REVIEW ARTICLE

**Marine Mammal Science**

# A review of big data analysis methods for baleen whale passive acoustic monitoring

## Katie A. Kowarski[1,2] (iD)    |    Hilary Moors-Murphy[3]

[1]Biology Department, Dalhousie University, Halifax, Nova Scotia, Canada

[2]JASCO Applied Sciences (Canada) Ltd, Dartmouth, Nova Scotia, Canada

[3]Bedford Institute of Oceanography, Dartmouth, Nova Scotia, Canada

**Correspondence**
Katie Kowarski, 202–32 Troop Avenue, Dartmouth, Nova Scotia B3B 1Z1, Canada.
Email: katie.kowarski@gmail.com

## Abstract

Many organizations collect large passive acoustic monitoring (PAM) data sets that need to be efficiently and reliably analyzed. To determine appropriate methods for effective analysis of big PAM data sets, we undertook a literature review of baleen whale PAM analysis methods. Methodologies from 166 studies (published between 2000–2019) were summarized, and a detailed review was performed on the 94 studies that recorded more than 1,000 hr of acoustic data ("big data"). Analysis techniques for extracting baleen whale information from PAM data sets varied depending on the research observed. A spectrum of methodologies was used and ranged from manual analysis of all acoustic data by human experts to completely automated techniques with no manual validation. Based on this assessment, recommendations are provided to encourage robust research methods that are comparable across studies and sectors, achievable across research groups, and consistent with previous work. These include using automated techniques when possible to increase efficiency and repeatability, supplementing automation with manual review to calculate automated detector performance, and increasing consistency in terminology and presentation of results. This work can be used to facilitate discussion for minimum standards and best practices to be implemented in the field of marine mammal PAM.

## 1 | INTRODUCTION

Recent technological advances have resulted in the generation of vast quantities of data, mostly digital, known as "big data." At a time when almost any information can be captured, stored, and shared digitally, big data has emerged in marketing, business, law, public administration, and science (Sivarajah et al., 2017). The potential benefits of big data are evident (Sivarajah et al., 2017), but the sheer volume of information creates hurdles in all aspects of big data management, including storage, integration, quality control, and visualization (Lee & Yoon, 2017; Sivarajah et al., 2017). Here, we focus on the challenge of effectively analyzing big data sets (data mining) in passive acoustic monitoring (PAM).

PAM of wildlife is an increasing source of big data in the field of wildlife biology. Acoustically active species are surveyed using acoustic recorders to understand the occurrence, distribution, density, and behavior of animals over space and time. First applied to terrestrial species including birds (Slabbekoorn & Halfwerk, 2009; Thorpe, 1961; Wimmer et al., 2013), apes (Kalan et al., 2015; Lieberman, 1968), bats (Gould, 1975; Johnson et al., 2002), and frogs (Schmidt, 1968; Zelick et al., 1999), PAM technology is now utilized extensively in the marine realm (Mouy et al., 2018; Stirling et al., 1983; Zimmer, 2011). Marine PAM allows for the collection of big data over space and time with minimal cost and effort when compared to boat- and aerial-based visual survey methods (Wimmer et al., 2010).

Marine acoustic signals of interest vary greatly from single distinct vocalizations like the North Atlantic right whale (*Eubalaena glacialis*) upcall (a contact call used by all individuals in the population throughout their range; Parks et al., 2011) to complex acoustic events such as the codas of the sperm whale, *Physeter macrocephalus* (Weilgart & Whitehead, 1997; Whitehead & Rendell, 2004). Some signals are organized in a repeated manner and are considered songs such as the complex hierarchical songs of the humpback whale, *Megaptera novaeangliae* (Payne & McVay, 1971) or the relatively simple songs of the fin whale, *Balaenoptera physalus* (Delarue, 2008; Watkins, 1981). Many baleen whale species produce songs and also produce social and/or feeding calls that are categorized as "non-song" vocalizations.

Methods to analyze big data (in PAM and other fields) efficiently and reliably include manual (human) content analysis, automated (computational/algorithmic) content analysis, or a combination of the two. Manual techniques are typically considered the benchmark for analysis, but humans can be inconsistent, subjective, and their level of experience can impact results (Baumgartner et al., 2020; Knight et al., 2017; Leroy et al., 2018b). Manual analysis of complete data sets is also difficult or even unrealistic to apply due to the sheer quantity of big data (Lewis et al., 2013). The development of automated methods has greatly increased the efficiency of big data analysis. A range of automated detectors have been employed in PAM analysis, including matched filters, energy summation, image processing, neural networks, wavelets, spectrogram correlation, and machine learning (Dugan et al., 2010; Frouin-Mouy et al., 2017; Klinck et al., 2012; Mellinger, Stafford, et al. et al., 2007; Thomas et al., 2019). Some automated detectors are designed to identify discreet acoustic signals, while others detect the chorusing events of many singers by comparing the frequency band of interest to ambient to determine an index of occurrence. Automated techniques have yet to be perfected as they often lack the context-sensitivity of human analysts, rendering them insufficient for reliable information extraction without human oversight (Lewis et al., 2013; Sivarajah et al., 2017).

Lewis et al. (2013) argued that the most efficient and reliable method to analyze big data combines manual and automated methods. This common "human-in-the-loop" automation involves some manual analysis to validate or supplement automation, allowing performance of the automated method to be evaluated by comparing the outcome of the two methods. Performance metrics include accuracy, sensitivity, specificity, receiver operating curve (ROC) characteristics, precision (P), recall (R), f-measure or F-score (F), the number of positive predictions, the number of

false positives, and false negatives (Knight et al., 2017; Sokolova & Lapalme, 2009). The evaluation metrics utilized depend on the nature of the data and the question at hand, with multiple measures often used (Sokolova & Lapalme, 2009). Emerging and experienced researchers must determine the appropriate methods to employ for effective analysis of big PAM data sets.

The present study undertakes a literature review on the reported methods of analysis in baleen whale PAM studies. Other marine animals, such as odontocetes, seals, and fish, were excluded from the review to keep the literature reviewed to a manageable size. First, we describe how the amount of PAM data varies by platform (e.g., surface or near-surface buoys, over-the-side hydrophones, bottom-mounted recorders). Then, we explore how big PAM data analysis methods vary (in terms of when manual, automated, and combined techniques are employed) depending on how many species are targeted and the research question of interest. We consider the metrics used to describe automated signal detector performance and the terminology applied to methods. This work aims to summarize past efforts for baleen whale PAM analyses and present recommendations for ways forward to standardize how results of PAM studies are reported to encourage application of robust, comparable, and reliable methodologies.

## 2 | METHODS

We used Web of Science, EndNote, and Covidence to perform a systematic literature search for peer-reviewed publications on baleen whale PAM. Detailed methods on the literature search and subsequent screening are provided in Supplementary A. The initial search resulted in 1,557 studies. Through title, abstract, and full text screening, the review was ultimately reduced to 166 studies on the topic of baleen whale PAM that were published in the years 2000–2019, presented original research (were not review or overview articles), did not implement real-time or near real-time analysis or aim to localize animals, and were not entirely based on describing automated detectors (Table S1). Real-time and localization studies were excluded to keep the review to a manageable size, as these topics have specific methodological requirements and the literature on these topics is large.

Data from the 166 studies (Table S2) were collected in two stages. In stage one, the overall data analysis method, data volume (in hours recorded), and the data collection platform (e.g., bottom mounted, over-the-side, tag) were recorded for each article (Table S3). The overall data analysis method was categorized as full manual (no automated detector, manually reviewed all acoustic data collected), partial manual (no automated detector, manually reviewed a subset of the acoustic data collected), automated (employed one or more automatic detectors, did not manually review any data), or manual and automated (employed some combination of automated and manual analysis). In the second stage of the literature review, questions specific to big data were investigated, where we defined big data as studies with 1,000 hr or more of acoustic data. Big data articles were subjected to a methodology-based review that considered how much data were manually analyzed, how many baleen whale species were targeted, how data for manual analysis were selected, and what metrics were used to report automated detector performance (Table S3). Each study's main research question was categorized as one of the following: occurrence (the spatial, temporal, and/or diel presence of vocalizations); characterization (the description of vocalization characteristics that can vary from frequency-time characteristics of discrete non-song vocalizations to characterizing changing patterns over time and space of songs); a combination of occurrence and characterization; or, other (e.g., anthropogenic impacts).

Data extracted from the literature reviewed (in stage one and two) were summarized by data volume (seven categories ranged from 0–10 to 100,000–860,000 hr), number of species targeted (studies targeting 1–2 species or studies targeting all baleen whale species possibly acoustically present in the PAM data set), and the overarching research question (as defined above) to determine commonalities and differences in data analysis methods across studies. Similarities and differences among studies were examined to determine if there are general best practices commonly used and to develop recommendations for future studies. Some variables, such as data volume and percent of data manually analyzed, were not always explicitly provided and were calculated or estimated based on available information where possible.

## 3 | RESULTS

The systematic literature review resulted in 166 articles (from 110 different first authors) on the topic of PAM and baleen whales. Detailed tabular data by study can be found in Supplementary Results. With increasing data volume included in a study, we observed a shift in both the data collection platform and the data analysis method. Studies with relatively small amounts of acoustic data were associated with varied data collection platforms, including tags directly attached to animals (Videsen et al., 2017), drifting buoys (Širović et al., 2006), and even snorkelers (Zoidis et al., 2008) (Figure 1). In contrast, studies using big data predominantly collected data from autonomous bottom, or near-bottom, mounted systems (e.g., Burnham & Duffus, 2019; Soldevilla et al., 2014; Wright et al., 2019; Figure 1). With increasing data volume, the predominant data analysis method switched from full manual analysis to a combination of manual and automated analysis (Figure 2).

### 3.1 | Big PAM data

The following results are restricted to articles with 1,000 hr or more of data (94 articles) and are considered big PAM data. Studies of big PAM data sets for baleen whales covered a range of species and research questions. The majority (71 out of 94 papers) targeted blue (*Balaenoptera musculus*), humpback, fin, or right whales with proportionally fewer
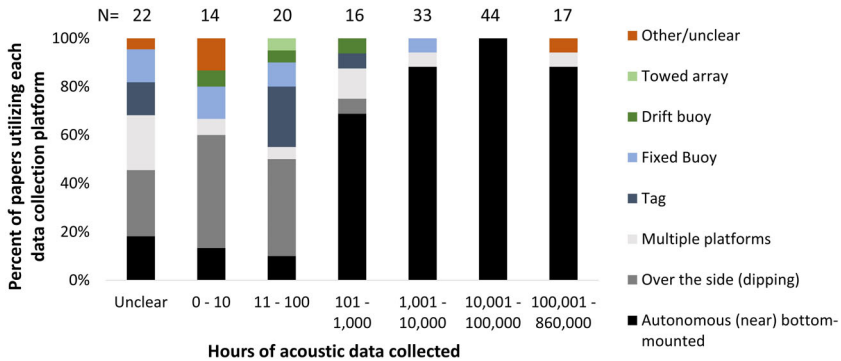


**FIGURE 1** Percent of baleen whale PAM papers (*N* = 166) utilizing each data collection platform for different data volumes (hours of acoustic data collected). Some papers had insufficient information to confidently determine data volume (unclear).



**FIGURE 2** Percent of baleen whale PAM papers (*N* = 166) utilizing each data analysis method for different data volumes (hours of acoustic data collected). Some papers had insufficient information to confidently determine data volume (unclear).
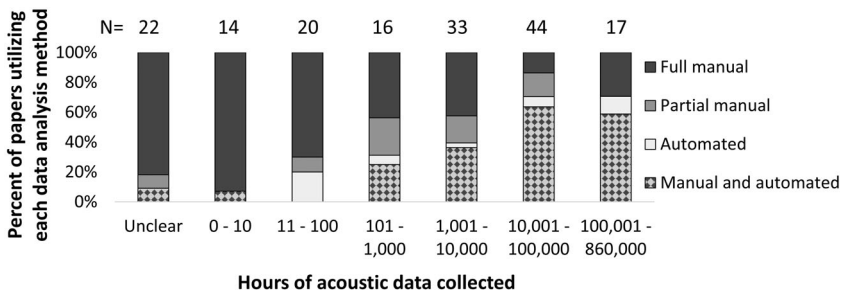
studies (24%) on bowhead (*Balaena mysticetus*), Bryde's (*Balaenoptera brydei*), gray (*Eschrichtius robustus*), or minke whales (*Balaenoptera acutorostrata*, *Balaenoptera bonaerensis*; Figure 3). Most research questions focused on vocalization characterization (16 articles), occurrence (46 articles), or a combination of the two (16 articles) and the frequency of questions varied across target species (Figures 3 and 4). Studies on humpback whales were more focused on vocalization characterization than studies on blue, right, and fin whales, which more commonly described occurrence (Figure 3). Six studies investigated less common research questions and were considered "other" (Figure 3). These studies investigated the impacts of anthropogenic activities on baleen whale acoustics (Castellote et al., 2012a; Cerchio et al., 2014; Melcon et al., 2012; Risch et al., 2012), the acoustic response of baleen whales to predation (Burnham & Duffus, 2019), and baleen whale density (Marques et al., 2011). Studies with "other" research questions were too varied to observe patterns and draw meaningful conclusions or recommendations; therefore, they will not be described further, limiting the total number of papers described below to 88 of the 94 big PAM data studies.

## 3.2 | Studies targeting 1–2 baleen whale species

Of the 88 articles that investigated characterization and/or occurrence, 75 targeted one species, 4 targeted two species (fin and blue whales), and 9 targeted all baleen whale species acoustically present in the acoustic recordings
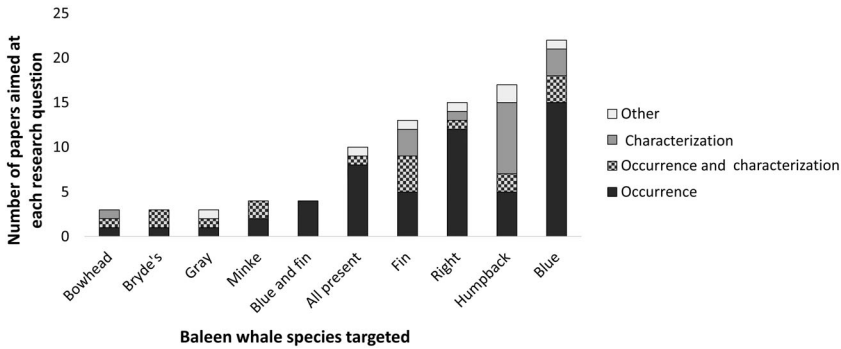


**FIGURE 3** Distribution big PAM data (1,000+ hr) baleen whale papers (*N* = 94) by each research question and species targeted where "all present" indicates that all acoustically active species in the acoustic data were targeted.
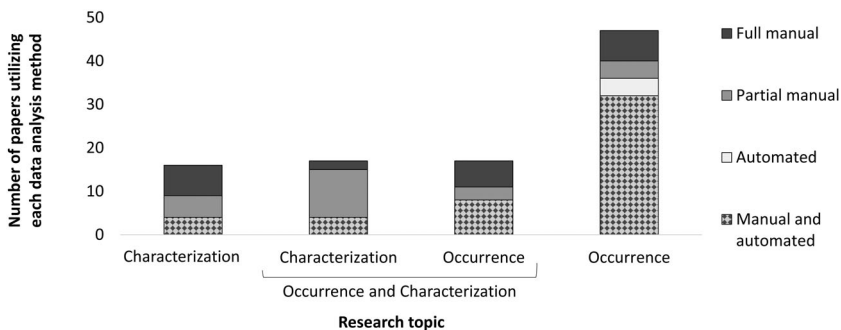


**FIGURE 4** Number of big PAM data baleen whale papers that used each data analysis technique for research For Peer Review questions involving characterization, occurrence, or both (*N* = 88). Papers investigating both occurrence and characterization were duplicated to show how each question was addressed separately within the same study.

(Figure 3). In this section we consider studies targeting 1 or 2 species (N = 79) and how data analysis methods varied across different research questions (Tables S5–S10). Studies were separated based on number of species because those looking at all species will face challenges (e.g., many different vocalization types requiring analysis of a potentially broader frequency range, more automated detectors, greater breadth of human expertise) beyond those faced by studies focused on 1 or 2 species.

## 3.3 | Characterization studies

Studies focused solely on characterization of vocalizations encompassed 19% (15/79) of the big PAM data papers reviewed that targeted one or two species (Figure 4). Ten of the papers focused on aspects of baleen whale song, including how songs vary in space, time, between populations, and in different behavioral contexts, while the remaining papers described characteristics of non-song vocalizations. Analysis methods included full manual analysis, partial manual analysis, and a combination of automated and manual analysis with specific analysis protocols varying across studies and the percent of data manually reviewed ranging from 3% to 100% (Table S5). The most common methodology for characterizing song was the full manual analysis of all acoustic data (Table S5). Most of these studies had data sets less than 2,000 hr in size (Cholewiak et al., 2018; Johnson et al., 2015; Mercado, 2018; Miksis-Olds et al., 2008; Miksis-Olds et al., 2018) though one study spanned 10 years and still seemed to have employed full manual review (Garland et al., 2015). Partial manual analysis was the most commonly employed methodology for the characterization of non-song vocalizations with protocols for selecting data varying across studies. For example, Fournet et al. (2018) selected 60 hr of data randomly while Jacobs et al. (2019) reviewed 4 of every 24 hr (Table S5).

## 3.4 | Occurrence studies

Understanding the occurrence of a vocal species was the most common research question for big PAM data set analysis, encompassing 59% (47/79) of the papers reviewed that targeted one or two species (Tables S6 to S8). The temporal breadth of occurrence studies ranged from 64 days to 10 years, with the majority (35/47) ranging from 1–4 years. All four analysis methods were used to determine occurrence: full manual (seven papers), partial manual (four papers), fully automated (four papers), and manual and automated (32 papers; Figure 4).

Manual analysis methods to determine species occurrence varied greatly across studies both for full manual and partial manual techniques (Table S6). For partial manual analysis studies, both the amount reviewed (10%–44% of data) and manner in which reviewed data were selected (randomly versus systematically) varied (Table S6). Similarly, full manual studies differed in how data were reviewed, with some authors interested in finding presence over some timeframe (e.g., manually reviewed every day until a vocalization was observed to determine daily presence), while others determined the occurrence of every individual vocalization (Table S6).

All studies that solely used automated techniques to determine species occurrence targeted fin and/or blue whales (Gavrilov et al., 2018; Leroy et al., 2016; Leroy et al., 2018a; Matsuo et al., 2013). For three out of four of the studies, automated detector performance was either not calculated or not reported. Gavrilov et al. (2018) reported both missed and false detection rates that had been previously calculated in earlier publications.

A combination of manual and automated analysis was the most common methodology applied to determine species occurrence (Figure 4). Factors that were inconsistent across studies included: how much data were manually analyzed, how files for manual analysis were selected, if automated detector performance was determined, and how automated detector performance was determined. These studies were broadly categorized into those that manually reviewed every automated detection and those that did not (Table S7). Only two studies could not be categorized, as they applied different analysis techniques to different portions of their data sets (Buchan et al., 2015; Kerosky et al., 2012).

Most manual and automated occurrence studies (21/32) completed some form of manual review of every automated detection or automated detection event (e.g., days with detections). Nine studies applied automated detectors and then manually checked all automated detections and removed false positives, but never determined how often their detector missed vocalizations (Mellinger, Nieukirk, et al., 2007; Morano, Rice, et al., 2012; Risch et al., 2014; Salisbury et al., 2016; Širović et al., 2009; Stafford et al., 2004, 2011; Stafford et al., 2005; Whitt et al., 2013). In addition to manually checking all detections and removing false positives, seven studies analyzed a subset of data without automated detections to determine the missed detection rate of their automated detector (Balcazar et al., 2015; Buchan et al., 2018; Davis et al., 2017; Hodge et al., 2015; Lammers et al., 2011; Mussoline et al., 2012; Tripovich et al., 2015). The amount of data reviewed and data selection protocol to determine missed detection rate varied greatly across these studies, with <1%–33% of data reviewed that were selected randomly, based on ambient sound levels, or selected systematically (e.g., every third day; Table S7).

Some studies that reviewed every automated detection undertook manual analysis independent of the automated detection results and did not use the analysis to determine detector performance. In addition to checking every detection, Bort et al. (2015) carried out a systematic partial manual analysis (reviewed every third day). Munger et al. (2008) carried out full manual analysis via long-term spectral averages where successive fast Fourier transforms are averaged into long-term spectrograms for review.

Ten occurrence studies that utilized both automated and manual techniques did so without checking every automated detection (see references in Table S7). Instead, they applied automated detectors and subsequently completed variable amounts of manual data review (<1%–7% or unclear to reader) to determine automated detector performance, optimize automated detectors, or create ROC curves (Table S7). The subset of acoustic data for manual review were randomly selected, spread over time, spread over locations, spread over number of detections, or not specified, depending on the study (Table S7).

The presentation of baleen whale occurrence results in big PAM data varied across studies, with results generally presented as one of two measures: number of vocalizations/detections or presence (Table S8). Half of studies (23/47) presented results as number of vocalizations or number of automated detections over some duration (e.g., number of automated detections per week; Table S8). In contrast, 18/47 studies presented results as presence within a timeframe, commonly over some larger timeframe, e.g., number of days with presence per week (Davis et al., 2017; Table S8). Risch et al. (2014), Morano, Rice et al. (2012), and Wright et al. (2018) presented results using both measures. A few studies presented results using completely different measures such as a daily index (Leroy, Samaran, et al., 2018; Nieukirk et al., 2012; Simon et al., 2010). The most common descriptors of occurrence were number of vocalizations (or automated detections) per month (nine studies) or per day (six studies) and number of hours (or percent of hours) with presence per day (four studies) or per month (three studies; Table S8).

Diel patterns were explored in 14 of the 47 occurrence papers with an additional two papers that commented on diel patterns in the occurrence and characterization papers. Most diel studies incorporated all days where vocalizations occurred while some limited diel analysis to a portion of the data (e.g., only months where the vocalizations were most common; Table S9). The number of light regimes in a 24-hr day to explore diel patterns varied across studies with the most common being three light regimes (light, dark, and twilight; four studies) or four light regimes (light, dark, dusk, dawn; three studies; Table S9). Light regimes were most frequently defined using nautical twilight determined by angle of the sun (11/15 studies; Table S9). The most common unit to compare vocalizations across light regimes was hourly mean adjusted number of vocalizations (or presence of vocalizations) per hr (7/15 studies; Table S9).

## 3.5 | Combined occurrence and characterization studies

A number of studies (22%; 17/79) explored the acoustic occurrence of baleen whale species and characterized some aspect(s) of the recorded vocalizations (Figure 4). While the overarching aim of the research was often focused on

either occurrence or characterization, both questions were addressed throughout each study. The characterization aspect of these studies largely targeted non-song vocalizations (12/17). Similar to when these research questions were explored independently, studies exploring both characterization and occurrence had a wide range of methodologies with few commonalities between them (Table S10). Some studies (7/17) undertook the same analysis methodology for both research questions (e.g., Leroy et al., 2017; Stafford et al., 2007), but the majority utilized two largely unrelated methodologies (Table S10). For the question of occurrence, the most common method was a combination of manual and automated detection (8/17), though specific protocols varied across studies. For example, Clark et al. (2002) manually validated 1 hr every 18 days of data, whereas Vu et al. (2012) manually reviewed all automated detections. For the question of characterization, the most common method was partial manual (11/17) with an emphasis on high signal-to-noise ratio (SNR) vocalizations, where what was considered "high" SNR varied across studies. Again, specific analysis protocols varied from a selection of 1 hr per day (Oswald et al., 2011) to a random selection of 6 days per month (Vu et al., 2012; Table S10). The percent of acoustic data manually reviewed was often unclear as it was guided by automated detectors, but values ranged from <1% to 100% to determine occurrence and from 1.9% to 11% for characterization analysis (Table S10).

## 3.6 | Studies targeting all species present

Of the 94 big PAM data baleen whale papers reviewed, only nine targeted all species present in the acoustic data and of those, only five described the occurrence of three or more species (Table S11; Figure 4). Multi-species studies used data ranging in timespan from 1 to 6 years with the majority (6/9 studies) spanning 1 to 2 years. Occurrence was investigated in every all-species study, with McDonald (2006) also investigating density and Nieukirk et al. (2004) also characterizing vocalizations. Methodologies to determine occurrence varied across studies, with the percent of data manually reviewed ranging from 5% to 100% (Table S11). The most commonly used method to investigate occurrence of all baleen whale species was a combination of manual and automated techniques (3/9 studies; Table S11) where protocols included manually reviewing a portion of each file (Hannay et al., 2013), reviewing 20 files per species with varying automated detector counts (Frouin-Mouy et al., 2017), or selecting data at random from within regions that were suspected to have unreliable automated detections (De Vreese et al., 2018).

## 3.7 | Describing automated detector performance

The description of automated detector performance in baleen whale big PAM studies was variable across the literature reviewed. Of the 53 studies that incorporated automated detectors, 25 made no commentary on automated detector performance. Three studies commented on detector performance (e.g., noted that automated detector(s) may have missed some vocalizations or performed poorly), but performance was not calculated, or, if it was, the results were not presented (Širović et al., 2004; Stafford et al., 2011, 2005). Five studies commented on how the automated detector(s) performed in previous studies but did not calculate performance for their data set (Table 1). Some aspect of automated detector performance was calculated and presented in only 20 of the 53 studies that used automated detectors (Table 1).

The majority of studies that incorporated automation (with or without manual review) investigated acoustic occurrence (Figure 4), with the exception of three that investigated characterization only (Delarue, 2008; Magnúsdóttir et al., 2015; Rekdahl et al., 2017). None of the characterization studies reported automated detector performance, which is unsurprising given that automated detector performance would not impact the reliability of their results as they simply used the automated detectors to find signals that were then manually characterized.

Different terminology was used across studies to describe the same automated detector performance metric (see performance metric definitions and formulas summarized in Figure 5); in some cases, the definition was assumed when equations were not provided. False negative detection errors, missed call rate, false negative percent,

**TABLE 1** Summary of metrics used to describe automated detector performance where "—" indicates the metric was not described. Metrics include false negative rate (FNR), false positive rate (FPR), precision (P), recall (R), and accuracy (ACC). This summary only includes the 25 studies that provided information on automated detector performance (28/53 studies did not and are not included here).

| Species | FNR (%) | FPR (%) | P (%) | R (%) | ACC (%) | Study |
|---|---|---|---|---|---|---|
| Blue | 1.0 | 1.0 | — | — | — | Gavrilov and McCauley (2013) |
| | — | — | 67.0–95.0 | 77.0–96.0 | | Širović et al. (2015) |
| | 0.0–9.4 | 17.6–100.0 | — | — | — | Balcazar et al. (2015) |
| | — | 14.6–36.6 | — | — | — | Tripovich et al. (2015) |
| | 6.0–33.0 | 3.0–13.0 | — | — | — | Buchan et al. (2015) |
| | ~1.0[a] | ~1.0[a] | — | — | — | Gavrilov et al. (2018) |
| | <0.002 | — | — | — | — | Buchan et al. (2018) |
| | <5.0[a] | <5.0[a] | — | — | — | Jolliffe et al. (2019) |
| Bowhead | — | — | 84.0 | 22.0 | — | Hannay et al. (2013) |
| Bryde's | 10.0 | 50.0 | — | — | — | Kerosky et al. (2012) |
| Fin | 21.0 | 20.0 | — | 79.0 | — | Morano, Salisbury, et al. (2012) |
| | — | 0.0 | — | 35.3 | — | Tsujii et al. (2016) |
| Humpback | 6.0–11.0 | — | — | — | — | Vu et al. (2012) |
| | — | — | — | 90.0 | — | Kowarski et al. (2018) |
| | — | 8.0 | — | 68.0 | 85.0 | Español Jiménez and van der Schaar (2018) |
| Minke | — | 5.0 | — | 22.0–100.0 | — | Oswald et al. (2011) |
| | 27.0 (0.0–51.0) | — | — | — | — | Risch et al. (2013) |
| | 27.0[a] | — | — | — | — | Risch et al. (2014) |
| | — | 11.0 | 20.0 | 74.0 | — | Risch et al. (2019) |
| Right | — | — | — | 15.0 | — | Mellinger et al. (2004) |
| | 7.0–19.0 | — | — | — | — | Mussoline et al. (2012) |
| | — | — | — | 75.0[a] | — | Morano, Rice, et al. (2012) |
| | 36.2–53.6 | — | — | 46.4–63.8 | — | Hodge et al. (2015) |
| | — | — | — | 75.0[a] | — | Salisbury et al. (2016) |
| | 31.0 | — | — | — | — | Davis et al. (2017) |

[a]Performance metric based on previous research.

misdetection rate, rate of missed detections, and missed detection rate were all assumed to refer to FNR. Rate of false detections, false detection rate, negative positive rate, false alarm rate, and false positive percent were all interpreted as FPR. Correct detection rate, percent identified, percent of calls found, true positive percent, and percent probability of detection were interpreted as R.

The performance metrics varied across studies. FNR, FPR, and R were the most commonly reported metrics (Table 1). P was reported in three instances, and ACC was only described on one occasion (Table 1). For some studies, describing FPR, R, or P was unnecessary as every automated detection was manually checked and the FP were removed (Table S7). Some authors plotted true positive rate against false positive rate to create ROC curves (Español Jiménez & van der Schaar, 2018; Hodge et al., 2015; Tsujii et al., 2016; Vu et al., 2012). The values of performance metrics varied greatly across metrics and species (Table 1). With an ideal value of 0%, FNR ranged from 0% to 54% and FPR from 0% to 100%. Conversely, with an ideal value of 100%, P and R ranged from 20% to 95% and 15% to 100%, respectively (Table 1). Wide ranging values were in some cases reported within studies. For example, Balcazar et al. (2015)

|  | Detector | |
|---|---|---|
|  | **Detected** | **Not detected** |
| **Truth — Detected** | TP — True positive | FN — False negative |
| **Not Detected** | FP — False positive | TN — True negative |

$$False\ negative\ rate = FNR = \frac{FN}{TP+FN}$$

$$False\ positive\ rate = FPR = \frac{FP}{FP+TN}$$

$$Recall = R = true\ positive\ rate = TPR = \frac{TP}{TP+FN}$$

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

$$Accuracy = ACC = \frac{TP+TN}{TP+TN+FP+FN}$$

$$F\text{-}score = 2\frac{P\ x\ R}{P+R}$$

$$Matthews\ Correlation\ Coefficient =$$
$$MCC = \frac{TP\ x\ TN - FP\ x\ FN}{\sqrt{(FP+TP)(TP+FN)(TN+FP)(TN+FN)}}$$

**FIGURE 5** Detector performance metric definitions and formulas for a binary detector/classifier. Where a detector (automated or manual) is binary if a positive or negative result is defined for each unit (e.g., acoustic file or period of time). Where number of detections is given rather than a binary output, TN is generally not defined. When considering automated detector performance, the manual analysis results (Truth) are compared to automated detector results (Detector).

reported an FNR of 0%–9% and FPR of 18%–100%, depending on the type of blue whale vocalization targeted by the automated detector, the location of the acoustic recorders, and the year of the recordings. Hodge et al. (2015) similarly concluded that right whale automated detector performance varies by location with R values of 46%–64% and FNRs of 36%–54%. Risch et al. (2013) found that minke whale FNR for automated detection ranged from 0%–51%, depending on vocalization quality in the recorded data. No species consistently had better automated detector performance than others, though such comparisons are challenging given the range in metrics employed, automated detectors applied, vocalization types, and variability in number of studies per species (Table 1).

## 3.8 | Terminology associated with data analysis methods

An inconsistency in terminology related to analysis methods in big PAM data was apparent throughout this literature review. For example, one study may describe false positives while another describes incorrect detections, two equivalent detector performance metrics. Human manual analysis of data was referred to in a variety of ways throughout the literature including "manually," "visually," and "aurally" "examined", "viewed," "logged," "inspected," "analyzed," and "detected." Automated analysis was described as "autodetection," "automated detection algorithms," "automatically detected," and "search algorithm." The terms "manual," "visual," and "aural" most commonly referred to some manner of human analysis, while the term "automated" indicated the use of an algorithm. "Detected," "detector," and "detection" were used interchangeably between manual and automated methods.

## 4 | DISCUSSION

This systematic literature review of baleen whale PAM studies revealed a shift in methodologies applied to increasing volumes of data collected. With the use of modern technologies allowing for the remote autonomous collection

of large volumes of acoustic data, the data analysis methods have appropriately shifted from predominantly relying on human analysts to methods incorporating both automated and manual techniques (Figure 2), a trend similarly observed in other fields tackling big data (Lewis et al., 2013; Sivarajah et al., 2017).

Big PAM baleen whale studies were largely limited to humpback, blue, fin, and right whales, with a small proportion of research investigating more than one species at a time. The bias in reporting toward these species likely reflects a combination of factors, including their conservation status, distributions that overlap or conflict with human activity, propensity towards being vocally active and thus captured on recorders, or public interest. For example, the acoustically prolific humpback whale song first became famous in the 1970s, and now has a great deal of public and scientific interest (Payne & McVay, 1971). Conversely, other baleen species are in locations where it is more difficult to collect data such as the bowhead whale in the Arctic (Blackwell, Richardson, Greene, & Streever, 2007; Charif et al., 2013; Stafford et al., 2012) or have vocalizations that are more rare, poorly understood, or difficult to study acoustically such as the Omura's whale (*Balaenoptera omurai*; Cerchio et al., 2015). As more work describes the sounds of these species, their presence in the literature will likely increase.

One of the great benefits of PAM data collection is that it captures signals from all species acoustically active in the area that vocalize within the parameters of the recording system. Therefore, by restricting research efforts to only describing the occurrence of a single species, a practice that likely reflects many factors including budget, time, access to different automated detectors, and priorities, researchers are not fully utilizing the investment they have made in data collection. Thus far, most multispecies studies involve fin and blue whales and low sampling rate data. This is unsurprising given that these species produce regular, predictable, species-specific acoustic signals for which reliable automated detectors can be more easily developed relative to other baleen whales such as humpback whales, which produce more varied signals that can change considerably over time (Rekdahl et al., 2017). Further, blue and fin whales produce low-frequency signals that can be captured by recorders sampling at 100 or 250 Hz, requiring less power and memory storage in recording devices. However, acoustic recorders are now sufficiently powerful to sample at higher rates for long durations, allowing for a greater range of signals to be captured and more species to be described (Moloney et al., 2018). While few studies in the present review explored all species acoustically present, high sampling rates in new generation recorders, an increasing knowledge of marine mammal vocalizations allowing more species to be identified, and an impetus towards ecosystem studies (Sherman et al., 2005) should lead towards more multispecies studies. The development of more automated detectors, or suites of automated detectors that capture and differentiate many types of acoustic signals could encourage more multispecies studies. Worth noting, the present review could not effectively differentiate between studies that investigated and reported on a single species and those that investigated all species but reported on species separately across publications. Therefore, the practice of multispecies analysis may be more common than is perceived here. As research groups include more species, putting greater strain on analysis efforts, the shift to utilizing a combination of automated and manual analysis techniques is expected to continue, as was observed in studies that investigated all acoustically active species in this review.

While a trend toward incorporating automation into big PAM baleen whale analysis methods was observed, protocols showed great variation across studies, which likely reflects differences in research goals, time, budget, acoustic signal(s) of interest, and experience of research groups. Some variation may be due to different requirements from the journal of choice for publication, with some potentially requiring more stringently described methodologies (e.g., automated detector performance evaluation) than others. The acoustic research community must work to standardize methods among different projects to increase comparability/consistency across studies (Mellinger, Stafford, et al., 2007). Therefore, we provide the following recommendations to encourage future research towards this goal.

## 4.1 | Recommendations

Recommendations for future work on baleen whale PAM are based on the present literature review as well as recommendations from other fields that utilize big data. Suggestions seek to find a balance between (1) being

reasonable and achievable for researchers, regardless of budget and time restraints; (2) maintaining consistency with previous baleen whale PAM literature; and (3) achieving standards utilized beyond the marine PAM field to encourage interdisciplinary consistency (e.g., PAM in the terrestrial realm). In many instances, our recommendations are limited by a lack of consistency observed in the literature (in big PAM data or other big data studies). These should therefore not be considered hard and fast rules, but rather a starting point to a larger methodological discussion within the PAM data community.

### 4.1.1    |    Recommendation 1: Methodology for characterization studies

For researchers seeking to characterize baleen whale acoustic signals or acoustic behavior, we recommend analysis driven by the SNR of acoustic signals. This was commonly employed in studies that investigated both occurrence and characterization in the present literature review. By restricting characterization to signals of high SNR, researchers reduce the chance that they misidentify, misclassify, or inaccurately characterize the signal (Mellinger, 2004). The method of calculating SNR and what is considered "high" must be clearly defined in the study. One approach is to compare signals to the sound level of the appropriate frequency band (Castellote et al., 2012b; Gavrilov et al., 2012), which can be made more specific by using the ambient sound immediately before the signal (for the same duration of the signal; Kowarski et al., 2019). There is arguably value to additionally characterizing low SNR signals if the quality and context is such that the researcher can reliably identify species. If the characteristics of lower SNR vocalizations are shared with the acoustics community, future researchers will be better able to classify faint signals in their own data, which can be particularly important for endangered or rare species. Where researchers include varying SNR vocalizations in characterizations, they should clearly specify and differentiate these in their results. Automated methods can be utilized to identify high SNR signals for manual review, though effective automated detectors may not be available before a signal has been adequately characterized. Where automation is not feasible to guide manual review, we recommend employing partial manual review to identify acoustic signals of interest with high SNR.

Partial manual review was the most common characterization method in the present review for non-song vocalizations and is arguably the most appropriate method when effective automation is unavailable, and the quantity of data is such that full manual review is unachievable. Specific protocols to select data for partial manual review and the amount of data reviewed were varied in the literature reviewed, likely reflecting different research questions, goals, acoustic signals of interest, time, and budget. While we recommend that protocols incorporate SNR, one must then decide how to ensure the sample is appropriate for the research question. For example, if a study is claiming to describe vocalizations of a species, selecting high SNR vocalizations from three days is arguably not representative as vocal behavior may vary across space and time. Rather, vocalizations should be analyzed across time of day, day of year, and locations to increase the chance of a representative sample. Ancillary information can also be considered when determining analysis protocols such as seasonal ice cover, known migration times, and information from visual surveys.

Studies that incorporated characterization of songs commonly employed full manual review, but these were almost entirely limited to relatively small studies (2,000 hr or less of acoustic data) and may become increasingly unrealistic for research groups to undertake in the future as data sets continue to grow. Furthermore, where a full manual review does not limit its characterization to high SNR events, it risks mischaracterization. Therefore, when full manual review is unachievable, we argue partial manual review driven by SNR (or vocalization quality) is sufficient for song characterization research.

Our recommended guidelines for characterization methodology are summarized as follows:

- Select vocalizations for analysis based on SNR. As a minimum the analysis should focus on high SNR vocalizations; however, describing vocalizations at all signal to noise ratios is recommended.

- If available, use automated techniques to identify high SNR vocalizations.
- Where automated methods are inapplicable, undertake a partial manual review, at minimum, to identify signals of interest. The manual review protocol must capture a representative sample and should consider ancillary information.
- Include a wide temporal window (over season, day, hour) and diverse spatial locations where possible to fully capture the variability on vocalization types.

### 4.1.2 | Recommendation 2: Methodology for occurrence studies

For the question of baleen whale spatial or temporal acoustic occurrence, we recommend utilizing a combination of manual and automated methods, the most common technique observed in the present literature review. This recommendation applies to studies solely investigating occurrence and those that include additional research questions. The sole use of manual review (full or partial) would only be necessary where automated detectors are ineffective (e.g., signal is too variable, there are other sounds that interfere with automated detection) or where automated detectors cannot be developed because too little is understood about the acoustic signal of interest. Full manual review would be unrealistic for many research groups, and partial manual review inevitably restricts findings to only a portion of the data. Automated detection without the verification of manual review was rare in the literature and should be avoided. Indeed, the present review revealed that no species has consistently had high automated detector performance reported such that one could argue no manual review is required. Conclusions drawn from research studies solely reliant on automated detectors that are known to have variable performance depending on factors such as location, season, and other ocean sounds should be viewed with caution (Erbs et al., 2017; Hodge et al., 2015; Širović et al., 2015).

The question then remains, what protocol should be followed when implementing a combination of manual and automated techniques to determine baleen whale occurrence and how can researchers effectively determine automatic detector performance? Many researchers in the present literature review opted to manually review every automated detection. This technique can result in an intensive workload where effort in terms of time and cost will vary depending on the experience level of analysts. Such effort is likely beyond the reach of some research groups and could become completely unattainable when investigating more than one or two species. Indeed, the present literature review found that no study investigating more than two species checked every automated detection. We therefore argue that, unless the number of automated detections is reasonably small or the research requirements are such that not a single false automated detection is acceptable in the results, the manual review of each automated detection or automated detection event is unnecessary. Rather, researchers should follow the trend of many big data studies and manually review a subset of data to determine automated detector performance metrics (Knight et al., 2017; see Recommendation 4; Lewis et al., 2013; Ofli et al., 2016). The topic of automated detector performance metrics is discussed further in Recommendation 3.

To effectively compare occurrence results between studies, researchers must not only aim for consistent methodologies but also consistent presentation of results, especially within species. The present literature review observed two main categories of units used in results: number of vocalizations or automated detections and presence of vocalizations or automated detections, where presence is given the same weight whether one detection happened within that timeframe or 200. Number of detections can allow a finer comparison of relative occurrence between e.g., two recording stations, but is susceptible to misrepresenting findings if detector performance (manual or automated) is variable across space or time. Presence is arguably more comparable across studies. The same automated detector or human analyst is rarely applied to multiple studies, and if they were, they would be expected to perform differently if the acoustic environment between the two studies is not comparable (Erbs et al., 2017). When there is a high density of vocalizations, this difference in performance is reduced by looking at presence rather than the number of vocalizations (e.g., Hodge et al., 2015 found that missed detections were high when assessed per

vocalization but low when assessed per day). What's more, presence allows TN to be defined and calculated, which is not the case when considering number of detections (Figure 5). We therefore argue that presence is the preferred unit for occurrence results. In the present review, presence results were most commonly reported as hours per day or per month. Maintaining this trend would increase comparability between studies, though the appropriate unit of time analyzed and presented will depend on the specific research questions being addressed, and, with data sets growing ever larger, days per week, month, season, or year may become more achievable and common units.

In contrast to most methodologies observed in the present literature review, those investigating diel occurrence showed some consistency between studies that should continue to be upheld in future research where authors seek to directly compare studies. Diel studies should incorporate all days where vocalizations were deemed present, define light regimes based on nautical twilight (angle of the sun), and present results as hourly mean adjusted number of vocalizations (or presence of vocalizations) per hour for each light regime. There is some discrepancy in how many light regimes are used. Most studies reviewed here used three or four, where those with three had one regime for twilight, while those with four separated twilight into dawn and dusk. Given that acoustic occurrence did vary between dawn and dusk in some studies (Kowarski et al., 2018; Kowarski et al., 2019; Mussoline et al., 2012), it seems reasonable to have them separated and to use four light regimes. These recommendations based on trends in the literature may not be applicable in all regions of the world (e.g., in high latitudes where using nautical twilight would be inappropriate) and should be adapted appropriately. Incorporating novel approaches to diel analysis could be useful, though we argue there is value in additionally presenting results as per the typical methods to allow comparison and discussion.

Our recommended guidelines for occurrence methodology are summarized as follows:

- Use a combination of automated and manual methods where the automated detector performance is determined from the manual review of a subset of data.
- Present results as presence (e.g., as hours per day or days per month).
- Define light regimes for diel analysis based on nautical twilight (where appropriate), and compare the mean adjusted number of vocalizations per hour for each light regime.

### 4.1.3 | Recommendation 3: Automated detector metrics

This review highlights previous findings that automated detectors in PAM vary in performance depending on location and timeframe (Erbs et al., 2017; Hodge et al., 2015; Širović et al., 2015). This is unsurprising given the wide range of acoustic conditions encountered in the marine realm. For example, the performance of an automated right whale upcall detector would be different in a month where right whales were the only species present than in a month where humpback whales were singing because humpbacks produce similar signals to right whales (Parks et al., 2011). We therefore propose that any application of automated techniques must be paired with manual review to determine automated detector performance, a conclusion similarly reached by Knight et al. (2017) after reviewing PAM avian literature. Where studies span numerous locations and seasons, researchers should further consider calculating automated detector performance for each circumstance as was completed by Erbs et al. (2017).

Our review revealed that automated detector performance metrics are underrepresented in the baleen whale PAM literature with more than half of studies that used automated detectors not describing performance. Studies that did describe performance used varying metrics that were often poorly defined in terms of formula and the unit of time being considered. This variation makes it difficult to compare results across studies, species, locations, and seasons. While FNR and FPR were the most commonly reported metrics in the present review, Knight et al. (2017) argue that P, R, F-score (Figure 5), and area under the (P and R) curve (AUC) are required for PAM studies to maintain standards with other disciplines. We concur that researchers should strive to present these metrics, particularly P and R, as F-score and AUC are aggregate metrics to represent both P and R. Where P provides the proportion of

automated detections (or detection events per unit time) that were correct and R provides the proportion of true detections (or detection events per unit time) that were captured by the automated detector (Figure 5). Another less common metric to summarize automated detector performance is Matthew's correlation coefficient (Figure 5) which has the benefit of incorporating TN, a metric not represented in P, R, F-score, or AUC, and is appropriate for the imbalanced data often seen in PAM (Boughorbel et al., 2017). Though, TN can only be determined where automated detector output is binary (e.g., presence/absence). A formula must always be presented alongside any performance metric as well as the unit of performance (e.g., it must be clear if the metric is describing the automated detector performance per vocalization or per unit time, relative to the manual analysis). Indeed, the variation in performance metric values in the present review may be partially due to studies evaluating different units.

The significance of automated detector metrics will vary between studies as there will almost always be a trade-off between P and R. For some studies, achieving high R will be important because the acoustic signals of interest are rare or the species is endangered, but this will likely result in a lower P (Buchan et al., 2018; Kerosky et al., 2012; Mellinger, Stafford, et al., 2007). Studies on species that are more vocally prolific may allow a lower R in favor of optimizing P (Mellinger, Stafford, et al., 2007). Authors should indicate whether they have developed automated detectors with the intent to optimize P, R, or some metric that balances both such as F-score, AUC, or MCC.

Once performance metrics are calculated, researchers must critically interpret them and present results in a meaningful way. Where automated detector performance is found to be low, the automated detector may be deemed ineffective and either the automated detector must be improved, or an appropriate manual analysis protocol must be undertaken. At what level automated detector performance is deemed too poor will depend on the research goals. Researchers may improve or optimize automated detector results post processing. For example, some automated detectors can have confidence levels imposed, or a minimum number of automated detections per timeframe requirement can be applied (Delarue et al., 2018; Mouy et al., 2016). Presenting automated detector results alongside performance metrics will help ensure results are interpreted correctly. For example, an R of 0.50 tells the reader that the results only represent half of the true occurrence of the acoustic signal. Further, where authors understand what factors impact automated detector performance, they should present such information along with their acoustic occurrence results (e.g., anthropogenic activities, weather events, ambient sound levels).

In addition to evaluating automated detectors, researchers should strive to validate the effectiveness of the manual review, given the potential for discrepancies between human analysts that produce the "truth data." This can be done by having multiple humans review the same data (at least for a subset) to calculate the agreement between them (e.g., Leroy, Thomisch, et al., 2018). Where only employing a single analyst is possible (given the limited resources of many research groups and scarcity of qualified individuals), papers should provide sufficient evidence that the experience of the analyst is such that the truth data can be considered sufficiently reliable.

Our recommended guidelines for automated detector metrics are summarized as follows:

- Pair automated detector implementation with manual review to determine automated detector performance metrics.
- Evaluate automated detector performance metrics separately for different acoustic circumstances (e.g., each season and location).
- At minimum, evaluate P and R for each automated detector.
- Include a formula and the unit investigated for all automated detector performance metrics reported.
- Evaluate the reliability of the human created "truth data."

### 4.1.4 | Recommendation 4: Data selection to calculate automated detector metrics

Regardless of the research question, one of the most variable protocols observed in the present literature review was how data were selected for manual analysis to validate automated detectors. The amount of acoustic data

reviewed ranged greatly with little to no consistency between studies. The manner data were selected for manual review was inconsistent. Depending on the study, data for manual validation of automated detections were selected completely randomly or was based on automated detector counts, time to capture the entire recording period, location to capture all recording sites, ambient sound, or suspicion that automated detections were incorrect. Given the known variability in automated detector performance that can occur across these variables, we recommend that researchers account for all of them when selecting data for manual validation. Proceeding otherwise could result in performance metrics that are not truly representative of the results.

The question then remains, how much data should be manually validated? This will likely be determined by the constraining factors of time, budget, and resources, but need to be sufficiently high enough to ensure a representative sample was manually analyzed. A manner to check if the sample data set is sufficiently large is to calculate the P and R of the automated detectors based on the sample, which may be 5% of acoustic data, then calculate P and R as if only 2.5% of acoustic was reviewed. If P and R do not vary between the two sample sizes, the plateau of performance has likely been reached, and the performance metric values are reliable.

Our recommendation for samples of data selected to validate automated detectors is:

- Samples must represent the entire breadth of conditions present in the data set of interest to effectively report on automated detector performance.

### 4.1.5 | Recommendation 5: Terminology

Terminology used to describe methods to analyze big PAM baleen whale data varied across studies, resulting in a lack of clarity when attempting to compare methods and results during the present review. For example, detections can refer to the outcome of human manual analysis, automated detectors, or both, depending on the study. Further, studies incorporating sighting surveys also have visual detections (a term also applied to manual PAM analysis). Knight et al. (2017) similarly found a lack of consistency in avian PAM literature terminology, particularly regarding automated detector performance metrics. Future research should strive to maintain consistency and clarity. To this end, we have compiled the following guidelines that are recommended for future studies:

- Clearly define the terminology that will be used in the paper. For example, if using an algorithm that first detects a signal, then classifies it as a vocalization, the author should clarify that the term "automated detector" refers to an automated detector-classifier. A small glossary section would add great clarity to PAM articles.
- Qualify any use of the term "detection." For example, "manual detection" describes a human-detected signal, "automated detection" describes an algorithm-detected signal, and "manually validated detection" describes a signal that was first detected automatically then validated by a human.
- Use the term "manual" or "human" to qualify detections from human analysts. These are more broadly applicable beyond the field of acoustics than "aural" or "visual," for example.
- Use the term "automated" to qualify results from a computer algorithm.

### 4.2 | Conclusions

We present a literature review of baleen whale PAM methods and propose guidelines for future work to enhance consistency and rigor of methods. The PAM community should work towards implementing consistent standards while considering the sometimes-limited resources of research groups and how requirements vary across research questions. Big data will likely continue to increase as technology matures and analysis methodologies must find the balance between adapting appropriately, while maintaining consistency. Results from PAM marine mammal research,

especially those investigating occurrence, can reliably inform management decisions. However, standards or best practices are required and must be upheld for findings to be considered reliable and taken into consideration as appropriate.

## AUTHOR CONTRIBUTIONS

**Hilary Moors-Murphy:** Conceptualization; funding acquisition; methodology; project administration; supervision; writing-review and editing.

## ORCID

*Katie A. Kowarski* https://orcid.org/0000-0002-1325-8321

## REFERENCES

Balcazar, N. E., Tripovich, J. S., Klinck, H., Nieukirk, S. L., Mellinger, D. K., Dziak, R. P., & Rogers, T. L. (2015). Calls reveal population structure of blue whales across the southeast Indian Ocean and the southwest Pacific Ocean. *Journal of Mammalogy*, *96*(6), 1184–1193. https://doi.org/10.1093/jmammal/gyv126

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

Blackwell, S. B., Richardson, W. J., Greene, C. R., Jr., & Streever, B. (2007). Bowhead whale (*Balaena mysticetus*) migration and calling behaviour in the Alaskan Beaufort Sea, Autumn 2001–04: An acoustic localization study. *Arctic*, *60*(3), 255–270. http://www.jstor.org/stable/40512894

Bort, J., Van Parijs, S. M., Stevick, P. T., Summers, E., & Todd, S. (2015). North Atlantic right whale *Eubalaena glacialis* vocalization patterns in the central Gulf of Maine from October 2009 through October 2010. *Endangered Species Research*, *26* (3), 271–280. https://doi.org/10.3354/esr00650

Boughorbel, S., Jarray, F., & El-Anbari, M. (2017). Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. *PLoS ONE*, *12*(6), e0177678. https://doi.org/10.1371/journal.pone.0177678

Buchan, S. J., Hucke-Gaete, R., Stafford, K. M., & Clark, C. W. (2018). Occasional acoustic presence of Antarctic blue whales on a feeding ground in southern Chile. *Marine Mammal Science*, *34*(1), 220–228.

Buchan, S. J., Stafford, K. M., & Hucke-Gaete, R. (2015). Seasonal occurrence of southeast Pacific blue whale songs in southern Chile and the eastern tropical Pacific. *Marine Mammal Science*, *31*(2), 440–458. https://doi.org/10.1111/mms.12173

Burnham, R. E., & Duffus, D. A. (2019). Acoustic predator–prey reaction: Gray whales' (*Eschrichtius robustus*) acoustic response to killer whales (*Orcinus orca*). *Aquatic Mammals*, *45*(3), 340–348. https://doi.org/10.1578/AM.45.3.2019.340

Castellote, M., Clark, C. W., & Lammers, M. O. (2012a). Acoustic and behavioural changes by fin whales (*Balaenaptera physalus*) in response to shipping and airgun noise. *Biological Conservation*, *147*(1), 115–122. https://doi.org/10.1016/j.biocon.2011.12.021

Castellote, M., Clark, C. W., & Lammers, M. O. (2012b). Fin whale (*Balaenoptera physalus*) population identity in the western Mediterranean Sea. *Marine Mammal Science*, *28*(2), 325–344. https://doi.org/10.1111/j.1748-7692.2011.00491.x

Cerchio, S., Andrianantenaina, B., Lindsay, A., Rekdahl, M., Andrianarivelo, N., & Rasoloarijao, T. (2015). Omura's whales (*Balaenoptera omurai*) off northwest Madagascar: Ecology, behaviour and conservation needs. *Royal Society Open Science*, *2*(10), 150301.

Cerchio, S., Strindberg, S., Collins, T., Bennett, C., & Rosenbaum, H. (2014). Seismic surveys negatively affect humpback whale singing activity off northern Angola. *PLoS ONE*, *9*(3), e86464. https://doi.org/10.1371/journal.pone.0086464

Charif, R. A., Rahaman, A., Muirhead, C. A., Pitzrick, M. S., Warde, A. M., Hall, J., Pyć, C. D., & Clark, C. W. (2013). Bowhead whale acoustic activity in the southeast Beaufort Sea during late summer 2008–2010. *Journal of the Acoustical Society of America*, *134*(6), 4323–4334. https://doi.org/10.1121/1.4824679

Cholewiak, D. M., Cerchio, S., Jacobsen, J. K., Urbán-R, J., & Clark, C. W. (2018). Songbird dynamics under the sea: acoustic interactions between humpback whales suggest song mediates male interactions. *Journal of Royal Society Open Science*, *5*(2), 171298. https://doi.org/10.1098/rsos.171298

Clark, C. W., Borsani, J. F., & Notarbartolo-di-Sciara, G. (2002). Vocal activity of fin whales, *Balaenoptera physalus*, in the Ligurian Sea. *Marine Mammal Science*, *18*(1), 286–295. https://doi.org/10.1111/j.1748-7692.2002.tb01035.x

Davis, G. E., Baumgartner, M. F., Bonnell, J. M., Bell, J., Berchok, C., Bort Thornton, J., Brault, S., Buchanan, G., Charif, R. A., Cholewiak, D., Clark, C. W., Corkeron, P. J., Delarue, J., Dudzinski, K., Hatch, L., Hildebrand, J., Hodge, L., Klinck, H., Kraus, S., … & Van Parijs, S. M. (2017). Long-term passive acoustic recordings track the changing distribution of North Atlantic right whales (*Eubalaena glacialis*) from 2004 to 2014. *Scientific Reports*, *7*(1), 1–12. https://doi.org/10.1038/s41598-017-13359-3

De Vreese, S., van der Schaar, M., Weissenberger, J., Erbs, F., Kosecka, M., Solé, M., & André, M. (2018). Marine mammal acoustic detections in the Greenland and Barents Sea, 2013–2014 seasons. *Scientific Reports*, *8*(1), 16882. https://www.nature.com/articles/s41598-018-34624-z

Delarue, J. (2008). *Northwest Atlantic fin whale vocalizations: Geographic variations and implications for stock assessments* [Unpublished master's thesis]. College of the Atlantic.

Delarue, J., Kowarski, K. A., Maxner, E. E., MacDonnell, J. T., & Martin, S. B. (2018). *Acoustic monitoring along Canada's East Coast: August 2015 to July 2017* (Environmental Studies Research Fund Report Number 215). JASCO Applied Sciences.

Dugan, P. J., Rice, A. N., Urazghildiiev, I. R., & Clark, C. W. (2010, May 7). North Atlantic right whale acoustic signal processing: Part I. Comparison of machine learning recognition algorithms [Paper presentation]. IEEE Long Island Systems, Applications and Technology Conference, Farmingdale, NY. https://doi.org/10.1109/LISAT.2010.5478268

Erbs, F., Elwen, S. H., & Gridley, T. (2017). Automatic classification of whistles from coastal dolphins of the southern African subregion. *Journal of the Acoustical Society of America*, *141*(4), 2489–2500. https://doi.org/10.1121/1.4978000

Español Jiménez, S., & van der Schaar, M. (2018). First record of humpback whale songs in Southern Chile: Analysis of seasonal and diel variation. *Marine Mammal Science*, *34*(3), 718–733. https://doi.org/10.1111/mms.12477

Fournet, M. E. H., Jacobsen, L., Gabriele, C. M., Mellinger, D. K., & Klinck, H. (2018). More of the same: Allopatric humpback whale populations share acoustic repertoire. *PeerJ*, *6*, e5365. https://doi.org/10.7717/peerj.5365

Frouin-Mouy, H., Kowarski, K. A., Martin, B., & Bröker, K. (2017). Seasonal trends in acoustic detection of marine mammals in Baffin Bay and Melville Bay, Northwest Greenland + Supplementary Appendix 1. *Arctic*, *70*(1), 59–76. https://doi.org/10.14430/arctic4632

Garland, E. C., Goldizen, A. W., Lilley, M. S., Rekdahl, M. L., Garrigue, C., Constantine, R., Hauser, N. D., Poole, M. M., Robbins, J., & Noad, M. J. (2015). Population structure of humpback whales in the western and central South Pacific Ocean as determined by vocal exchange among populations. *Conservation Biology*, *29*(4), 1198–1207. https://doi.org/10.1111/cobi.12492

Gavrilov, A. N., & McCauley, R. D. (2013). Acoustic detection and long-term monitoring of pygmy blue whales over the continental slope in southwest Australia. *Journal of the Acoustical Society of America*, *134*(3), 2505–2513. https://doi.org/10.1121/1.4816576

Gavrilov, A. N., McCauley, R. D., & Gedamke, J. (2012). Steady inter and intra-annual decrease in the vocalization frequency of Antarctic blue whales. *Journal of the Acoustical Society of America*, *131*(6), 4476–4480. https://doi.org/10.1121/1.4707425

Gavrilov, A. N., McCauley, R. D., Paskos, G., & Goncharov, A. (2018). Southbound migration corridor of pygmy blue whales off the northwest coast of Australia based on data from ocean bottom seismographs. *Journal of the Acoustical Society of America*, *144*(1), EL281–EL285. https://doi.org/10.1121/1.5063452

Gould, E. (1975). Neonatal vocalizations in bats of eight genera. *Journal of Mammalogy*, *56*, 15–29.

Hannay, D. E., Delarue, J., Mouy, X., Martin, B. S., Leary, D., Oswald, J. N., & Vallarta, J. (2013). Marine mammal acoustic detections in the northeastern Chukchi Sea, September 2007–July 2011. *Continental Shelf Research*, *67*, 127–146. https://doi.org/10.1016/j.csr.2013.07.009

Hodge, K. B., Muirhead, C. A., Morano, J. L., Clark, C. W., & Rice, A. N. (2015). North Atlantic right whale occurrence near wind energy areas along the mid-Atlantic US coast: Implications for management. *Endangered Species Research*, *28*(3), 225–234. https://doi.org/10.3354/esr00683

Jacobs, E., Duffy, M., Magolan, J., Galletti Vernazzani, B., Cabrera, E., Landea, R., Buchan, S., & Sayigh, L. (2019). First acoustic recordings of critically endangered eastern South Pacific southern right whales (*Eubalaena australis*). *Marine Mammal Science*, *35*(1), 284–289. https://doi.org/10.1111/mms.12519

Johnson, H. D., Stafford, K. M., George, J. C., Ambrose, W. G., Jr., & Clark, C. W. (2015). Song sharing and diversity in the Bering Chukchi Beaufort population of bowhead whales (*Balaena mysticetus*), spring 2011. *Marine Mammal Science*, *31*(3), 902–922. https://doi.org/10.1111/mms.12196

Johnson, J. B., Menzel, M. A., Edwards, J. W., & Ford, W. M. (2002). A comparison of 2 acoustical bat survey techniques. *Wildlife Society Bulletin*, *30*(3), 931–936.

Jolliffe, C. D., McCauley, R. D., Gavrilov, A. N., Jenner, K. C. S., Jenner, M.-N. M., & Duncan, A. J. (2019). Song variation of the South Eastern Indian Ocean pygmy blue whale population in the Perth Canyon, Western Australia. *PLoS ONE*, *14*(1), e0208619.

Kalan, A. K., Mundry, R., Wagner, O. J., Heinicke, S., Boesch, C., & Kühl, H. S. (2015). Towards the automated detection and occupancy estimation of primates using passive acoustic monitoring. *Ecological Indicators*, *54*, 217–226.

Kerosky, S. M., Širović, A., Roche, L. K., Baumann-Pickering, S., Wiggins, S. M., & Hildebrand, J. A. (2012). Bryde's whale seasonal range expansion and increasing presence in the Southern California Bight from 2000 to 2010. *Deep Sea Research Part I: Oceanographic Research Papers*, *65*, 125–132. https://doi.org/10.1016/j.dsr.2012.03.013

Klinck, H., Mellinger, D. K., Klinck, K., Bogue, N. M., Luby, J. C., Jump, W. A., Geoffrey B. Shilling, G. B., Litchendorf, T., Wood, A. S., Schorr, G. S., & Baird, R. W. (2012). Near-real-time acoustic monitoring of beaked whales and other cetaceans using a Seaglider™. *PLoS ONE*, *7*(5), e36128. https://doi.org/10.1371/journal.pone.0036128

Knight, E. C., Hannah, K. C., Foley, G. J., Scott, C. D., Brigham, R. M., & Bayne, E. (2017). Recommendations for acoustic recognizer performance assessment with application to five common automated signal recognition programs. *Avian Conservation & Ecology*, *12*(2), 14. https://doi.org/10.5751/ACE-01114-120214

Kowarski, K. A., Evers, C., Moors Murphy, H., Martin, S. B., & 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. *Marine Mammal Science*, *34*(1), 169–189. https://doi.org/10.1111/mms.12447

Kowarski, K. A., Moors-Murphy, H., Maxner, E. E., & Cerchio, S. (2019). Western North Atlantic humpback whale fall and spring acoustic repertoire: Insight into onset and cessation of singing behavior. *Journal of the Acoustical Society of America*, *145*(4), 2305–2316. https://doi.org/10.1121/1.5095404

Lammers, M. O., Fisher-Pool, P. I., Au, W. W. L., Meyer, C. G., Wong, K. B., & Brainard, R. E. (2011). Humpback whale *Megaptera novaeangliae* song reveals wintering activity in the Northwestern Hawaiian Islands. *Marine Ecology Progress Series*, *423*, 261–268. https://doi.org/10.3354/meps08959

Lee, C. H., & Yoon, H.-J. (2017). Medical big data: Promise and challenges. *Kidney Research Clinical Practice*, *36*, 3–11. https://dx.doi.org/10.23876%2Fj.krcp.2017.36.1.3

Leroy, E. C., Samaran, F., Bonnel, J., & Royer, J.-Y. (2016). Seasonal and diel vocalization patterns of Antarctic blue whale (*Balaenoptera musculus intermedia*) in the southern Indian Ocean: A multi-year and multi-site study. *PLoS ONE*, *11*(11), e0163587. https://doi.org/10.1371/journal.pone.0163587

Leroy, E. C., Samaran, F., Bonnel, J., & Royer, J.-Y. (2017). Identification of two potential whale calls in the southern Indian Ocean, and their geographic and seasonal occurrence. *Journal of the Acoustical Society of America*, *142*(3), 1413–1427. https://doi.org/10.1121/1.5001056

Leroy, E. C., Samaran, F., Stafford, K. M., Bonnel, J., & Royer, J.-Y. (2018a). Broad-scale study of the seasonal and geographic occurrence of blue and fin whales in the Southern Indian Ocean. *Endangered Species Research*, *37*, 289–300. https://doi.org/10.3354/esr00927

Leroy, E. C., Thomisch, K., Royer, J.-Y., Boebel, O., & Van Opzeeland, I. (2018b). On the reliability of acoustic annotations and automatic detections of Antarctic blue whale calls under different acoustic conditions. *Journal of the Acoustical Society of America*, *144*(2), 740–754. https://doi.org/10.1121/1.5049803

Lewis, S. C., Zamith, R., & Hermida, A. (2013). Content analysis in an era of big data: A hybrid approach to computational and manual methods. *Journal of Broadcasting & Electronic Media*, *57*(1), 34–52. https://doi.org/10.1080/08838151.2012.761702

Lieberman, P. (1968). Primate vocalizations and human linguistic ability. *Journal of the Acoustical Society of America*, *44*(6), 1574–1584.

Magnúsdóttir, E. E., Miller, P. J. O., Lim, R., Rasmussen, M. H., Lammers, M. O., & Svavarsson, J. (2015). Humpback whale (*Megaptera novaeangliae*) song unit and phrase repertoire progression on a subarctic feeding ground. *Journal of the Acoustical Society of America*, *138*(5), 3362–3374. https://doi.org/10.1121/1.4935517

Marques, T. A., Munger, L., Thomas, L., Wiggins, S., & Hildebrand, J. A. (2011). Estimating North Pacific right whale *Eubalaena japonica* density using passive acoustic cue counting. *Endangered Species Research*, *13*(3), 163–172. https://doi.org/10.3354/esr00325

Matsuo, I., Akamatsu, T., Iwase, R., & Kawaguchi, K. (2013, March 5–8). *Automated acoustic detection of fin whale calls off Kushiro-Tokachi at the deep sea floor observatory* [Paper presentation]. IEEE International Underwater Technology Symposium (UT), Tokyo, Japan.

McDonald, M. A. (2006). An acoustic survey of baleen whales off Great Barrier Island, New Zealand. *New Zealand Journal of Marine and Freshwater Research*, *40*(4), 519–529. https://doi.org/10.1080/00288330.2006.9517442

Melcon, M. L., Cummins, A. J., Kerosky, S. M., Roche, L. K., Wiggins, S. M., & Hildebrand, J. A. (2012). Blue whales respond to anthropogenic noise. *PLoS ONE*, *7*(2), e32681. https://doi.org/10.1371/journal.pone.0032681

Mellinger, D. K. (2004). A comparison of methods for detecting right whale calls. *Canadian Acoustics*, *32*(2), 55–65. https://jcaa.caa-aca.ca/index.php/jcaa/article/view/1588

Mellinger, D. K., Nieukirk, S. L., Matsumoto, H., Heimlich, S. L., Dziak, R. P., Haxel, J., Fowler, M., Meinig, C., & Miller, H. V. (2007). Seasonal occurrence of North Atlantic right whale (*Eubalaena glacialis*) vocalizations at two sites on the Scotian Shelf. *Marine Mammal Science*, *23*(4), 856–867. https://doi.org/10.1111/j.1748-7692.2007.00144.x

Mellinger, D. K., Stafford, K. M., Moore, S. E., Dziak, R. P., & Matsumoto, H. (2007). An overview of fixed passive acoustic observation methods for cetaceans. *Oceanography*, *20*(4), 36–45.

Mellinger, D. K., Stafford, K. M., Moore, S. E., Munger, L. M., & Christopher, G. (2004). Detection of North Pacific right whale (*Eubalaena japonica*) calls in the Gulf of Alaska. *Marine Mammal Science*, 20(4), 872–879.

Mercado, E., III. (2018). Recurrent acoustic patterns produced by a singing humpback whale (*Megaptera novaeangliae*). *Western Indian Ocean Journal of Marine Science, Special Issue*, 1, 41–51.

Miksis-Olds, J. L., Buck, J. R., Noad, M. J., Cato, D. H., & Dale Stokes, M. (2008). Information theory analysis of Australian humpback whale song. *Journal of the Acoustical Society of America*, 124(4), 2385–2393.

Miksis-Olds, J. L., Nieukirk, S. L., & Harris, D. V. (2018). Two unit analysis of Sri Lankan pygmy blue whale song over a decade. *Journal of the Acoustical Society of America*, 144(6), 3618–3626. https://doi.org/10.1121/1.5084269

Moloney, J., Cole, A., Hillis, C. A., Kowarski, K. A., MacDonald, B., Pedersen, G., Tassara, L., & Camus, L. (2018, November 5–9). *Observer/AMAR G4 passive acoustic monitoring system for autonomous maritime observation*. Proceedings of AUV 2018, Porto, Portugal. http://www.jasco.com/s/PID5703971_Moloney_et_al_2019_Observer-AMAR-G4.pdf

Morano, J. L., Rice, A. N., Tielens, J. T., Estabrook, B. J., Murray, A., Roberts, B. L., & Clark, C. W. (2012a). Acoustically detected year-round presence of right whales in an urbanized migration corridor. *Conservation Biology*, 26(4), 698–707. https://doi.org/10.1111/j.1523-1739.2012.01866.x

Morano, J. L., Salisbury, D. P., Rice, A. N., Conklin, K. L., Falk, K. L., & Clark, C. W. (2012b). Seasonal and geographical patterns of fin whale song in the western North Atlantic Ocean. *Journal of the Acoustical Society of America*, 132(2), 1207–1212. https://doi.org/10.1121/1.4730890

Mouy, X., Mouy, P.-A., & Hannay, D. E. (2016). Automatic detection and visualization of marine mammal sounds in British-Columbia, Canada. *Canadian Acoustics*, 44(3). https://jcaa.caa-aca.ca/index.php/jcaa/article/view/2993

Mouy, X., Rountree, R., Juanes, F., & Dosso, S. E. (2018). Cataloging fish sounds in the wild using combined acoustic and video recordings. *Journal of the Acoustical Society of America*, 143(5), EL333–EL339. https://doi.org/10.1121/1.5037359

Munger, L. M., Wiggins, S. M., Moore, S. E., & Hildebrand, J. A. (2008). North Pacific right whale (*Eubalaena japonica*) seasonal and diel calling patterns from long-term acoustic recordings in the southeastern Bering Sea, 2000–2006. *Marine Mammal Science*, 24(4), 795–814. https://doi.org/10.1111/j.1748-7692.2008.00219.x

Mussoline, S. E., Risch, D., Hatch, L. T., Weinrich, M. T., Wiley, D. N., Thompson, M. A., Corkeron, P. J., & Van Parijs, S. M. (2012). Seasonal and diel variation in North Atlantic right whale up-calls: Implications for management and conservation in the northwestern Atlantic Ocean. *Endangered Species Research*, 17, 17–26. https://doi.org/10.3354/esr00411

Nieukirk, S. L., Mellinger, D. K., Moore, S. E., Klinck, K., Dziak, R. P., & Goslin, J. (2012). Sounds from airguns and fin whales recorded in the mid-Atlantic Ocean, 1999–2009. *Journal of the Acoustical Society of America*, 131(2), 1102–1112. https://doi.org/10.1121/1.3672648

Nieukirk, S. L., Stafford, K. M., Mellinger, D. K., Dziak, R. P., & Fox, C. G. (2004). Low-frequency whale and seismic airgun sounds recorded in the mid-Atlantic Ocean. *Journal of the Acoustical Society of America*, 115(4), 1832–1843. https://doi.org/10.1121/1.1675816

Ofli, F., Meier, P., Imran, M., Castillo, C., Tuia, D., Rey, N., Briant, J., Millet, P., Reinhard, F., & Parkan, M. (2016). Combining human computing and machine learning to make sense of big (aerial) data for disaster response. *Big Data*, 4(1), 47–59. https://doi.org/10.1089/big.2014.0064

Oswald, J. N., Au, W. W. L., & Duennebier, F. (2011). Minke whale (*Balaenoptera acutorostrata*) boings detected at the Station ALOHA Cabled Observatory. *Journal of the Acoustical Society of America*, 129(5), 3353–3360. https://doi.org/10.1121/1.3575555

Parks, S. E., Searby, A., Celerier, A., Johnson, M. P., Nowaceck, D. P., & Tyack, P. L. (2011). Sound production behavior of individual North Atlantic right whales: Implications for passive acoustic monitoring. *Endangered Species Research*, 15, 63–76. https://doi.org/10.3354/esr00368

Payne, R., & McVay, S. (1971). Songs of humpback whales. *Science*, 173(3997), 585–597. https://science.sciencemag.org/content/173/3997/585

Rekdahl, M., Tisch, C., Cerchio, S., & Rosenbaum, H. (2017). Common nonsong social calls of humpback whales (*Megaptera novaeangliae*) recorded off northern Angola, southern Africa. *Marine Mammal Science*, 33(1), 365–375. https://doi.org/10.1111/mms.12355

Risch, D., Castellote, M., Clark, C. W., Davis, G. E., Dugan, P. J., Hodge, L. E. W., … Nieukirk, S. L. (2014). Seasonal migrations of North Atlantic minke whales: Novel insights from large-scale passive acoustic monitoring networks. *Movement Ecology*, 2, 24. https://doi.org/10.1186/s40462-014-0024-3

Risch, D., Clark, C. W., Dugan, P. J., Popescu, M., Siebert, U., & Van Parijs, S. M. (2013). Minke whale acoustic behavior and multi-year seasonal and diel vocalization patterns in Massachusetts Bay, USA. *Marine Ecology Progress Series*, 489, 279–295. https://doi.org/10.3354/meps10426

Risch, D., Corkeron, P. J., Ellison, W. T., & Van Parijs, S. M. (2012). Changes in humpback whale song occurrence in response to an acoustic source 200 km away. *PLoS ONE*, 7(1), e29741. https://doi.org/10.1371/journal.pone.0029741

Risch, D., Wilson, S. C., Hoogerwerf, M., van Geel, N. C. F., Edwards, E. W. J., & Brookes, K. L. (2019). Seasonal and diel acoustic presence of North Atlantic minke whales in the North Sea. *Scientific Reports*, *9*, 3571. https://www.nature.com/articles/s41598-019-39752-8

Salisbury, D. P., Clark, C. W., & Rice, A. N. (2016). Right whale occurrence in the coastal waters of Virginia, USA: Endangered species presence in a rapidly developing energy market. *Marine Mammal Science*, *32*(2), 508–519. https://doi.org/10.1111/mms.12276

Schmidt, R. S. (1968). Chuckle calls of the leopard frog (*Rana pipiens*). *Copeia*, *1968*, 561–569.

Sherman, K., Sissenwine, M., Christensen, V., Duda, A., Hempel, G., Ibe, C., Levin, S., Lluch-Belda, D., Matishov, G., & McGlade, J. (2005). A global movement toward an ecosystem approach to management of marine resources. *Marine Ecology Progress Series*, *300*, 275–279.

Simon, M., Stafford, K. M., Beedholm, K., Lee, C. M., & Madsen, P. T. (2010). Singing behavior of fin whales in the Davis Strait with implications for mating, migration and foraging. *Journal of the Acoustical Society of America*, *128*, 3200–3210. https://doi.org/10.1121/1.3495946

Širović, A., Hildebrand, J. A., & Theile, D. (2006). Baleen whales in the Scotia Sea during January and February 2003. *Journal of Cetacean Research and Management*, *8*(2), 161–171.

Širović, A., Hildebrand, J. A., Wiggins, S. M., McDonald, M. A., Moore, S. E., & Thiele, D. (2004). Seasonality of blue and fin whale calls and the influence of sea ice in the Western Antarctic Peninsula. *Deep Sea Research Part II*, *51*(17–19), 2327–2344. https://doi.org/10.1016/j.dsr2.2004.08.005

Širović, A., Hildebrand, J. A., Wiggins, S. M., & Thiele, D. (2009). Blue and fin whale acoustic presence around Antarctica during 2003 and 2004. *Marine Mammal Science*, *25*(1), 125–136. https://doi.org/10.1111/j.1748-7692.2008.00239.x

Širović, A., Rice, A., Chou, E., Hildebrand, J. A., Wiggins, S. M., & Roch, M. A. (2015). Seven years of blue and fin whale call abundance in the Southern California Bight. *Endangered Species Research*, *28*(1), 61–76. https://doi.org/10.3354/esr00676

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, *70*, 263–286. https://doi.org/10.1016/j.jbusres.2016.08.001

Slabbekoorn, H., & Halfwerk, W. (2009). Behavioural ecology: Noise annoys at community level. *Current Biology*, *19*(16), R693–R695. https://doi.org/10.1016/j.cub.2009.07.002

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, *45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

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

Stafford, K. M., Bohnenstiehl, D. R., Tolstoy, M., Chapp, E., Mellinger, D. K., & Moore, S. E. (2004). Antarctic-type blue whale calls recorded at low latitudes in the Indian and eastern Pacific Oceans. *Deep Sea Research Part I: Oceanographic Research Papers*, *51*(10), 1337–1346. https://doi.org/10.1016/j.dsr.2004.05.007

Stafford, K. M., Chapp, E., Bohnenstiel, D. R., & Tolstoy, M. (2011). Seasonal detection of three types of "pygmy" blue whale calls in the Indian Ocean. *Marine Mammal Science*, *27*(4), 828–840. https://doi.org/10.1111/j.1748-7692.2010.00437.x

Stafford, K. M., Moore, S. E., Berchok, C. L., Wiig, O., Lydersen, C., Hansen, E., Kalmbach, D., & Kovacs, K. M. (2012). Spitsbergen's endangered bowhead whales sing through the polar night. *Endangered Species Research*, *18*(2), 95–103. https://doi.org/10.3354/esr00444

Stafford, K. M., Moore, S. E., & Fox, C. G. (2005). Diel variation in blue whale calls recorded in the eastern tropical Pacific. *Animal Behaviour*, *69*(4), 951–958. https://doi.org/10.1016/j.anbehav.2004.06.025

Stafford, K. M., Moore, S. E., Spillane, M., & Wiggins, S. (2007). Gray whale calls recorded near Barrow, Alaska, throughout the winter of 2003–04. *Arctic*, *60*(2), 167–172. https://doi.org/10.14430/arctic241

Stirling, I., Calvert, W., & Cleator, H. (1983). Underwater vocalizations as a tool for studying the distribution and relative abundance of wintering pinnipeds in the high Arctic. *Arctic*, *36*(3), 262–274. https://doi.org/10.14430/arctic2275

Thomas, M., Martin, S. B., Kowarski, K. A., Gaudet, B. J., & Matwin, S. (2019, September 16–20). *Marine mammal species classification using convolutional neural networks and a novel data representation* [Paper presentation]. European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD), Würzburg, Germany.

Thorpe, W. H. (1961). *Bird-song: The biology of vocal communication and expression in birds*. Cambridge University Press.

Tripovich, J. S., Klinck, H., Nieukirk, S. L., Adams, T., Mellinger, D. K., Balcazar, N. E., Klinck, K., Hall, E. J. S., & Rogers, T. L. (2015). Temporal segregation of the Australian and Antarctic blue whale call types (*Balaenoptera musculus spp.*). *journal of Mammalogy*, *96*(3), 603–610. https://doi.org/10.1093/jmammal/gyv065

Tsujii, K., Otsuki, M., Akamatsu, T., Matsuo, I., Amakasu, K., Kitamura, M., Kikuchi, T., Miyashita, K., & Mitani, Y. (2016). The migration of fin whales into the southern Chukchi Sea as monitored with passive acoustics. *ICES Journal of Marine Science*, *73*(8), 2085–2092. https://doi.org/10.1093/icesjms/fsv271

Videsen, S. K. A., Bejder, L., Johnson, M., & Madsen, P. T. (2017). High suckling rates and acoustic crypsis of humpback whale neonates maximise potential for mother–calf energy transfer. *Functional Ecology*, *31*(8), 1561–1573. https://doi.org/10.1111/1365-2435.12871

Vu, E. T., Risch, D., Clark, C. W., Gaylord, S., Hatch, L. T., Thompson, M. A., Wiley, D. N., & Van Parijs, S. M. (2012). Humpback whale (*Megaptera novaeangliae*) song occurs extensively on feeding grounds in the Northwest Atlantic Ocean. *Aquatic Biology*, *14*(2), 175–183. https://doi.org/10.3354/ab00390

Watkins, W. A. (1981). Activities and underwater sounds of fin whales. *Scientific Reports of the Whales Research Institute, Tokyo*, *33*, 83–117.

Weilgart, L. S., & Whitehead, H. (1997). Group-specific dialects and geographical variation in coda repertoire in South Pacific sperm whales. *Behavioral Ecology and Sociobiology*, *40*(5), 277–185. https://doi.org/10.1007/s002650050343

Whitehead, H., & Rendell, L. (2004). Movements, habitat use and feeding success of cultural clans of South Pacific sperm whales. *Journal of Animal Ecology*, *73*(1), 190–196. https://doi.org/10.1111/j.1365-2656.2004.00798.x

Whitt, A. D., Dudzinski, K., & Laliberté, J. R. (2013). North Atlantic right whale distribution and seasonal occurrence in nearshore waters off New Jersey, USA, and implications for management. *Endangered Species Research*, *20*(1), 59–69. https://doi.org/10.3354/esr00486

Wimmer, J., Towsey, M., Planitz, B., Roe, P., & Williamson, I. (2010, December 7–10). *Scaling acoustic data analysis through collaboration and automation* [Paper presentation]. IEEE 6th International Conference on e-Science, Brisbane, Australia. https://doi.org/10.1109/eScience.2010.17

Wimmer, J., Towsey, M., Roe, P., & Williamson, I. (2013). Sampling environmental acoustic recordings to determine bird species richness. *Ecological Applications*, *23*(6), 1419–1428. https://doi.org/10.1890/12-2088.1

Wright, D. L., Berchok, C. L., Crance, J. L., & Clapham, P. J. (2019). Acoustic detection of the critically endangered North Pacific right whale in the northern Bering Sea. *Marine Mammal Science*, *35*(1), 311–326. https://doi.org/10.1111/mms.12521

Wright, D. L., Castellote, M., Berchok, C. L., Panirakis, D., Crance, J. L., & Clapham, P. J. (2018). Acoustic detection of North Pacific right whales in a high-traffic Aleutian Pass, 2009–2015. *Endangered Species Research*, *37*, 77–90.

Zelick, R., Mann, D., & Popper, A. N. (1999). Acoustic communication in fishes and frogs. In R. R. Fay & A. N. Popper (Eds.), *Comparative hearing: Fish and amphibians* (pp. 363–411). Springer-Verlag. https://doi.org/10.1007/978-1-4612-0533-3_9

Zimmer, W. M. X. (2011). *Passive acoustic monitoring of cetaceans*. Cambridge University Press.

Zoidis, A. M., Smultea, M. A., Frankel, A. S., Hopkins, J. L., Day, A., McFarland, A. S., Whitt, A. D., & Fertl, D. (2008). Vocalizations produced by humpback whale (*Megaptera novaeangliae*) calves recorded in Hawaii. *Journal of the Acoustical Society of America*, *123*(3), 1737–1746. https://doi.org/10.1121/1.2836750

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.