. Am. **120**, 2340–2354.

Clark, C. W., and Altman, N. S. (**2006**). "Acoustic detections of blue whale (*Balaenoptera musculus*) and fin whale (*B. physalus*) sounds during a SURTASS LFA exercise," IEEE J. Ocean. Eng. **31**, 120–128.

Daoust, P. Y., Couture, E. L., Wimmer, T., and Bourque, L. (**2017**). "Incident report: North Atlantic right whale mortality event in the Gulf of St. Lawrence, 2017," Collaborative report by Canadian Wildlife Health Cooperative, Marine Animal Response Society, and Fisheries and Oceans Canada.

Darling, J. D., Goodwin, B., Goodoni, M. K., Taufmann, A. J., and Taylor, M. G. (**2019**). "Humpback whale calls detected in tropical ocean basin between known Mexico and Hawaii breeding assemblies," J. Acoust. Soc. Am. **145**, EL534–EL540.

Davis, R., Baumgartner, M., Comeau, A., Cunningham, D., Davies, K., Furlong, A., Johnson, H., L'Orsa, S., Ross, T., and Taggart, C. (**2016**). "Tracking whales on the Scotian Shelf using passive acoustic monitoring on ocean gliders," in *Proceedings of OCEANS 2016 MTS/IEEE* September 19–23, Monterey, CA, pp. 1–4.

Delarue, J. (**2008**). "Northwest Atlantic fin whale vocalizations: Geographic variations and implications for stock assessments," M.S. thesis, College of the Atlantic, Bar Harbor, ME.

Delarue, J., Kowarski, K. A., Maxner, E. E., MacDonnell, J. T., and Martin, S. B. (**2018**). "Acoustic monitoring along Canada's east coast: August 2015 to July 2017," Technical report by JASCO Applied Sciences for Environmental Studies Research Fund, Dartmouth, NS, Canada.

Delarue, J., MacDonnell, J. T., Martin, S. B., Mouy, X., and Hannay, D. E. (**2014**). "Northeastern Chukchi Sea joint acoustic monitoring program 2012–2013," Technical report for ConocoPhillips Company, Shell Exploration & Production Company, and Statoil USA E&P, Inc. by JASCO Applied Sciences.

Delarue, J., Todd, S. K., van Parijs, S. M., and Di Iorio, L. (**2009**). "Geographic variation in Northwest Atlantic fin whale (*Balaenoptera physalus*) song: Implications for stock structure assessment," J. Acoust. Soc. Am. **125**, 1774–1782.

DFO (**2019a**). "Fishing closures for North Atlantic right whale protection," http://www.dfo-mpo.gc.ca/fisheries-peches/commercial-commerciale/atl-arc/narw-bnan/index-eng.html (Last viewed August 13, 2019).

DFO (**2019b**). "Review of North Atlantic right whale occurrence and risk of entanglements in fishing gear and vessel strikes in Canadian waters," Canadian Science Advisory Secretariat Science Advisory Report 2019/028.

Erbs, F., Elwen, S. H., and Gridley, T. (**2017**). "Automatic classification of whistles from coastal dolphins of the southern African subregion," J. Acoust. Soc. Am. **141**, 2489–2500.

Fisheries (**2019**). "2017–2019 North Atlantic right whale unusual mortality event," https://www.fisheries.noaa.gov/national/marine-life-distress/2017-2019-north-atlantic-right-whale-unusual-mortality-event (Last viewed September 19, 2019).

Fregosi, S., Harris, D. V., Matsumoto, H., Mellinger, D. K., Negretti, C., Moretti, D. J., Martin, S. W., Matsuyama, B., Dugan, P. J., and Klinck, H. (**2020**). "Comparison of fin whale 20 Hz call detections by deep-water mobile autonomous and stationary recorders," J. Acoust. Soc. Am. **147**, 961–977.

Frouin-Mouy, H., Kowarski, K. A., Martin, B., and Bröker, K. (**2017**). "Seasonal trends in acoustic detection of marine mammals in Baffin Bay and Melville Bay, northwest Greenland + Supplementary Appendix 1," Arctic **70**, 59–76.

Hannay, D. E., Mouy, X., and Li, Z. (**2016**). "An automated real-time vessel sound measurement system for calculating monopole source levels using a modified version of ANSI/ASA S12.64-2009," Can Acoust **44**, 166–167.

Hodge, K. B., Muirhead, C. A., Morano, J. L., Clark, C. W., and Rice, A. N. (**2015**). "North Atlantic right whale occurrence near wind energy areas along the mid-Atlantic US coast: Implications for management," Endanger Species Re.s **28**, 225–234.

JASCO Applied Sciences (**2018**). *AMAR G4: Autonomous Multichannel Acoustic Recorder Generation 4* (JASCO Applied Sciences, Halifax, Canada).

Johnson, H. D. (**2018**). "WhaleMap," https://whalemap.ocean.dal.ca/WhaleMap/ (Last viewed August 7, 2019).

Klinck, H., Mellinger, D. K., Klinck, K., Bogue, N. M., Luby, J. C., Jump, W. A., Shilling, G. B., Litchendorf, T., Wood, A. S., Schorr, G. S., and Baird, R. W. (**2012**). "Near-real-time acoustic monitoring of beaked whales and other cetaceans using a Seaglider^TM," PLoS ONe **7**, e36128.

Kowarski, K. A., Evers, C., Moors-Murphy, H., Martin, B., and Denes, S. L. (**2018**). "Singing through winter nights: Seasonal and diel occurrence of humpback whale (*Megaptera novaeangliae*) calls in and around the Gully MPA, offshore eastern Canada," Mar. Mamm. Sci. **34**, 169–189.

Kowarski, K. A., Moors-Murphy, H., Maxner, E. E., and Cerchio, S. (**2019**). "Western North Atlantic humpback whale fall and spring acoustic repertoire: Insight into onset and cessation of singing behavior," J. Acoust. Soc. Am. **145**, 2305–2316.

Küsel, E. T., Munoz, T., Siderius, M., Mellinger, D. K., and Heimlich, S. (**2017**). "Marine mammal tracks from two-hydrophone acoustic recordings made with a glider," Ocean Sci. J. **13**, 273–288.

Laurinolli, M. H., Hay, A. E., Desharnais, F., and Taggart, C. T. (**2003**). "Localization of North Atlantic right whale sounds in the Bay of Fundy using a sonobuoy array," Mar. Mamm. Sci. **19**, 708–723.

Lorenzo-Lopez, A. (**2019**). (private communication).

Marotte, E. C., and Moors-Murphy, H. (**2015**). "Seasonal occurrence of blue whale (*Balaenoptera musculus*) vocalizations in the Gully Marine Protected Area," Can. Acoust. **43**, 50–51.

Martin, B., Kowarski, K. A., Mouy, X., and Moors-Murphy, H. (**2014**). "Recording and identification of marine mammal vocalizations on the scotian shelf and slope," in *Proceedings of 2014 Oceans MTS/IEEE*, September 14–19, St. John's, Canada, pp. 1–6.

Mellinger, D. K., Stafford, K. M., Moore, S. E., Dziak, R. P., and Matsumoto, H. (**2007**). "An overview of fixed passive acoustic observation methods for cetaceans," Oceanography **20**, 36–45.

Mohammad, B., and McHugh, R. (**2011**). "Automatic detection and characterization of dispersive North Atlantic right whale upcalls recorded in a shallow-water environment using a region-based active contour model," IEEE J. Ocean. Eng. **36**, 431–440.

Moloney, J., Cole, A., Hillis, C. A., Kowarski, K. A., MacDonald, B., Pedersen, G., Tassara, L., and Camus, L. (**2018**). "Observer/AMAR G4 Passive Acoustic Monitoring System for Autonomous Maritime Observation," in *Proceedings of AUV 2018*, November 6–9, Porto, Portugal.

Moore, S. E., Howe, B. M., Stafford, K. M., and Boyd, M. L. (**2007**). "Including whale call detection in standard ocean measurements: Application of acoustic Seagliders," Mar. Technol. Soc. J. **41**, 53–57.

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.

Parks, S. E., Searby, A., Celerier, A., Johnson, M. P., Nowaceck, D. P., and Tyack, P. L. (**2011**). "Sound production behavior of individual North Atlantic right whales: Implications for passive acoustic monitoring," Endanger. Species Res. **15**, 63–76.

Parks, S. E., and Tyack, P. L. (**2005**). "Sound production by North Atlantic right whales (*Eubalaena glacialis*) in surface active groups," J. Acoust. Soc. Am. **117**, 3297–3306.

Pine, M. K., Hannay, D. E., Insley, S. J., Halliday, W. D., and Juanes, F. (**2018**). "Assessing vessel slowdown for reducing auditory masking for marine mammals and fish of the western Canadian Arctic," Mar. Pollut. Bull. **135**, 290–302.

Risch, D., Clark, C. W., Dugan, P. J., Popescu, M., Siebert, U., and van Parijs, S. M. (**2013**). "Minke whale acoustic behavior and multi-year seasonal and diel vocalization patterns in Massachusetts Bay, USA," Mar. Ecol. Prog. Ser. **489**, 279–295.

Silva, T. L., Mooney, T. A., Sayigh, L. S., and Baumgartner, M. F. (**2019**). "Temporal and spatial distributions of delphinid species in Massachusetts Bay (USA) using passive acoustics from ocean gliders," Mar. Ecol. Prog. Ser. **631**, 1–17.

Širović, A., Hildebrand, J. A., Wiggins, S. M., and Thiele, D. (**2009**). "Blue and fin whale acoustic presence around Antarctica during 2003 and 2004," Mar. Mamm. Sci. **25**, 125–136.

Širović, A., Rice, A., Chou, E., Hildebrand, J. A., Wiggins, S. M., and Roch, M. A. (**2015**). "Seven years of blue and fin whale call abundance in the Southern California Bight," Endanger. Species Res. **28**, 61–76.

Soldevilla, M. S., Rice, A. N., Clark, C. W., and Garrison, L. P. (**2014**). "Passive acoustic monitoring on the North Atlantic right whale calving grounds," Endangered Species Research **25**(2), 115–140.

Spaulding, E., Robbins, M., Calupca, T., Clark, C. W., Tremblay, C., Waack, A., Warde, A., Kemp, J., and Newhall, K. (**2009**). "An autonomous, near-real-time buoy system for automatic detection of North Atlantic right whale calls," in *Proceedings of the 157th Meeting on Acoustics*, October 26–30, San Antonio, TX, 010001.

Thomisch, K., Boebel, O., Clark, C. W., Hagen, W., Spiesecke, S., Zitterbart, D. P., and van Opzeeland, I. (**2016**). "Spatio-temporal patterns in acoustic presence and distribution of Antarctic blue whales Balaenoptera musculus intermedia in the Weddell Sea," Endanger. Spec. Res. **30**, 239–253.

Transport Canada (**2019a**). "Government of Canada introduces new, additional measures to protect the North Atlantic right whale," https://www.canada.ca/en/transport-canada/news/2019/07/government-of-canada-introduces-new-additional-measures-to-protect-the-north-atlantic-right-whale.html (Last viewed 20 Aug 2020).

Transport Canada (**2019b**). "Static protected area. Protecting North Atlantic right whales from collisions with ships in the Gulf of St. Lawrence," https://www.tc.gc.ca/en/services/marine/navigation-marine-conditions/protecting-north-atlantic-right-whales-collisions-ships-gulf-st-lawrence.html#toc_2. (Last viewed August 13, 2019).

Tremblay, C. J., van Parijs, S. M., and Cholewiak, D. (**2019**). "50 to 30-Hz triplet and singlet down sweep vocalizations produced by sei whales (Balaenoptera borealis) in the western North Atlantic Ocean," J. Acoust. Soc. Am. **145**, 3351–3358.

van Parijs, S. M., Clark, C. W., Sousa-Lima, R. S., Parks, S. E., Rankin, S., Risch, D., and Van Opzeeland, I. C. (**2009**). "Management and research applications of real-time and archival passive acoustic sensors over varying temporal and spatial scales," Mar. Ecol. Prog. Ser. **395**, 21–36.

Wimmer, T. (**2020**). (private communication).

Wimmer, J., Towsey, M., Planitz, B., Roe, P., and Williamson, I. (**2010**). "Scaling acoustic data analysis through collaboration and automation," in *Proceedings of the IEEE 6th International Conference on e-Science* December 7–10, Brisbane, Australia, pp. 308–315.

Wright, D. L., Berchok, C. L., Crance, J. L., and Clapham, P. J. (**2019**). "Acoustic detection of the critically endangered North Pacific right whale in the northern Bering Sea," Mar. Mamm. Sci. **35**, 311–326.