Data Analysis Component of Comparative Passive Acoustic Monitoring (PAM) Technology Assessment

FINAL REPORT

Project number: SDP 200-206
Start date: August 8, 2019
Reporting period: August 8, 2019 – May 31, 2020
Recipient name: Fundy Ocean Research Centre for Energy
Project lead: Daniel J. Hasselman

Prepared by: Daniel J. Hasselman[1*], Tyler Boucher[1], and Jessica Douglas[1]

[1]Fundy Ocean Research Center for Energy, Halifax, NS
*Corresponding author: dan.hasselman@fundyforce.ca

Submission date: June 1, 2020

Contents

Executive Summary

Passive Acoustic Monitoring (PAM) technologies are commonly used to monitor echolocating marine mammals in tidal channels, but their detection efficiency is hampered by a series of factors in high flow environments (e.g., flow noise, ambient noise) that can ultimately impede monitoring efforts. In partnership with The Pathway Program, the Fundy Ocean Research Center for Energy (FORCE) conducted an assessment for two 'stand-alone' (i.e., CPOD, FPOD) and three 'conventional' (i.e., AMAR, SoundTrap, icListen) PAM instruments to understand the operational limitations of these 'off the shelf' technologies. The PAM instruments were mounted to a subsea platform and deployed at the FORCE tidal demonstration site. A series of passive drifts were then conducted over the platform from a vessel across a range of tidal flow conditions while playing synthetic clicks ('pseudo clicks') emitted from an icTalk. This data was supplemented with that collected from real harbour porpoise transiting the FORCE site.

Pseudo clicks were insufficiently similar to real harbour porpoise click trains to be classified by either the CPOD or FPOD; limiting comparisons of these devices to real harbour porpoise clicks collected during the deployment. 'Conventional' PAM technologies (AMAR, SoundTrap, icListen) detected pseudo clicks over short ranges (median detection range: ~40m) due to the lower source level of the icTalk (~130 dB re 1μPa at 1m) relative to real porpoise clicks (~160 dB re 1μPa at 1m). The low source level of the pseudo clicks necessitated increasing the sensitivity of the detectors, which in turn increased the frequency of false-positive detections. True detections of pseudo clicks decreased with increasing flow speed, with few detections above current velocity of 2 m/sec. The icListen appeared to be less impacted by ambient noise at high frequencies, which allowed this instrument to detect more pseudo clicks than the other conventional PAM technologies assessed in this study. While all five PAM technologies were able to detect real harbour porpoise clicks, the false positive detection rates for the three 'conventional' instruments were higher than the 'stand-alone' instruments, creating additional post-processing steps. Reducing the sensitivity of the 'conventional' instruments decreased the instance of false positive detections, suggesting that further efforts on the classification of detections could reduce the rate of false positive detections while keeping recall high.

The choice of which PAM device to use depends on the scientific questions being asked. A primary objective of The Pathway Program is to establish a regulator-approved monitoring solution that can be used by tidal energy developers for monitoring the near-field (0 - 100m) region of their tidal energy device at the FORCE demonstration site. To that end, if the appropriate sensitivity settings are coupled with a very good classifier, 'conventional' PAM instruments could be used for monitoring tidal turbines at the FORCE site. Of those instruments tested, the icListen HF hydrophone was less impacted by ambient noise at the FORCE site, and is recommended for advancement to subsequent testing phases focused on how to configure and deploy this technology for marine mammal monitoring in high flow environments.

Introduction and Objectives

Passive Acoustic Monitoring (PAM) technologies are frequently used to monitor echolocating marine mammals (primarily porpoise and dolphin) in high flow environments that are sought after for instream tidal power development (Adams et al., 2019; Malinka et al., 2018). These technologies generally fall into two categories: i) 'conventional' instruments that frequently require separate hardware (recording) and software (signal detection and classification) systems (e.g., Ocean Sonics icListen HF), and ii) 'stand-alone' instruments that allow the pressure time series to be analysed in real time using some prescribed criteria for signal detection and classification; permitting the raw data to be discarded while retaining the associated metadata (e.g., Chelonia Ltd. 'CPOD') (Hasselman et al., 2020). Regardless of the technology category, the detection efficiency of PAM instruments for monitoring vocalizing marine mammals in tidal channels is impacted by a variety of factors, including the vocalizing bandwidth for the species of interest and the potential masking of these sounds by flow noise and ambient sound (e.g., sediment transport on the seafloor), as well as the propagating environment, reverberation, sensor placement and sensor deployment methodology (Hasselman et al., 2020). Although 'conventional' PAM instruments generally have a greater detection range (0-500 m) than 'stand-alone' instruments (0-300 m), this depends on the conditions under which tests are conducted (Benjamins et al., 2017; Kyhn et al., 2008, 2012; Polagye et al., 2012; Porskamp et al., 2015; Roberts & Read, 2015; Tollit & Redden, 2013). For instance, in tidal channels the detection range is dictated by the high sea water absorption coefficient at a specified frequency (e.g., 130 kHz), and signal attenuation due to bubble scattering and turbulent mixing in high flow environments that can further decrease detection range (Hasselman et al., 2020).

The Fundy Ocean Research Centre for Energy (FORCE) operates an instream tidal power demonstration site in Minas Passage, Bay of Fundy, and conducts a series of monitoring programs that include the use of PAM technologies for monitoring harbour porpoise (*Phocoena phocoena*) activity. This region is dominated by the highest tidal amplitudes in the world (>6 m; Karsten et al., 2008) with tidal flow speeds regularly exceeding 5 m/sec (Karsten et al., 2011). This culminates in a dynamic marine environment dominated by turbulent conditions and excessive flow noise (Martin et al., 2018) that presents challenges for monitoring harbor porpoise using PAM technologies. FORCE uses CPODs (a 'stand-alone' instrument) mounted on Streamlined Underwater Buoyancy Systems (SUBS) that are moored to the seabed to monitor harbour porpoise at its demonstration site (Figure 1). However, this technology and deployment methodology has several limitations (Adams et al., 2019) related to performance (i.e., increased % time lost) during the high current speeds that represent a substantial proportion of the tidal cycle at the FORCE site (Joy et al., 2018). Flow-induced noise in tidal channels can cause the maximum number of recordable clicks per minute to be exceeded on 'stand-alone' instruments, resulting in saturation of the detection buffer, and generating 'lost time' (Tollit & Redden, 2013); ultimately leading to the under-reporting of harbour porpoise detections that can obfuscate the results of monitoring efforts. The deployment methodology

also factors into the generation of 'lost time', as bottom-mounted 'stand-alone' instruments generally have greater detection minutes per day than moored systems (Wilson et al., 2013), where noise generated by the mooring system being 'blown down' against the seabed during periods of high flow may saturate the detection buffer of the instrument (Porskamp et al., 2015). For instance, in the Minas Passage, the mean percent lost time across five moored CPODs was 22.3%, with the percent lost time surpassing 65% when current speeds during the ebb and flood tide exceeded 2.9 m/sec and 3.5 m/sec, respectively (Joy et al., 2018).
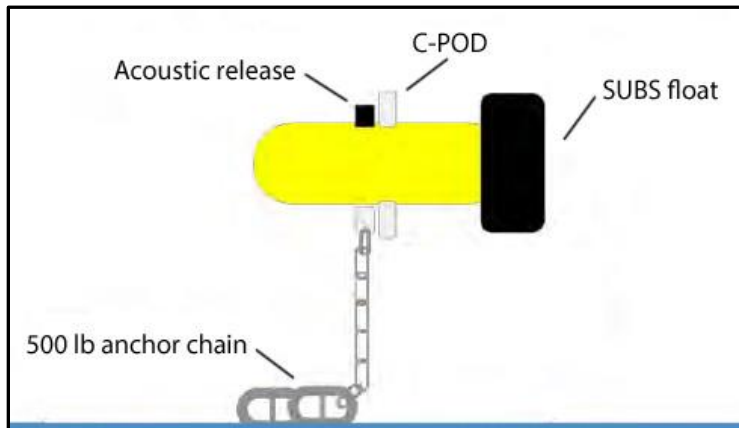

Figure 1: Diagram of FORCE's SUBS package deployed on seabed with CPOD and mooring design.

The limitations that have been identified when using CPODs to monitor harbour porpoise in the Minas Passage warrant a critical assessment of the capabilities of alternative PAM technologies that may provide a more suitable means for monitoring in high flow environments. To that end, the primary objective of this study is to understand the operational limitations of several 'off the shelf' PAM technologies across a range of tidal flow conditions so that an informed decision can be made about which technology is best suited for monitoring harbour porpoise in the Minas Passage, and specifically at the FORCE demonstration site. This study constitutes the crucial first phase of comparative tests for PAM technologies under Phase 3 ('Technology Validation') of The Pathway Program[1], and serves to identify which PAM instrument should be advanced to successive testing phases focused on how to configure and deploy PAM technology for monitoring harbour porpoise in high flow environments.

SMRU Canada Ltd. conducted the data analyses component of this work, and their final report is included herein as an Appendix. The body of this report outlines the laboratory and field components of this project, and only reflects the main points of the SMRU report in the Results and Conclusions/Recommendations sections. For a more thorough understanding of the results, readers are encouraged to read the Appendix.

---

[1] https://oera.ca/research/pathway-program-towards-regulatory-certainty-instream-tidal-energy-projects

Methodology

Laboratory procedures

Five PAM technologies were identified for assessment, and included three 'conventional' instruments (i.e., Ocean Instruments SoundTrap 300 HF, OceanSonics icListen HF, and JASCO Applied Sciences AMAR-G4) and two 'stand-alone' instruments (i.e., Chelonia Ltd. CPOD and CPOD-F, hereafter 'FPOD'). Prior to field testing, an acoustic time-synchronization was conducted in a controlled laboratory setting (i.e., Aquatron Pool Tank at Dalhousie University) using synthetic harbour porpoise clicks (hereafter 'pseudo clicks') emitted by an OceanSonics icTalk (130 kHz) as a positive control. This test also served to verify that each PAM instrument was operating as expected (confirmed by SMRU Canada Ltd.). Briefly, each PAM instrument was deployed under surface floats that were distributed around the periphery of the Aquatron Pool Tank (Figure 2a) while the icTalk emitted pseudo clicks while deployed below a floating platform located approximately equidistant from each PAM instrument (Figure 2b).

Figure 2: (a) Schematic of the Aquatron Pool Tank showing the placement of PAM instruments and icTalk used for acoustic time synchronization, and (b) photograph of the laboratory based assessment showing the location of surface floats and the floating platform where the icTalk was deployed (photo credit: Tyler Boucher).

Field testing

After confirmation that all PAM instruments were operating properly, they were mounted to one of FORCE's FAST (Fundy Advanced Sensor Technology) subsea platforms and prepared for deployment (Figure 3). The platform was deployed at the FORCE site (Figure 4) over two periods (July 10 - 29 and September 5 - 13, 2019) to record pseudo clicks and real harbour porpoise clicks over a range of tidal flow conditions. Passive drifts were conducted on July 11 and September 6, 2019 over the platform from the RV Nova Endeavor with the icTalk deployed over the side of the vessel, and were conducted across an entire tidal cycle (i.e., flood and ebb tide) to determine the ability of each PAM instrument to detect this positive control signal across the range of flow conditions experienced at the FORCE site. The center frequency of the pseudo clicks from the icTalk was 130 kHz, with pseudo clicks produced every 0.3 seconds at

peak-to-peak source levels of 130 dB re 1µPa. A handheld GPS (Garmin Oregon 600) recorded the vessel tracks during the passive drifts, and an acoustic Doppler current profiler (ADCP) mounted on the Cape Sharp Tidal Venture turbine deployed in the vicinity of the FAST platform recorded current velocity. Additional CPODs were deployed around the site as part of FORCE's ongoing Environmental Effects Monitoring Program, and the CPOD deployed at location D1 (~84 m from the platform) was used to ground truth the data collected by the CPOD mounted on the FAST platform (Figure 4).
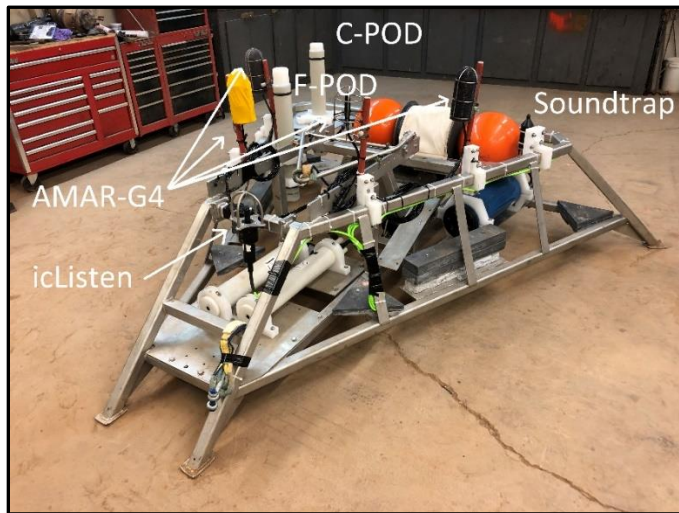

Figure 3: Final configuration of the FAST platform with five PAM devices mounted on platform frame using mounting brackets. The AMAR-G4 has four hydrophones (i.e., 'channels'), three of which were protected by coverings (two different densities of foam, and a yellow 'sock').
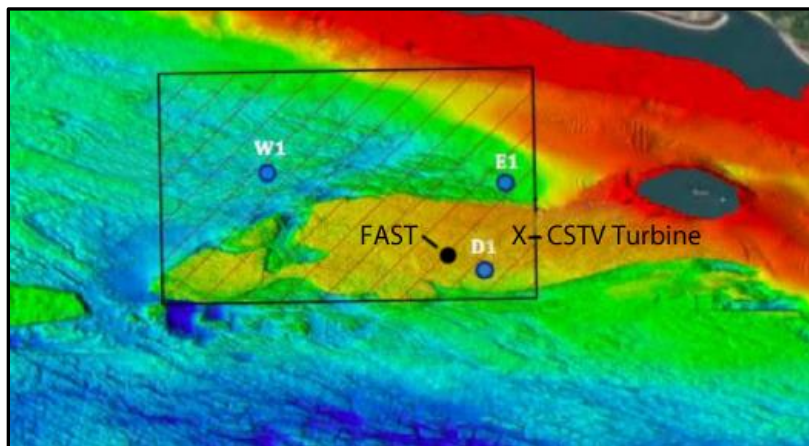

Figure 4: Map of the FORCE tidal demonstration site showing the approximate locations for CPODs deployed on SUBS (W1, E1, D1), the deployment location of the FAST subsea platform with PAM devices (●) and the location of the CSTV turbine (X).

Data analysis

Upon platform recovery, data were downloaded and provided to SMRU for standard QA/QC procedures and analyses. This included checking time stamps and start/stop times for all data sets, a series of click annotation steps to establish a 'gold standard' for comparison, time synchronization procedures, and subjecting the data to click detectors and classifiers (PAMGuard; 6 dB threshold) to determine instances of true positive, false positive and false negative click detections. The performance (i.e., accuracy) of the different PAM devices was then compared by calculating an equally weighted F-score that accounts for both precision (i.e., the proportion of detections that were true positives) and recall (i.e., the proportion of true positive detections captured by the classifier). Details are provided in the Appendix.

Results

A total of 11,678 pseudo clicks from the icTalk were annotated across 52 passive drifts during the deployments in July and September 2019. Instances where pseudo clicks were not recorded during some passive drifts were likely due to the low source level of the icTalk, flow noise, and the passive drift not passing close to the FAST platform, or some combination thereof. Pseudo clicks were insufficiently similar to real harbour porpoise click trains to be classified by either the CPOD or FPOD; limiting comparisons of these devices to real harbour porpoise clicks collected during the deployment (see below). This is because the classifier in these 'stand-alone' instruments focuses on identifying click trains, not individual clicks.

Figure 5 provides an example of a passive drift (September 2019) with the icTalk where automated detections were numerous. In general, similar plots revealed that i) detections increased as the icTalk drifted away from the platform rather than towards it (i.e., more distant detections 'downstream' than 'upstream'), ii) the SoundTrap and icListen documented more false positive detections than the AMAR, but the AMAR had fewer overall detections (including true positives) than the SoundTrap or icListen, and 3) true positive detections were more likely at current velocities <1.7 m/sec. The detection rate decreases markedly at current velocities >2.0 m/sec.

Across both deployments, the icListen HF hydrophone had the highest recall (62%) and precision (15%) of pseudo clicks, leading to the greatest F-score (0.25) (Table 1). However, large variance in detector performance was observed between the two deployments, suggesting subtle differences in platform deployment orientation and an important role for how PAM instruments are deployed. The icListen also had the greatest median detection range (43.9 m) than the AMAR (38.8 m) or SoundTrap (39 m). These are smaller detection ranges than expected and are likely attributable to the low source level of the icTalk. The results also revealed that the icListen had consistently lower Sound Pressure Levels (SPL) during high flow conditions, suggesting that this instrument has lower sensitivity to flow noise. This explains why the icListen records a higher number of true positive detections (i.e., because the noise is

lower, the signal to noise ratio is higher, and the detection probability is better). Many of the false positive detections occurred during periods when flow velocity was <2 m/sec, and when sand particles with diameters of 0.25mm are in suspension and generate click-like sounds at 130 kHz.
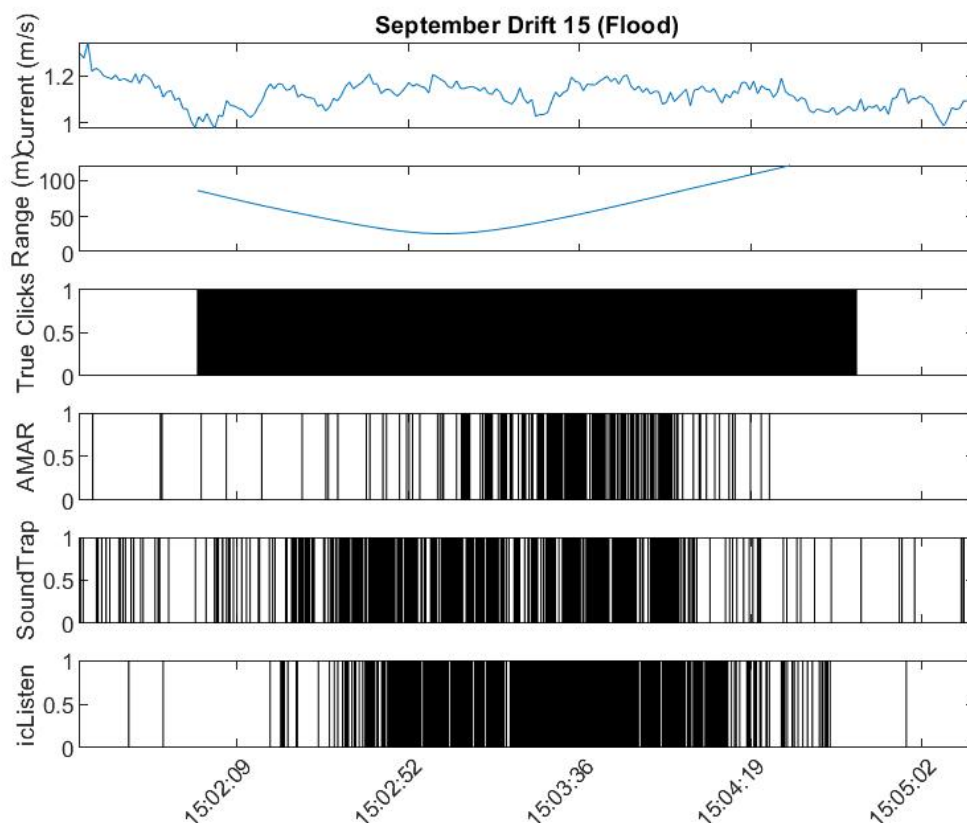

Figure 5: Data collected for a single drift experiment including current speed (top panel), range from the icTalk to the FAST platform (second panel), time of icTalk pseudo click projections (third panel), and time of PAMGuard click detections as recorded on the three 'conventional' PAM instruments (AMAR, SoundTrap, and icListen).

Table 1: Results of detector metrics at the click scale for the July and September deployments combined.

Hardware	Threshold (dB)	Deployment	Annotated clicks	TP	FP	FN	Precision	Recall	F-score
AMAR	6	1&2	11,678	3,689	25,025	7989	0.13	0.32	0.18
SoundTrap	6	1&2	11,678	3,431	152,806	8247	0.02	0.29	0.04
icListen	6	1&2	11,678	7,256	39,800	4422	0.15	0.62	0.25

The pseudo click data generated by the icTalk was supplemented with 17,826 individual clicks (281 click trains) from real harbour porpoise collected opportunistically recorded during both deployments. Table 2 provides the resulting detection metrics. Most devices had the 11, 461 minutes of data available for analysis except for the icListen which only had 2,107 minutes due to memory capacity limitations of the instrument. While all five PAM technologies were able to detect real harbour porpoise clicks, the false positive detection rates for the three 'conventional' instruments were higher than the 'stand-alone' instruments, creating additional post-processing steps. However, reducing the sensitivity of the 'conventional' instruments decreased the instance of false positive detections, suggesting that further efforts on the classification of detections could reduce the rate of false positive detections while keeping recall high.

Table 2: Results of detector metrics at the minute scale using porpoise clicks.

Hardware	Threshold (dB)	Deployment	Minutes of Data	Annotated DPM	TP	FP	FN	Precision	Recall	F-score
AMAR	6	2	11,461	10	10	5682	0	0.00	1.00	0.00
SoundTrap	6	2	11,461	10	10	9548	0	0.00	1.00	0.00
icListen	6	2	2,017	2	2	1871	0	0.00	1.00	0.00
CPOD	NA	2	11,461	10	6	62	4	0.09	0.60	0.15
FPOD	NA	2	11,461	10	4	72	6	0.05	0.40	0.09
CPOD-D1	NA	2	11,461	10	0	63	10	0.00	0.00	0.00
AMAR	21	2	11,461	10	8	616	2	0.01	0.80	0.03

Conclusions and Recommendations

Monitoring echolocating marine mammals using PAM technologies in locations that are dominated by high current velocities is inherently challenging. The choice of which PAM device to use, and its associated hardware and software, depends on the scientific questions being asked; particularly those by regulatory agencies if the monitoring is related to industry. A primary objective of The Pathway Program is to establish a regulator-approved monitoring solution that can be used by tidal energy developers for monitoring the near-field (0 - 100m) region of their tidal energy devices at the FORCE demonstration site.

This study focused on assessing the efficacy of five 'off the shelf' PAM technologies for monitoring in the high flow environment of the Minas Passage. These comparisons were made difficult by the different ways these devices function (classifiers) and the challenges of generating detectable pseudo clicks in a high flow environment from a device with a low source level. The CPOD and FPOD could not be directly compared with the 'conventional' instruments

because the classifiers in those 'stand-alone' instruments rely on click trains for logging detections, as opposed to individual clicks. Further, the threshold for detections had to be set at a very sensitive level (i.e., low threshold; 6 dB) to detect the low level of the pseudo clicks generated by the icTalk. However, this came at the expense of an increased frequency of false positive detections. Under these conditions the icListen HF had the greatest recall, precision, accuracy (i.e., F-score), and detection range of all 'conventional' PAM instruments assessed. Moreover, the icListen was less sensitive to high frequency ambient noise in Minas Passage, increasing the signal to noise ratio and allowing the detection of more true positive pseudo click detections.

Each of the 'conventional' PAM instruments examined in this study experienced excessively high rates of false positive detections stemming from the need to set the PAMGuard detection threshold at a very sensitive level (6 dB). At that unrealistically sensitive threshold, each of these instruments would provide misleading monitoring results. As such, there is a need to develop better classifier and detection algorithms so that 'conventional' PAM instruments are better suited for monitoring tidal turbines at the FORCE tidal demonstration site. Indeed, the performance of the icListen in high flow environments could be improved if the appropriate sensitivity settings are coupled with an appropriate classifier algorithm. This is something The Pathway Program is now pursuing under Phase 2 ('Data Automation') of the program. As such, the icListen HF hydrophone is recommended for advancement to subsequent testing phases of The Pathway Program focused on how to configure and deploy this technology for marine mammal monitoring in high flow environments.
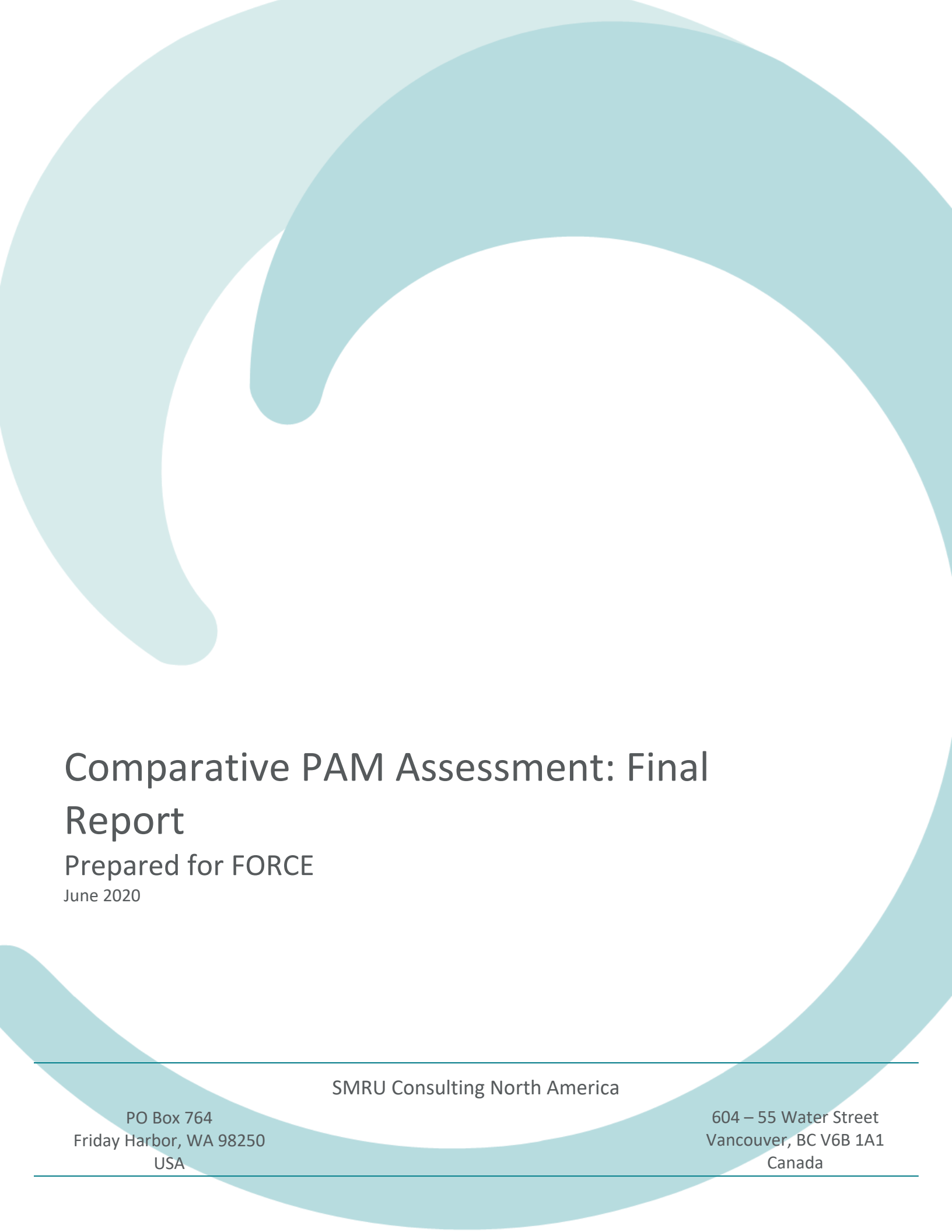
References

Adams, M., Sanderson, B., Porskamp, P., & Redden, A. (2019). Comparison of co-deployed drifting passive acoustic monitoring tools at a high low tidal site: C-PODs and icListern hydrophones. *Jourmal of Ocean Technology*, 1–33.

Benjamins, S., van Geel, N., Hastie, G., Elliott, J., & Wilson, B. (2017). Harbour porpoise distribution can vary at small spatiotemporal scales in energetic habitats. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *141*, 191–202. https://doi.org/10.1016/j.dsr2.2016.07.002

Hasselman, D. J., Barclay, D. R., Cavagnaro, R., Cotter, E., Gillespie, D. M., Hastie, G. D., Horne, J. K., Joslin, J., Lindsay, J., Mueller, R. P., Sparling, C. E., & Williamson, B. J. (2020). Environmental monitoring technologies and techniques for detecting interactions of marine animals with MRE devices. In *2020 State of the Science Report: Environmental Effects of Marine Renewable Energy Development Around the World* (p. 279). OES-Environmental.

Joy, R., Wood, J., & Tollit, D. (2018). *FORCE echolocating marine mammal environmental effects monitoring program - 2nd year (2018) monitoring report*.

Karsten, R., Greenberg, D., & Tarbotton, M. (2011). *Assessment of the potential of tidal power from Minas Passage and Minas Basin*.

Karsten, R. H., McMillan, J. M., Lickley, M. J., & Haynes, R. D. (2008). Assessment of tidal current energy in the Minas Passage, Bay of Fundy. *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy*, *222*(5), 493–507. https://doi.org/10.1243/09576509JPE555

Kyhn, L. A., Tougaard, J., Teilmann, J., Wahlberg, M., Jørgensen, P. B., & Bech, N. I. (2008). Harbour porpoise (Phocoena phocoena) static acoustic monitoring: laboratory detection thresholds of T-PODs are reflected in field sensitivity. In *Journal of the Marine Biological Association of the United Kingdom* (Vol. 88, Issue 6, pp. 1085–1091). https://doi.org/10.1017/s0025315408000416

Kyhn, L. A., Tougaard, J., Thomas, L., Duve, L. R., Stenback, J., Amundin, M., Desportes, G., & Teilmann, J. (2012). From echolocation clicks to animal density-acoustic sampling of harbor porpoises with static dataloggers. *The Journal of the Acoustical Society of America*, *131*(1), 550–560. https://doi.org/10.1121/1.3662070

Malinka, C. E., Gillespie, D. M., Macaulay, J. D. J., Joy, R., & Sparling, C. E. (2018). First in situ passive acoustic monitoring for marine mammals during operation of a tidal turbine in Ramsey Sound, Wales. *Marine Ecology Progress Series*, *590*, 247–266. https://doi.org/10.3354/meps12467

Martin, B., Whitt, C., & Horwich, L. (2018). *Acoustic data analysis of the OpenHydro opencentre turbine at FORCE: final report*.

Polagye, B., Thomsen, F., Bassett, C., Graber, J., & Cavagnaro, R. (2012). *Study of the acoustic effect of hydrokinetic tidal turbines in Admiralty Inlet , Puget Sound*. http://www.osti.gov/bridge/purl.cover.jsp?purl=/1039434/

Porskamp, P. H. J., Broome, J. E., Sanderson, B. G., & Redden, A. M. (2015). Assessing the performance of two passive acoustic monitoring technologies for porpoise detection in a high flow tidal site. *Canadian Acoustics*, *43*(3), 2–3.

Roberts, B. L., & Read, A. J. (2015). Field assessment of C-POD performance in detecting echolocation click trains of bottlenose dolphins (*Tursiops truncatus*). *Marine Mammal Science*, *31*(1), 169–190.

Tollit, D. J., & Redden, A. M. (2013). *Passive acoustic monitoring of cetacean activity patterns and movements in Minas Passage: pre-turbine baseline conditions (2011-2012)* (Issue July).

Wilson, B., Benjamins, S., & Elliott, J. (2013). Using drifting passive echolocation loggers to study harbour porpoises in tidal-stream habitats. *Endangered Species Research*, *22*(2), 125–143. https://doi.org/10.3354/esr00538

Appendix – SMRU Final Report

Comparative PAM Assessment: Final Report
Prepared for FORCE

June 2020

Comparative PAM Assessment

23 June 2020

Prepared by SMRU Consulting NA

Authors:

Jason Wood, PhD
Senior Research Scientist

Jessica Thompson
Junior Scientist

Kaitlin Palmer, PhD
Senior Research Scientist

Dominic Tollit, PhD
Senior Research Scientist

Executive Summary

In support of the Pathways Program, an assessment of porpoise passive acoustic detectors was undertaken. Five acoustic devices (CPOD, FPOD, AMAR, SoundTrap and icListen) were attached to a lander and deployed at the FORCE site. Fifty-two drifts of an icTalk past the lander were conducted in July and September 2019. During the drifts, the icTalk projected porpoise like clicks (pseudo clicks). The timing of the pseudo clicks was manually annotated in the AMAR dataset (N = 11,678) so that the acoustic detectors on all five devices could be compared to a 'true' set of pseudo clicks. This required a high level of time synchronization between the acoustic units. Although time synchronization was achieved, the CPOD and FPOD systems did not classify these pseudo clicks as porpoise clicks trains as the pseudo clicks were not sufficiently like real porpoise click trains. Consequently, comparisons between the CPOD/FPOD and the other devices could only be done with a limited number of opportunistically detected real porpoise clicks. We manually annotated 281 porpoise click trains, but only 60 of these could be used for comparison between units due to lack of CPOD/FPOD data from the first deployment.

Our key findings are as follows.
1) All three wav-based devices (AMAR, SoundTrap, icListen) successfully detected pseudo clicks with median detection ranges ~40 m. This short range is due to the low source level of the projector (~130 dB re 1μPa at 1 m). With real porpoise clicks (source level 160 dB re 1μPa at 1 m) directed at a device, this median detection range is predicted to be ~340 m, based on simple sound propagation and very sensitive detector settings.
2) To compensate for the low source levels of pseudo clicks, the sensitivity setting of the detectors were high. This high sensitivity resulted in many false positive detections.
3) No pseudo clicks were detected beyond 133 m.
4) True detections of pseudo clicks decreased with increasing current velocity with few detections above ~2 m/s.
5) False positive detections of pseudo clicks decreased with increased current velocity. There were few false positives above ~2 m/s. Many false positives are thought to result from sediment transport of small (~0.25mm diameter) particles.
6) The icListen seems to be less impacted by ambient noise at high frequencies which allows this device to detect more pseudo clicks than the other wav-based devices. However, we do not know if this is inherent to the device, or due to where it was located on the lander.
7) There was high variability in detection rates of pseudo clicks across the two deployments showing that how and where units are deployed can have a large influence on detection probabilities.
8) At the scale of detection positive minutes and using real porpoise clicks, all five devices were able to detect the majority of minutes when porpoise were present, however, due to the sensitivity of the detectors on the wav-based devices, their false positive rates were two orders of magnitude higher than the CPOD and PFOD. This creates additional data post-processing steps (e.g. classification and removal of false positives), often requiring human validation.

9) At lower sensitivity, the wav-based devices are still able to detect most minutes with porpoise and report fewer false positives. This suggests that further efforts on classification of detections can bring false positive rates down, while keeping recall high.

Based on the above, we provide the following recommendations.

1) The choice of PAM device needs to be based on the scientific questions being asked. A PAM unit suited for long term mid-field monitoring will likely be different than one needed to monitor near field evasion of harbor porpoise. If there is an additional need to collect information on ambient noise levels or detect vocalizations of cetaceans that do not echolocate or to localize the location of a porpoise, then the use of wav-based devices is required.

2) The complexity and difficulty of maintaining and deploying PAM systems in these challenging tidal environments should not be underestimated. Of key importance for long term studies is consistency in the data quality over time so that different time periods can be confidently compared, while recognizing inherent limitations and running costs. How devices are deployed is probably of greater importance than which of theses five PAM devices is used.

3) All five of the devices evaluated have different pros and cons in terms of ability to detect porpoise accurately, cost of devices and deployment, deployment duration, amount of post processing needed, etc. These all need to be considered in the decision on which device to use.

4) Using unrealistically sensitive detection settings, the wav-based devices detected more real porpoise than the CPOD and FPOD, but at the cost of higher post-processing needs, which currently require extensive human-based validation. At a lower detection sensitivity and with a very good click classifier, these wav-based devices could be used at the FORCE site for environmental effects monitoring.

5) At the current time, without the development and testing of bespoke click classifiers, the use of any of the three wav-based devices will result in extensive post-processing costs. Machine learning is rapidly improving but this post-processing is often labour intensive and can constrain project budgets. On very high sensitivity settings (not typically used) the wav-based devices do out-perform the CPOD and FPOD. Switching away from the current multi-unit CPOD-based environmental monitoring, based on these results, clearly requires further goal-based and cost-benefit considerations. A longer time series of real porpoise click monitoring is also advised.

Table of Contents

List of Figures

List of Tables

1   Introduction

The Pathway Program is a multi-year initiative to define, test, and validate an environmental effects monitoring solutions for the instream tidal energy industry in Canada. The main objectives of the Program are to i) define a DFO-approved Environmental Effects Monitoring (EEM) solution for the tidal energy industry, ii) apply machine learning to data analysis to reduce reporting time and compliance costs, iii) minimize initial capital costs to developers, iv) develop regional capability to manage, process, analyze and report EEM data, and v) develop intellectual property that regional companies can exploit commercially in multiple marine industries, both regionally and globally.

In support of the Pathways Program objective i), SMRU Consulting have been contracted to compare the harbor porpoise detection performance of multiple passive acoustic monitoring (PAM) systems. These systems include CPODs (Chelonia Ltd.) which have been used for porpoise monitoring at the FORCE site to date, as well as FPODs (a new variant of CPODs from Chelonia Ltd.), AMARs (Autonomous Acoustic Recorders made by JASCO Applied Sciences), SoundTraps (made by Ocean Instruments), and icListens (made by Ocean Sonics).

There are some obvious differences between these PAM devices. The Chelonia devices run an impulse detector in real time and only store limited information about those impulses that are detected, thus greatly reducing data storage needs. As with the wav-based devices, stored impulse information is later processed for the presence of animal click trains by the manufacturer's proprietary classification software. Wav-based devices, on the other hand, sample the acoustic environment many thousands of times per second and record the pressure at each sample. The pressure values are converted into bits and stored as audio (i.e. wav) files for later processing. Click detectors are then run on the stored audio for the presence of impulses. This process, like the click detection process on the CPOD/FPOD, occurs in the time domain. Following the impulse detection phase, customizable frequency-domain classifiers are run on the impulse detections to identify potential porpoise clicks.

The biggest difference between the CPOD/FPOD and the wav-based PAM devices is how the detected clicks are classified. CPOD/FPOD classifiers are largely based on time domain information, including click duration, amplitude, rate, etc. and the focus on the classifier is on identifying click trains, not individual clicks. In contrast, the software typically used to detect and classify the wav-based PAM data focuses on frequency domain information from individual clicks, not click trains. This makes for a fundamental challenge in comparing the performance of time domain and frequency domain click classifiers. CPOD/FPOD are focussed on click detection, while wav-based devices have a wider range of uses (including collecting ambient noise levels).

2 Methods

The FORCE team mounted the acoustic instrumentation on their lander (Figure 1) and deployed it over two periods: the 10-29th of July 2019 and the 5-13th of September 2019. As part of the ongoing FORCE Environmental Effects Monitoring Program, an additional CPOD was deployed at location D1 which was ~84 m from the lander. An acoustic doppler current profiler (ADCP) was used to measure current velocity and mounted on the CSTV turbine, which was also close to the location of the lander.

Twice during the lander deployments, the FORCE team drifted in the RV Nova Endeavor over the lander with icTalk transducers deployed over the side of the vessel. Drift experiments occurred on July 11th and September 6th, 2019. The center frequency of the pseudo clicks projected by the icTalk was 130 kHz and pseudo clicks were produced every 0.3 s at peak-to-peak source levels of 130 dB re1µPa. A handheld GPS (Garmin Oregon 600) onboard the vessel deploying the icTalk recorded the vessel's location.

Figure 1 PAM instrumentation mounted on the FORCE lander.

2.1   QA/QC

The data sent to SMRU Consulting from FORCE were assessed for overall data quality. Time stamps in the data were checked to ensure the correct time zone was used. The start and stop times of the data sets were also checked as were times when it was noted that the lander was deployed and recovered to ensure timing was not off on a large scale.

2.2   Click Annotation

Two click validated annotation data sets were created: 1) icTalk clicks (called 'pseudo' clicks in this report) and 2) real porpoise clicks. During icTalk drifts, the AMAR data were manually viewed in Audacity (acoustic software: https://www.audacityteam.org/) and every click generated by the icTalk that could be identified in the spectrogram was annotated such that the time of every 'true' pseudo click was then known. These annotations represented our validated dataset or, 'gold standard', against which we compared the pseudo click detection rates for the other instruments.

For real porpoise clicks, we used the PAMGuard click detector (threshold of 6 dB) with the default porpoise classifier on the AMAR data. The resulting detections were then validated by a trained operator to ensure only true positive clicks were retained. Through this process we obtained two validated data sets, one for the pseudo clicks and one for the real porpoise clicks.

2.3   Time Synchronization

One of the goals of this assessment is to compare click detections across these PAM devices using the annotated 'true' clicks from the AMAR. With the icTalk generating clicks every 0.3 seconds deciding which detection is a true positive (TP), a false positive (FP), or if the true click was missed (false negative: FN) required synchronizing the datasets to an accuracy better than ~20 milliseconds when comparing pseudo clicks. This was achieved by finding a unique click train which could be used to find the delay in time between the AMAR data and the other data sets.

2.4   Click Detectors and Classifiers

The two Chelonia devices (CPOD and FPOD) do not record raw audio data but run an impulse detector in real time and record only information about detected impulsive sounds. Theses detections were then run through Chelonia software to classify series of impulses (trains) into one of four categories: 'sonar', 'narrow-bandwidth high frequency', 'other cetacean', and 'unknown'. Narrow-bandwidth high frequency trains are likely to be porpoise. The classification algorithms for Chelonia instruments are not publicly available nor are users able to adjust sensitivity settings. All the other data sets were run through a PAMGuard click detector and classifier. PAMGuard is open source acoustic software that is used widely in passive acoustics and is maintained by the Sea Mammal Research Unit at the University of St. Andrews (https://www.pamguard.org/).

The ADCP current data; three-dimensional location of the lander and the icTalk during drifts (to calculate slant distances); click annotations; and the click detections from PAMGuard were imported into Matlab for further analyses using custom scripts.

By comparing the timing of automated classified detections with the timing of 'true' detections, the automated detections could be assigned as TP, FP or FN.

Where:
TP = number of true positives (i.e. a detection of a 'true' signal)
FP = number of false positives (i.e. a detection of something other than a signal of interest)
FN = number of false negatives (i.e. a signal that was not detected)

A standard way to then assess and compare auto-detectors is to calculate precision and recall. Precision (P) represents the proportion of detections that were true positive and is defined as follows:

$$P = \frac{TP}{TP + FP}$$

Recall (R) represents the proportion of true positive detections captured by the classification system and is defined as follows:

$$R = \frac{TP}{TP + FN}$$

Precision and recall values can be combined into a single F score that describes the overall performance of the classifier in the following way.

$$F = \frac{2 * P * R}{P + R}$$

The F score is a measure of the accuracy of the detector and varies from 0 to 1 with 1 being the best score. In the equation above, precision and recall are weighted equally; however, this can be changed to weight either precision or recall higher depending on the goals of the project. The full equation for the F score to do so is as follows.

$$F_\beta = (1 + \beta^2) * \frac{P * R}{(\beta^2 * P) + R}$$

Where β is the weight one attaches to recall vs. precision. A value of β = 2 means that recall is being weighted twice what precision is. This situation would be appropriate for studies where porpoise detections are rare and there is a need to record as many as possible. Conversely a value of β = 0.5 means that precision is weighted twice what recall is. This would be appropriate for longer term studies where manual annotation of detections is not possible and/or where porpoise detections are numerous. F scores were calculated with equal weighting for this report.

2.4.1 Pseudo Clicks

To evaluate whether each automated detection represented a true positive or false positive, we used a simple temporal criterion. If a detection happened within 0.02 seconds of a validated click, then the detection was considered a true positive. If not, the detection was considered a false positive. Likewise, when there was no detection within 0.02 seconds of a true click, a false negative was recorded. This provided detection metrics at the scale of individual clicks.

Precision and recall were also calculated on the minute and ten-minute scale. This is because our analyses of FORCE click data for the ongoing EEMP have used Detection Positive Minutes (DPM) or DPM10 (over 10 minutes) to deal with the autocorrection and zero inflation of the data. Both these values are standard output of the CPOD and FPOD systems used by numerous studies to investigate habitat use of porpoises. It was therefore considered useful to compare detection metrics at these temporal scales.

2.4.2 Porpoise Clicks

To evaluate whether automated detections represented TP or FP for real porpoise clicks, we used a similar temporal criterion. If a detection (of an individual click or click train) happened in the same minute as a true click, then a true positive was recorded. If not, the detection was considered false positive. False negatives were again recorded where no automated detection fell within the temporal criterion of a true click annotation. Through this process, it was possible to compare results from the time domain (CPOD/FPOD) and frequency domain (wav-based instruments) classifiers.

2.5 Effect of Ambient Noise on Detections

The probability of a detection occurring is affected by the signal to noise ratio (SNR). The lower the ratio, the less likely a detection will occur. The wav-based devices can measure ambient noise, therefore, we calculated ambient noise on each wav-based device to determine if noise levels on different devices might contribute to different detection results. We calculated 30 s average ambient noise levels for each device using 11 July and 6 September 2019 data across broadband (10 Hz – 200 kHz), and ~decade band (10 – 100 Hz, 100 – 1,000 Hz, 1 – 10 kHz, 10 – 200 kHz) Sound Pressure Levels (SPL) using Matlab code (FFT = Sample rate, 50% overlap, Welch's average method. Merchant et al. 2015). The sensitivity of devices was not adjusted across frequency but were assumed to be flat from the reported device sensitivity at 250 Hz.

3   Results

3.1   QA/QC

An error caused data loss for Deployment 1 CPOD/FPOD lander data, but data were successfully collected on other devices (Table 1). Data were collected on all six devices during Deployment 2 (Table 2). When data were available, they passed initial QA/QC controls.

Table 1 Data coverage and time zones for each instrument during Deployment 1.

Instrument	Start	End	Time Zone in file name
AMAR	6/4/2019 17:48	7/29/2019 21:27	UTC
SoundTrap	7/11/2019 2:00	7/14/2019 19:38	ADT
icListen	7/11/2019 5:00	7/12/2019 14:56	UTC
CPOD	7/9/2019 19:19	7/10/2019 9:37	No Data
FPOD	6/4/2019 12:06	6/4/2019 19:54	No Data
CPOD-D1	5/3/2019 16:03	8/14/2019 16:23	UTC

Table 2 Data coverage and time zone for each instrument during Deployment 2.

Instrument	Start	End	Time Zone in file name
AMAR	9/5/2019 13:48	9/13/2019 10:12	UTC
SoundTrap	9/5/2019 20:00	9/13/2019 9:53	ADT
icListen	9/5/2019 23:00	9/7/2019 8:35	UTC
CPOD	9/5/2019 10:01	9/13/2019 17:34	UTC
FPOD	9/5/2019 10:34	9/16/2019 12:28	UTC
CPOD-D1	8/14/2019 19:36	12/13/2019 16:32	UTC

3.2   Click Annotation

3.2.1   Pseudo Clicks

A total of 11,678 pseudo clicks were manually annotated across the 52 icTalk drifts that occurred in July and September (Table 3). There were several drifts when pseudo clicks were not detected on the AMAR. These were likely due to the low source level of the icTalk, noise from high current velocity, boat noise, and/or the icTalk not passing very close to the lander.

Table 3 Number of pseudo clicks annotated and used as a 'gold standard'.

Deployment	Click annotations	Drifts	Drifts with annotations
1	4,785	24	12
2	6,893	28	22

3.2.2   Porpoise Clicks

A total of 17,826 individual porpoise clicks in 281 click trains covering 85 minutes were manually annotated in the July and September datasets (Table 4). Although more data were analyzed from September, there were more clicks detected in July, as expected by earlier EEMP findings (Tollit et al. 2019).

Table 4 Number of porpoise clicks annotated and used as a 'gold standard'.

Deployment	Minutes of Data annotated	Porpoise Click Trains	Minutes with Porpoise
1	2,038	221	75
2	11,461	60	10

3.3   Time Synchronization

Time synchronization was achieved by locating series of clicks in the July and September datasets with unique temporal and amplitude patterns that could be matched across recording devices. These were either 70 kHz depth sounder pings or real porpoise echolocation clicks. The precision of this time synchronization between the devices is thought to be < 10 msec. All units were then synchronized to the timing on the AMAR as that dataset was used for our annotations, except for the CPOD located at D1. At ~84 m distance from the lander, the D1 CPOD was simply too far to have recorded the same click trains or depth sounder pings. We therefore had to assume the time stamps from that device were accurate.

3.4   Click Detector and Classifier

Due to the low source level of the icTalk, the click detector threshold had to be set at 6dB, a very low threshold for clicks, for PAMGuard to detect any pseudo clicks. This resulted in thousands of clicks being detected per minute (there are lots of impulsive click like sounds underwater). The standard PAMGuard harbor porpoise click classifier was therefore used to remove all clicks except those classified as harbor porpoise. There were also many clicks classified as porpoise clicks that were clearly echoes of the initial click (i.e. were lower amplitude than the first click and too soon to be the next pseudo click). Therefore, when click detections were less than 15 msec apart, the lower amplitude click in the pair was discarded.

3.5   Detection Assessment Using Pseudo Clicks at the Click Scale

Of the total 11,678 pseudo clicks annotated, the icListen had the highest recall (62%; Table 5). CPOD and FPOD results are not presented here as they are not directly comparable.

Because there was a large variance in performance of detectors between the two deployments, the detector results are provided separately for July (Table 6) and September (Table 7). Recall was much higher in the July deployment across all three devices. But the September deployment had much higher Precision for only the AMAR and SoundTrap. We believe these might be related to differences

in deployment orientation and location or possibly different sediment loads during these to time periods.

Table 5 Results of detector metrics at the click scale for July and September combined.

Hardware	Threshold (dB)	Deployment	Annotated clicks	TP	FP	FN	Precision	Recall	F-score
AMAR	6	1&2	11,678	3,689	25,025	7989	0.13	0.32	0.18
SoundTrap	6	1&2	11,678	3,431	152,806	8247	0.02	0.29	0.04
icListen	6	1&2	11,678	7,256	39,800	4422	0.15	0.62	0.25

Table 6 Results of detector metrics at the click scale for July only.

Hardware	Threshold (dB)	Deployment	Annotated clicks	TP	FP	FN	Precision	Recall	F-score
AMAR	6	1	4,785	1,909	22,541	2,876	0.08	0.40	0.13
SoundTrap	6	1	4,785	1,855	139,866	2,930	0.01	0.39	0.03
icListen	6	1	4,785	4,178	23,042	607	0.15	0.87	0.26

Table 7 Results of detector metrics at the click scale for September only.

Hardware	Threshold (dB)	Deployment	Annotated clicks	TP	FP	FN	Precision	Recall	F-score
AMAR	6	2	6,893	1,780	2,484	5,113	0.42	0.26	0.32
SoundTrap	6	2	6,893	1,576	12,940	5,317	0.11	0.23	0.15
icListen	6	2	6,893	3,078	16,758	3,815	0.16	0.45	0.23

3.5.1 Examples of Click Detections

Plots for each drift were made to contextualize the detection metrics. Figure 2 represents a drift in which very few automated detections were made and Figure 3 shows a drift where detections were numerous. These plots show several clear patterns. 1) Detections seem to occur more as the icTalk is drifting away from the lander, than towards it. This is visible by comparing the second panel in each figure (range from the icTalk to the lander) to detection records for each instrument (bottom three panels) 2) The SoundTrap and icListen documented more false positive detections outside of the annotated click times than the AMAR. 3) Conversely, the AMAR has fewer overall detections (including of TP) than either the SoundTrap or the icListen. Last, close inspection of a 30 s period around the point of closest approach in drift 18 indicates that true positive detections were more likely at current velocities < 1.7 m/s (Figure 4).

Figure 2 Data collected for a single drift experiment including current speed (top panel), range from the icTalk to the lander (second panel), time of icTalk pseudo-click projections (third panel) and time of PAMguard click detections as recorded on each of the three hydrophone systems (AMAR, SoundTrap, and icListen).
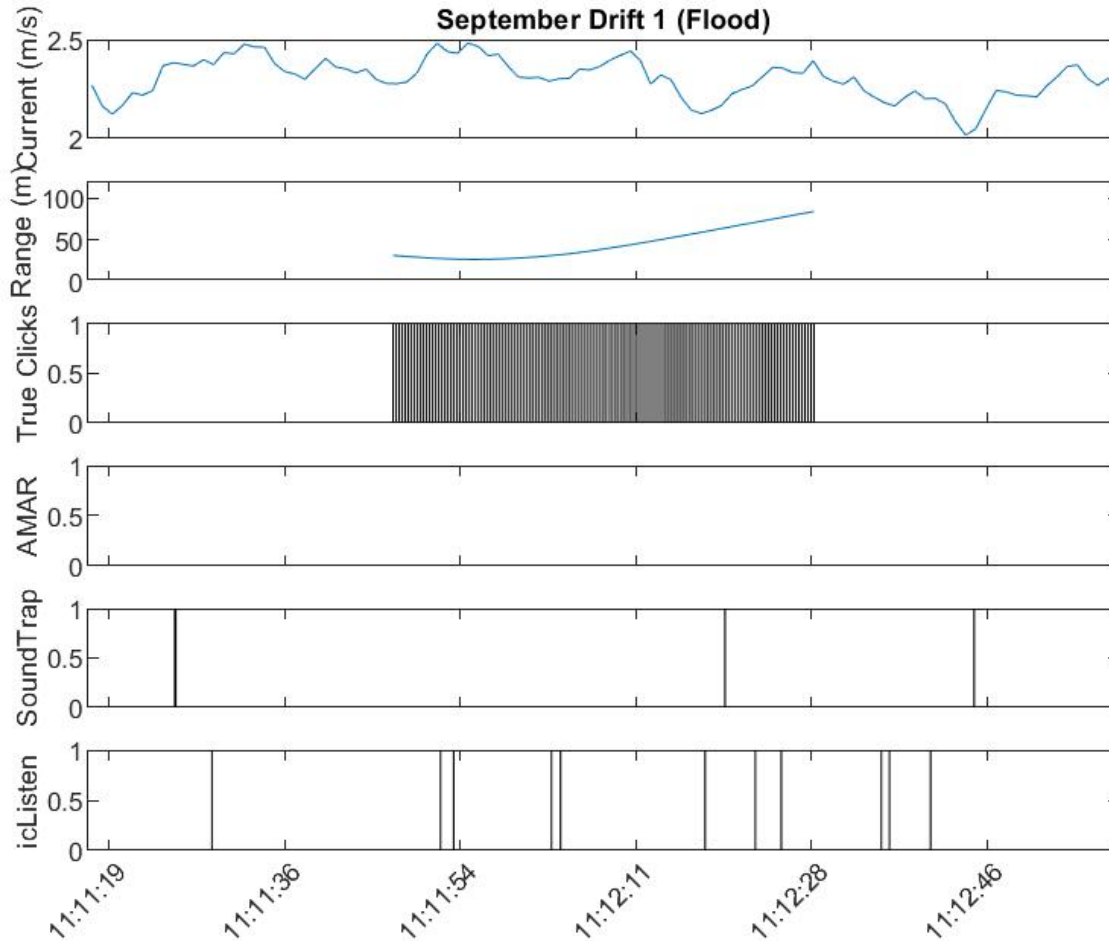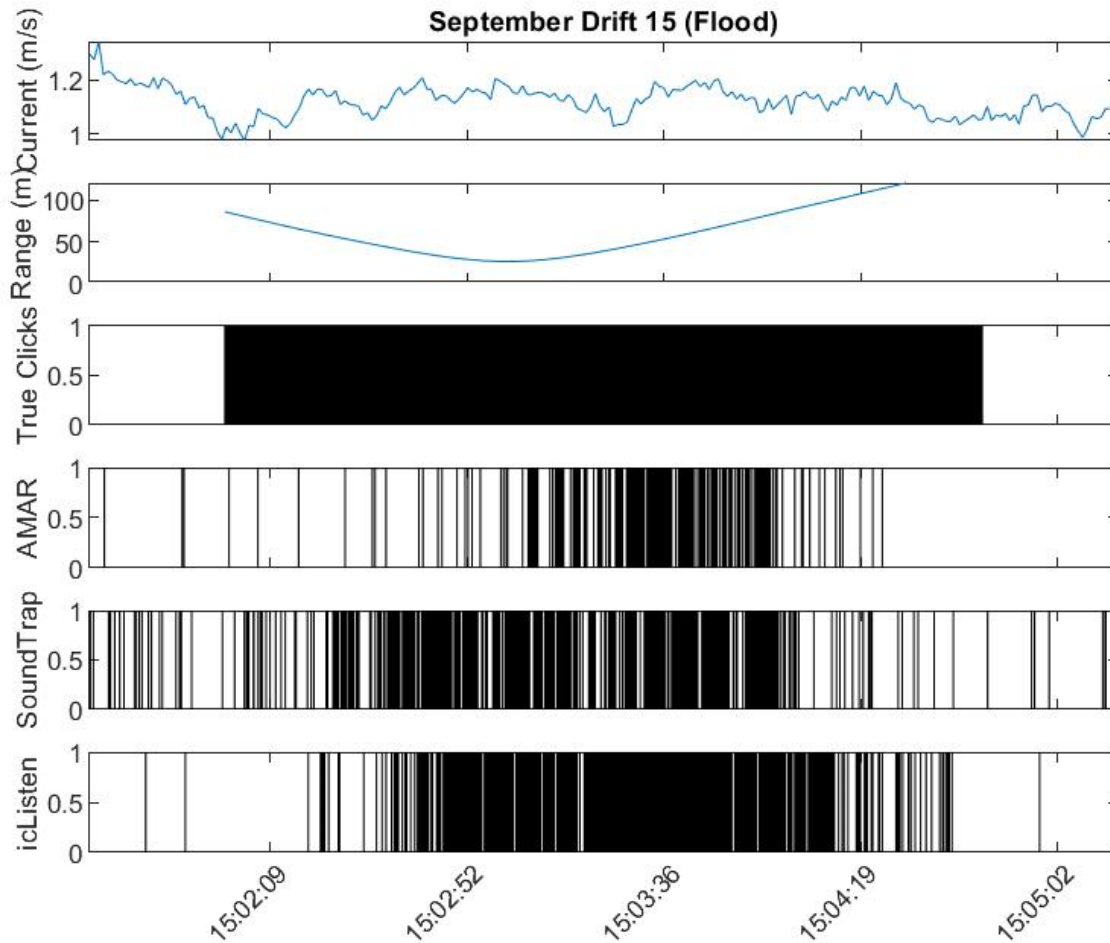
Figure 3 Data collected for a single drift experiment including current speed (top panel), range from the icTalk to the lander (second panel), time of icTalk pseudo-click projections (third panel) and time of PAMguard click detections as recorded on each of the three hydrophone systems (AMAR, SoundTrap, and icListen).
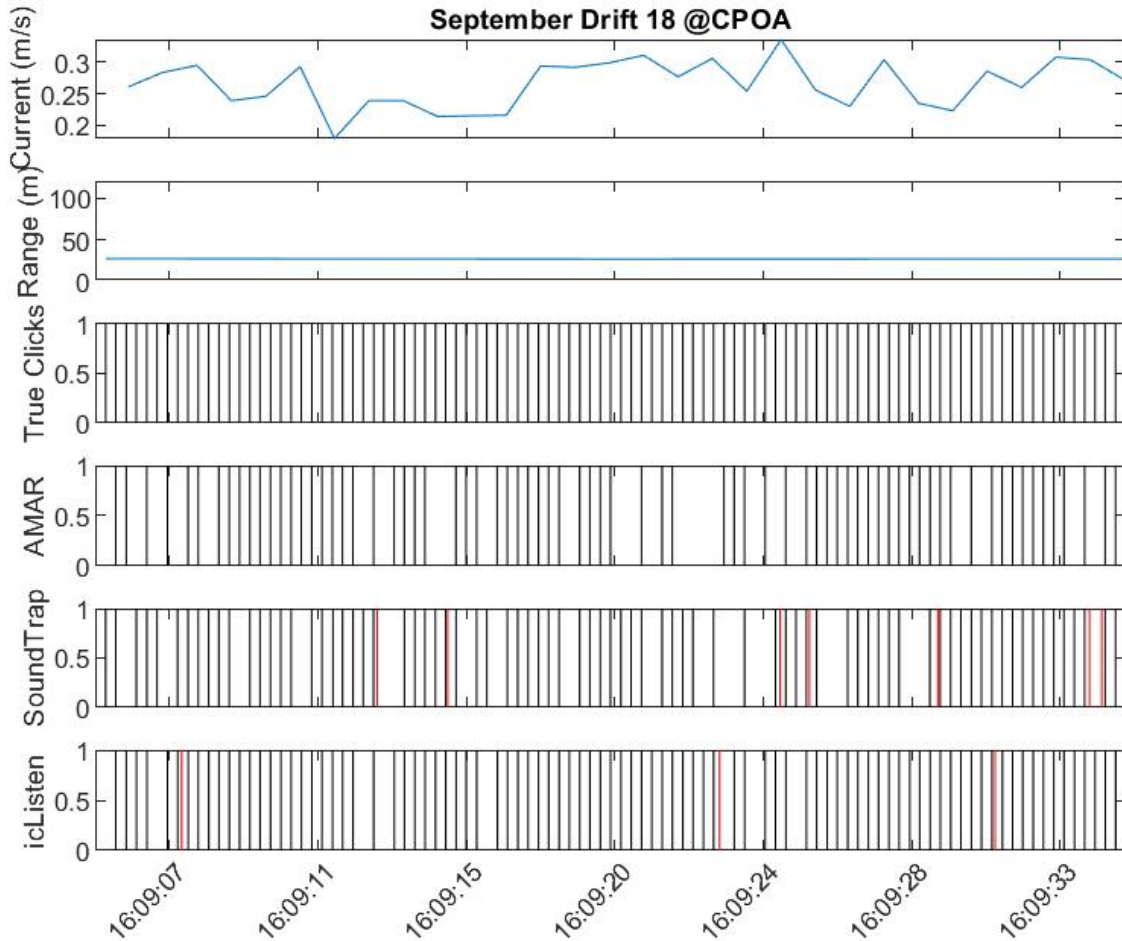
Figure 4 30 second subset of data from the closest point of approach during drift 18 including current speed (top panel), range from the icTalk to the lander (second panel), time of icTalk pseudo-click projections (third panel) and time of PAMguard detections as recorded on each of the three hydrophone systems (AMAR, SoundTrap, and icListen). Here false positive detections are shown in red and true positive detections are shown in black.

3.5.2 Effect of Distance and Current Velocity on Detections

Using data from July and September, the human annotated data had the largest median detection range of ~49 m (Figure 5). The median auto detector range varied from 39 to 44 m. These detection ranges are smaller than would be expected for porpoise clicks due to the low source level of the icTalk. If we assume a porpoise click source level of 160 dB re 1µPa at 1 m, a transmission loss coefficient of 20 dB and absorption of 38 dB/km (molecular absorption at 130 kHz), the median range of (a very sensitive) auto detection for a porpoise click directed straight at a monitoring device would be ~340 m.

As with increasing distance, increasing current velocity also reduces detection rates. Numbers of true positive detections are shown in Figure 6 as a function of both current velocity and distance between the icTalk and the sensor. Not every combination of detection distance and current velocity had data collected due to limited field time, so the outline (i.e. white areas) in Figure 6 is not indicative of trends in the data. Rather, looking within a single row (e.g. range bin) or a single column (e.g. current velocity bin), a general pattern of decreasing detections with increasing range and current velocity can be seen. Beyond ~2 m/s detection rates drop considerably. A major caveat in these conclusions is that we do not know when all pseudo clicks were produced and subsequently at what range and current velocity they were produced at. We were therefore unable to normalize the detection counts in Figure 6 by the number of pseudo clicks produced in each combination of range and current velocity. Consequently, some of the pattern in Figure 6 may be driven by the sampling regime, rather than measured detection rates.
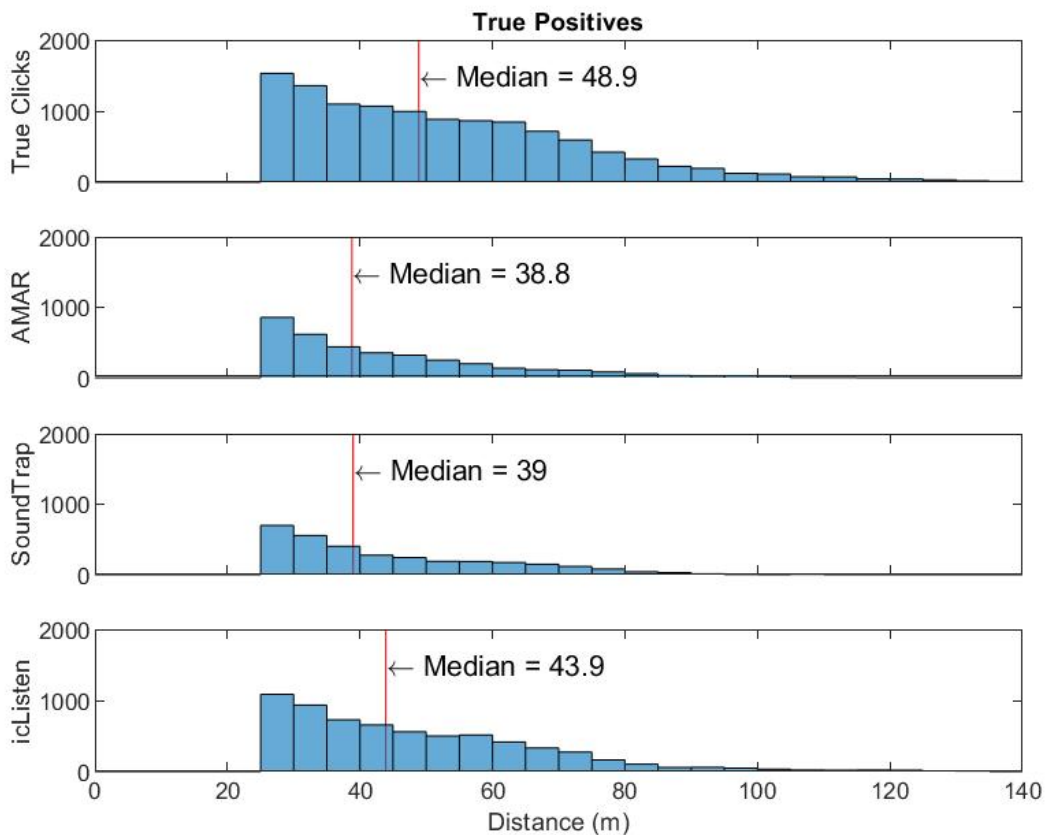

Figure 5 Histograms (counts) of pseudo click detection distance for: Human annotations (i.e. 'true clicks', first plot at top), AMAR autodetections (second plot), SoundTrap autodetections (third plot) and icListen autodetections (fourth plot).
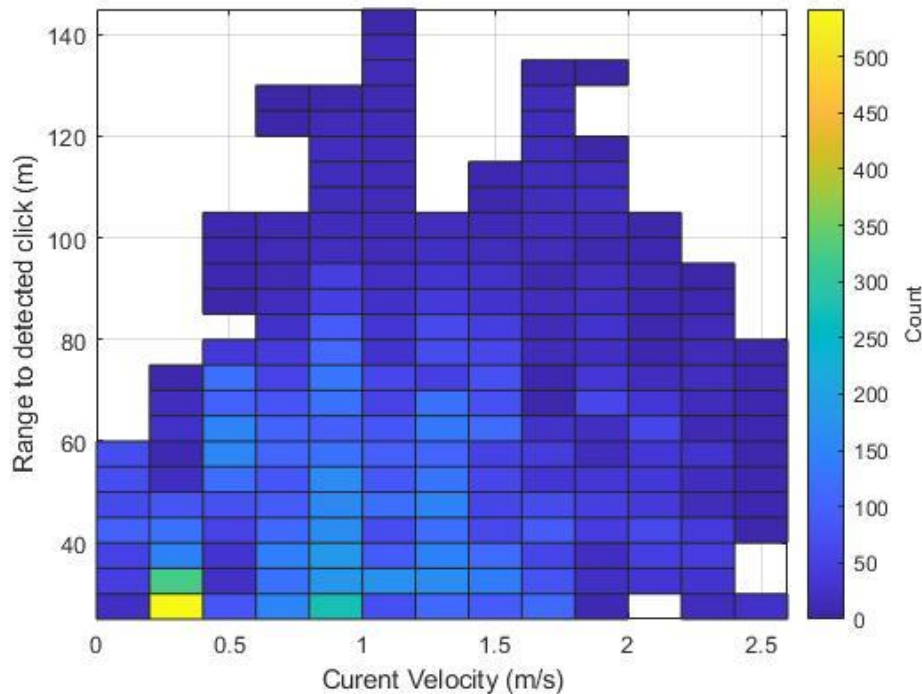
Figure 6 Two-dimensional histogram (counts) of hand annotated pseudo clicks by detection distance and current velocity. The color bar provides the scale of counts in each combination of range and current velocity.

3.5.3 Effect of Ambient Noise on Detections

There were three clear time periods with increased SPL (Figure 7 and Figure 8). These corresponded to periods with high current velocity. Also evident in both the July and September plots are short term increases in SPL (~15:00 to 21:00 on 11 July; ~11:00 to 19:00 on 6 September). These were due to the boat noise associated with the fieldwork for this project. The 100-1,000 Hz frequency band agrees the most for these three devices because the sensitivity of these systems at 250 Hz was used for calibration. In the highest two frequency bands, the icListen consistently reports lower SPL compared to the AMAR and SoundTrap devices. Given this apparent lower sensitivity to noise in the icListen at these high frequencies, this would explain the higher number of TP in the icListen (i.e. because the noise is lower, the signal to noise ratio ends up higher and detection probability is better). With similar SPL levels in these high frequencies, the AMAR and SoundTrap report lower TP but TP levels that are consistent between these two devices (Table 5).

If the high frequency sensitivity to noise of the hydrophones explains differences in TP rates, what is driving the FP rates, and specifically why is there such a large difference in FP rates between the July and September deployment for the SoundTrap (Table 6 and Table 7)? In high flow environments like the FORCE site, sediment transfer can cause a significant number of click-like sounds. We therefore plotted the FP rate against current velocity (Figure 9). FP rates are concentrated at current velocities < 2 m/s. At lower current velocities, only small particles are lifted into suspension and moved, while at higher current velocities, larger particles are moved into suspension. The frequency of the click like sounds is related to the diameter of the particle size. Sand with a diameter of 0.25 mm will produce

clicks at 130 kHz. These are likely what is causing many of the FP rates at lower current velocities seen in Figure 9. As current velocity increases, larger particles are brought into suspension which create lower frequency, but higher amplitude click like sounds. These swamp the clicks-like sound from the finer grained materials, hence the pattern seen in Figure 9.

There is also an asymmetry in sediment flow in tidal channels and the complexity of the lander structure could have shielded the instruments from some degree of sedimentation transport noise. The SoundTrap was deployed on the lower end of the lander (Figure 1). If the lander was oriented differently during the July and September deployments, the SoundTrap may have been more exposed to sediment noise. Likewise, the lower reported noise from the icListen could have been due to the shape of the hydrophone, instrument sensitivities, and/or where/how it was fastened to the lander.
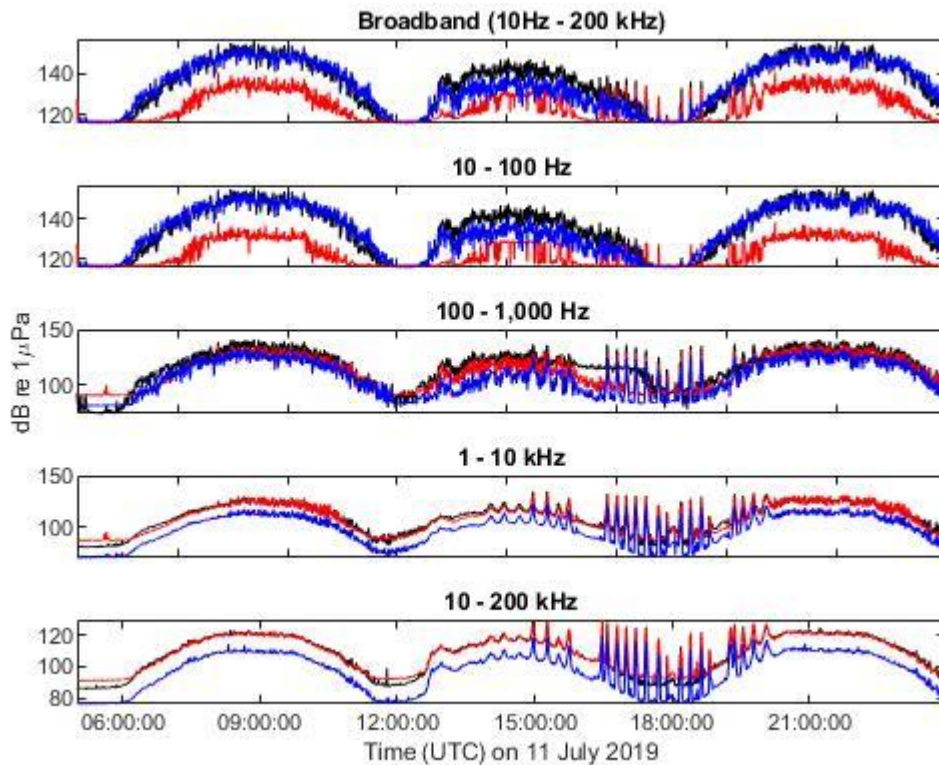

Figure 7 Broadband and decade band Sound Pressure Levels during 11 July 2019. Black trace is AMAR. Red trace is SoundTrap. Blue trace is icListen.
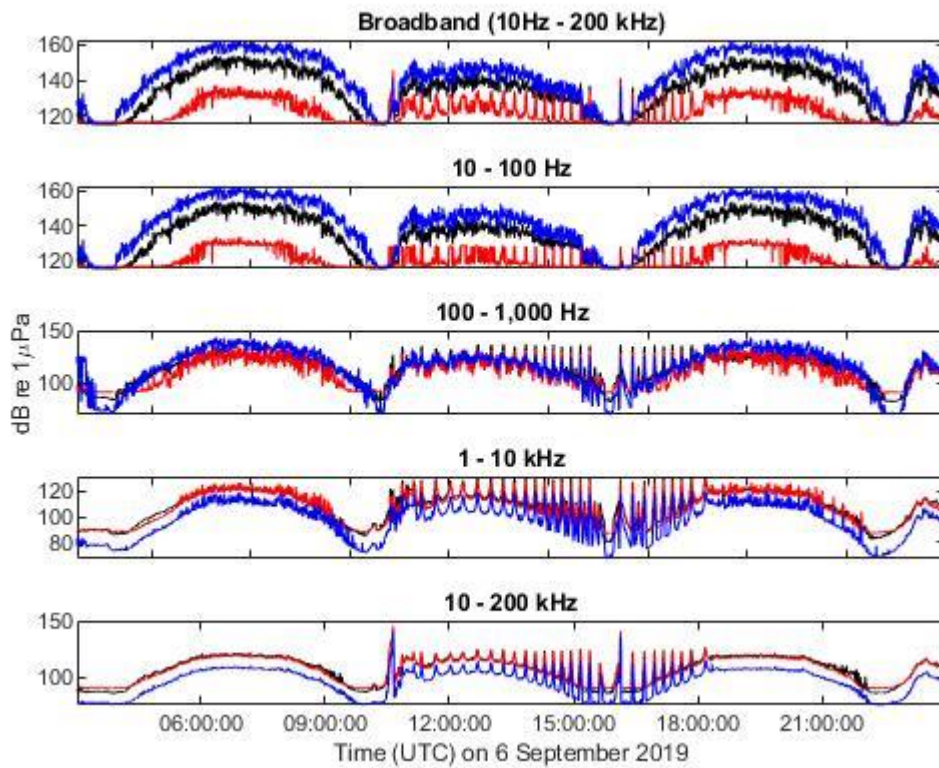
Figure 8 Broadband and decade band Sound Pressure Levels during 6 September 2019. Black trace is AMAR. Red trace is SoundTrap. Blue trace is icListen.
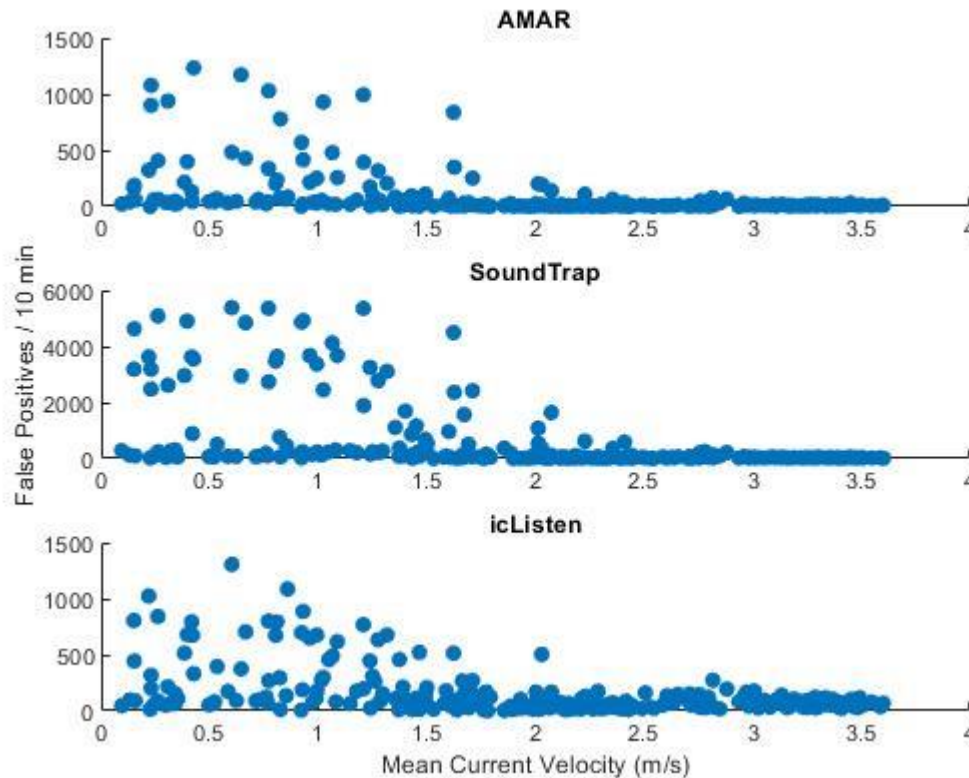
Figure 9 Scatter plots of FP rate / 10 min versus mean current velocity (over the same 10 min). Plots split by device. AMAR (top), SoundTrap (middle), icListen (bottom).

3.6 Detection Assessment Using Pseudo Clicks at the Minute Scale

Of the 92 one-minute periods with pseudo clicks, the icListen detected 89 of these with 97% recall but the documented highest number of FP detections (Table 8). This resulted in the AMAR having a slightly higher F score.

Table 8 Results of detector metrics at the minute scale using pseudo clicks.

Hardware	Threshold (dB)	Deployment	Annotated DPM	TP	FP	FN	Precision	Recall	F-score
AMAR	6	1&2	92	72	1567	20	0.04	0.78	0.08
SoundTrap	6	1&2	92	86	2222	6	0.04	0.93	0.07
icListen	6	1&2	92	89	2307	3	0.04	0.97	0.07

3.7 Detection Assessment at the 10 Minute Scale

Of the 42 ten-minute periods with pseudo clicks, the icListen and SoundTrap detected all of them (Recall of 100%) and the AMAR only missed 2 (Table 9). The Precision on all units was still low due to high false positive rates.

Table 9 Results of detector metrics at the 10-minute scale using pseudo clicks.

Hardware	Threshold (dB)	Deployment	Annotated DPM10	TP	FP	FN	Precision	Recall	F-score
AMAR	6	1&2	42	40	237	2	0.14	0.95	0.25
SoundTrap	6	1&2	42	42	215	0	0.16	1.00	0.28
icListen	6	1&2	42	42	215	0	0.16	1.00	0.28

3.8 Detection Assessment with Porpoise Clicks

Due to the nature of the pseudo clicks; it was not possible to properly incorporate CPOD/FPOD data into pseudo click assessments. Unlike PAMGuard click classifier, Chelonia classifiers rely on click train assessments. The icTalk produced clicks with uniform intervals and, as such, the Chelonia classifier was unable to cluster pseudo-clicks into trains. Since only click trains are reported by the click train classifier, an apples-to-apples comparison of the Chelonia and wav-based hydrophone systems was not possible. Instead, we focused on real porpoise clicks that were opportunistically collected during this study.

As data from the CPOD and FPOD on the lander were only available during the September deployment, we focused our efforts there. Table 10 provides the resulting detection metrics. Most devices had the core 11,461 minutes of data used for this analysis except for the icListen which only had 2,017 minutes of data due to limits on memory capacity. During the 11,461 minutes of data (~8 days) there were only 10 minutes with porpoise detections. The ~1.4 days of data used for the icListen had only 2 minutes with porpoise detections. Using the very sensitive 6 dB threshold in PAMGuard, all three wav-based devices had a recall of 100% but a precision of 0% due to the very large number of false positive detections. In contrast the CPOD and FPOD had lower recall but also much lower (2 orders of magnitude) FP rates. The CPOD located at D1 did not detect a single TP. This is to be expected given the distance from the lander to D1 and the need for the highly directional clicks from porpoise to line up between these two locations. In addition, the inability to time synchronize the CPOD at D1 may have also contributed to the lack of TP.

To compare the wav-based devices more closely with the CPOD/FPOD, we re-ran the AMAR data through PAMGuard with a more realistic and less sensitive detection threshold of 21 dB. In this case the AMAR had a recall of 80% but a FP rate much more in line with the CPOD/FPOD. This suggests that with more effort on classifiers for the wav-based devices, one may be able to achieve high recall while also minimizing FP.

Table 10 Results of detector metrics at the minute scale using porpoise clicks.

Hardware	Threshold (dB)	Deployment	Minutes of Data	Annotated DPM	TP	FP	FN	Precision	Recall	F-score
AMAR	6	2	11,461	10	10	5682	0	0.00	1.00	0.00
SoundTrap	6	2	11,461	10	10	9548	0	0.00	1.00	0.00
icListen	6	2	2,017	2	2	1871	0	0.00	1.00	0.00
CPOD	NA	2	11,461	10	6	62	4	0.09	0.60	0.15
FPOD	NA	2	11,461	10	4	72	6	0.05	0.40	0.09
CPOD-D1	NA	2	11,461	10	0	63	10	0.00	0.00	0.00
AMAR	21	2	11,461	10	8	616	2	0.01	0.80	0.03

4 Discussion

Passive acoustic monitoring for cetacean clicks in tidal sites is challenging. Obtaining meaningful results requires robust hardware capable of withstanding harsh conditions for long periods of time. Additionally, the large amounts of data in each deployment mean that automatic detection and classification algorithms are always required to identify porpoise echolocation signals. Beyond that, hardware and software need to be carefully matched to the needs of the scientific questions being asked, especially those by the regulatory agencies, if the project is industry related. There are also decisions needed that are based on cost and logistical constraints.

The focus of this project has been on assessing the relative efficacy of five PAM devices in detecting porpoise clicks at the FORCE site. There were inherent limitations in comparing these PAM devices due to the very different nature of how these devices work and our ability to produce porpoise like sounds (pseudo clicks). CPODs and FPODs use time domain classifiers that search for click trains. As a result, they were not 'fooled' by the pseudo clicks we generated. This meant we could not compare these devices directly with the wav-based devices using pseudo clicks, nor could we estimate their detection range.

Due to the low source level of the pseudo clicks, we had to run the PAMGuard auto detectors at a very low (i.e. very sensitive) threshold. Generally, there is a trade off in auto detectors. The more sensitive you make them, the more TP you capture, but that comes at the expense of high FP rates. This was the case when focused on pseudo clicks at the click scale. Using all the data available, the wav-based units were able to detect from ~2/3 to ~1/3 of the pseudo clicks but only from 2 to 15% of detections were actual pseudo clicks. At the time scale of 1 and 10 minutes, the Recall improved to > 78% but Precision remained low due to high FP rates.

The median detection range of the wav-based units using a very sensitive detector was ~40 m, which when extrapolated to the source level of a real porpoise click would equate to ~340 m median detection range for porpoise for these devices under the conditions experienced during these two deployments. Detection rate and detection range drop as current velocity increases, especially beyond 2 m/s. Interestingly, the FP rate also drops to low levels above current velocities of 2 m/s as larger particles are lifted into suspension which makes lower frequency but higher amplitude clicks.

The lower susceptibility of the icListen to high frequency noise seems to give it the ability to detect larger numbers of TP pseudo clicks. However, like the other wav-based devices with the detection and classification settings used (i.e. highly sensitive 6 dB), this also results in large numbers of FP. With these settings, from an ecological monitoring perspective, the wav-based devices would be unusable, unless a validated click classifier could be incorporated. Table 11 shows why. The detectors/classifiers are so sensitive for the wav-based devices that their FP rate is so high that porpoise would be estimated at this site from 50-93% of the time. Even the CPOD and FPOD, with much lower FP rates, likely overestimate porpoise presence, in this small dataset, by a factor of 5 or 6. At the level of FP rates presented in in the wav-based devices, the FP rates would completely swamp the real seasonal, diurnal and other patterns seen at this site (Tollit et al. 2019) and there would be no chance of detecting turbine effects on porpoise presence. The results for the AMAR data with a less sensitive

detector setting (21 dB) show that FP rates can be drastically reduced and still maintain decent recall. However, the FP rates will still need to drop further for environmental monitoring efforts as they still suggest porpoise presence 50 times higher than we think they are present.

Table 11 Ratio of FP to minutes of data (i.e. FP rate) and ratio of annotated DPM with porpoise to minutes of data. These are calculated from numbers presented in Table 10.

Hardware	Threshold (dB)	Ratio FP/Minutes of data	Actual Porpoise presence
AMAR	6	49.6%	0.1%
SoundTrap	6	83.3%	0.1%
icListen	6	92.8%	0.1%
CPOD	NA	0.5%	0.1%
FPOD	NA	0.6%	0.1%
AMAR	21	5.4%	0.1%

5   Conclusions

Each of the systems compared represents different costs, accuracy, precision, recall, and processing power. CPODs are popular for their ease of use, low cost and on-board post-processing, making them accessible to more researchers. Their impressive storage and battery life also make them a good choice for long-term and large-scale studies (Palmer *et al.* 2019, Jaramillo-Legorreta *et al*. 2017). Conversely, the nature of the Chelonia tools mean that users are not able to 'tune' the instruments to their needs. The 'black-box' nature of the devices (hardware and software), combined with the difficulty in producing porpoise like pseudo-clicks on demand, makes it difficult to validate the devices and can therefor cause concerns about their use. CPOD/FPODs are not used to collect ambient noise level information and can not be used to detect whistle and moans from other cetaceans.

Audio recorders including the SoundTrap, icListen, and AMAR devices collect many orders of magnitude more data than Chelonia systems. This enables researchers to develop quality detection and classification algorithms capable of far better recall than any CPOD or FPOD. This also enables researchers to study ambient noise levels and monitor for species not represented in CPOD or FPOD data including baleen whales and large odontocetes. These benefits, however, come at the expense of increased data storage and processing needs and the need for a more skilled operator to accurately interpret the data (as well check and remove false positives). With these considerations in mind, users must ultimately work closely with regulators and scientists to identify appropriate and cost-effective solutions to ecological monitoring.

Based on the results of this project, we conclude that any of these 5 PAM devices could be used for monitoring porpoise presence in tidal sites. At their most basic, they are all acoustic pressure sensors and will all suffer from flow noise and sediment noise. While some design features may give any one device a slightly better detection probability, of greater importance is how and where these units are deployed in tidal sites. If it is decided that wav-based devices should be used, then there needs to be concerted effort to greatly reduce false positive rates while keeping true positive rates high.

6   References

Jaramillo-Legorreta, A, Cardenas-Hinojosa, G, Nieto-Garcia, E, Rojas-Bracho, L, Ver Hoef, J, Moore, J, Tregenza, N, Barlow, J, Gerrodette, T, Thomas, L and Taylor, B (2017) Passive acoustic monitoring of the decline of Mexico's critically endangered vaquita. Conservation Biology, 31(1), pp.183-191.

Tollit, D, Joy, R, Wood, J, Redden, AM, Booth, C, Boucher, T, Porskamp, P, and Oldreive, M (2019) Baseline presence of and effect of tidal turbine installations and operations on harbour porpoise in Minas Passage, Bay of Fundy, Canada. The Journal of Ocean Technology, 14, pp. 24-48.

Merchant, ND, Fristrup, KM, Johnson, MP, Tyack, PL, Witt, MJ, Blondel, P, and Parks, E (2015) Measuring acoustic habitats. Methods in Ecology and Evolution, 6, pp. 257-265.

Palmer, KJ, Brookes, KL, Davies, IM, Edwards, E, and Rendell, L. (2019) Habitat use of a coastal delphinid population investigated using passive acoustic monitoring. Aquatic Conservation: Marine and Freshwater Ecosystems, 29, pp. 254-270.