Caller ID for Risso's and Pacific White-Sided Dolphins

Mahdi Al-Badrawi (Mahdi.AlBadrawi@unh.edu)
University of New Hampshire
Yue Liang
University of New Hampshire
Kerri D. Seger
Applied Ocean Sciences
Christopher M. Foster
University of New Hampshire
Nicholas J. Kirsch
University of New Hampshire

Research Article

Keywords: VMD, temporal properties, vocalizations, marine mammals

Posted Date: August 30th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-831254/v1

Version of Record: A version of this preprint was published at Scientific Reports on March 16th, 2022. See the published version at https://doi.org/10.1038/s41598-022-08184-2.

Caller ID for Risso's and Pacific White-sided Dolphins

Mahdi Al-Badrawi[1,2,*+], Yue Liang[2,+], Kerri D. Seger[3,+], Christopher M. Foster[2,+], and Nicholas J. Kirsch[2,+]

[1]Center for Acoustics Research and Education, University of New Hampshire, Durham, NH 03824, USA
[2]Electrical and Computer Engineering, University of New Hampshire, Durham, NH, 03824, USA
[3]Applied Ocean Sciences, Fairfax Station, VA, 22039, USA
[*]Mahdi.AlBadrawi@unh.edu
[+]these authors contributed equally to this work

ABSTRACT

Climate change affects the distributions of marine mammals[1], and some temperate water species are spreading northward into the Arctic Ocean[2,3]. Tracking expanding species is crucial to conservation efforts and using automatic detectors and classifiers to track the locations of their vocalizations could help. Risso's (Gg) and Pacific white-sided (Lo) dolphins were documented spreading poleward[2] and make very similar sounds, making it difficult for both human analysts and classification algorithms to tell them apart. Variational Mode Decomposition (VMD) has provided both an easier visualization tool[4] for human analysts and offers promising capabilities in separating call types of similar spectral and temporal properties. Here we show a new visualization tool and feature extraction technique using VMD that achieves 81.3% accuracy, even when using audio files with faint signals and high background noise levels and without context clues. Because not all dolphins whistle[5–7], being able to distinguish between just their pulsed signals is important for tracking them using as many files as possible from under-sampled areas of the ocean. Automating the VMD method and expanding it to other dolphin species that have very similar pulsive signals will lead to a faster understanding of ecosystem dynamics under a changing climate than can currently be achieved.

Placing underwater recorders throughout the ocean to listen for and track different marine species is a key tool for identifying animals that are shifting habitat ranges as a result of climate change. However, doing this with clicks, buzzes, burst pulses, or pulsed calls (henceforth called "pulsed signals") from dolphins is particularly difficult. The Risso's dolphin (*Grampus griseus*, or "Gg") and Pacific white-sided dolphin (*Lagenorhynchus obliquidens*, or "Lo") are two dolphins that have pulsed signals that look similar in spectrograms, which are the most common visual analysis tool in bioacoustics. Their pulsed signals are above the human hearing range and occur more often then their whistles[8]. Acoustical surveys are a better monitoring tool for tracking population shifts than visual surveys because they can be performed year-round in harsh and remote conditions[9]. Therefore, automated classifiers for pulsed signals in acoustic

data are the way forward for efficiently analyzing acoustic data to track different dolphin species around the globe. It is especially critical for tracking any poleward habitat shifts[3,10,11] so we can prepare for and predict ecosystem-wide effects.

Pulsed signals from both dolphin species (Lo and Gg) have been found in the Arctic Ocean when surface temperatures are warmer in the Gulf of Alaska during year with a negative Pacific Decadal Oscillation (PDO) and a Bering Sea Cold Regime that can foster greater food supply in the pelagic zone[2] where they both hunt. With climate change, the prevalence of Bering Sea Cold Regimes decreases[12], the PDO cycle weakens and shortens[13], and the Arctic Ocean continues to contain less ice[14], so it will be important to track whether both species, only one of the species, or neither species thrive in the new ecosystem(s) into which they are expanding. The effect that each species could have on the Arctic ecosystems could be different, largely because they feed on different types of prey. Risso's dolphins (Gg) mainly feed on squid, octopus, and cuttlefish[15] while Pacific white-sided dolphins (Lo) feed somewhat on squid, but mostly on small schooling fish like salmon[16], capelin, sardine, and herring[17](Fig. 1). All of these prey except octopus and cuttlefish are also commercially important in the Northeast Pacific for human consumption[18] (Fig. 1). If even one species becomes a continuous presence in the Arctic, they will be in direct competition with residential Arctic species that rely on the same food resources. Therefore, being able to differentiate between their pulsed signals to track their movements through the Arctic Ocean is critical for conservation efforts of endemic Arctic species, for understanding the restructuring of the Arctic food web, and for properly calculating Maximum Sustainable Yields for fisheries management.

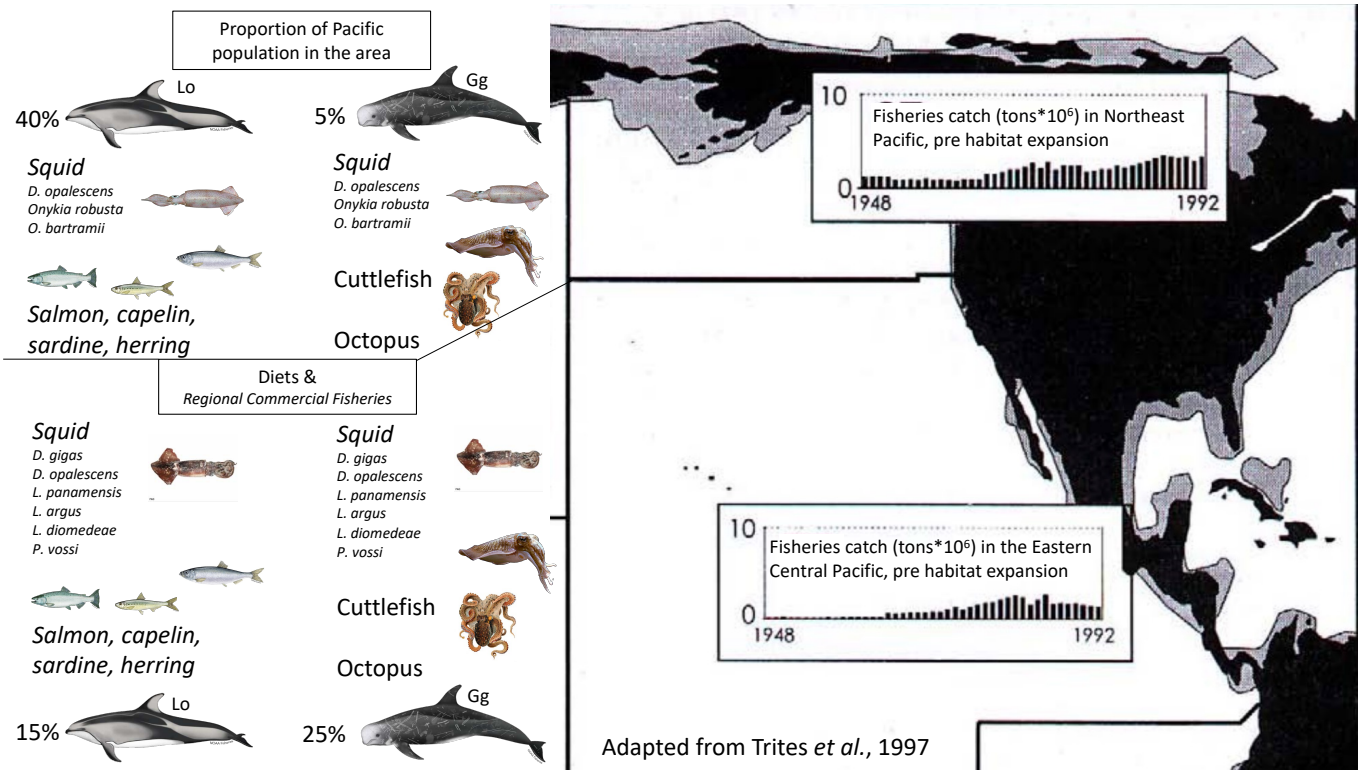

Figure 1. Illustration of overlap between Lo and Gg and their prey. Late 1900s Gg and Lo distributions in two commercially important fishing regions are shown (Northeast and Eastern Central Pacific) with with their diets of commercially important species (italicized) listed. Map adapted from Trites *et al.* 1997[19], fisheries and diet data from several sources[15–18]. Images from NOAA, SeaLifeBase, and Wikipedia.

The pulsed signals of both Gg and Lo have not been thoroughly studied throughout their habitat ranges. A single tank study of four Lo animals demonstrated that individuals produce different peak frequencies up to 7 kHz apart[20]. Only one Gg has been tested in a tank, providing no peak frequency variability measurements between animals, but the single animal produced a wide variety of peak frequencies across its many clicks[21]. The potential for geographical acoustic variability further complicates automated detection and classification of both species beyond the Eastern Tropical Pacific (ETP). But, one study has defined peak and notch patterns in the pulsed signals of each species in the ETP[22] (Fig. 2) and that work has been successfully used to associate click types with behavior in a subsequent study[23], instilling confidence that there are some distinctive peak and notch pattern differences between the two species worth using in our new method.

Table 1 displays the peaks and notches of energy in pulsed signals from Gg and Lo in the ETP that Soldevilla *et al.*[22] documented in their test data. Note the standard deviation around each peak and notch mean measurement and how often each one was actually present in a pulsed signal. The standard distributions overlap between all but the 33.7 kHz (Gg) and 30.5 kHz (Lo) peaks, adding difficulties to the classification (see Fig. 4(f,h)). These peak and notch parameters were established using data files with high signal-to-noise ratio (SNR). In reality, much of the underwater acoustic data that exist contains noise that masks portions of the pulsed signals, and in our sparse dataset, discarding lower quality files is an undesirable option. BANTER software[8] has successfully differentiated between both Gg and Lo signals and got 80% accuracy for Gg and 90% accuracy for Lo by pairing whistles with pulsed signals for contextual clues. However, the sparse dataset with short (4.5 s long) files used in this study only contained two files with whistles and files were spaced up to two minutes apart from one another, so using contextual clues was not an option.

Table 1. Mean, standard deviation (*SD*), and percent of occurrence (*O*) for the peak and notch patterns established by Soldevilla *et al.* (2008)[22]. Adapted from *Test data* part of their Table IV.

Class	Parameters	Peaks #				Notches #		
		1	2	3	4	1	2	3
Gg	*Mean* (*SD*)	22.4 (0.8)	25.5 (1.0)	30.5 (1.1)	38.8 (1.1)	19.6 (1.3)	27.7 (1.1)	35.9 (1.2)
	O %	72	45	82	48	46	64	54
Lo	*Mean* (*SD*)	22.2 (0.6)	26.6 (0.9)	33.7 (1.4)	37.3 (1.4)	19.0 (1.1)	24.5 (0.9)	29.7 (1.4)
	O %	89	76	45	62	51	75	66

Therefore, the goals of creating a new viable classification method for pulsed signals includes: not relying on whistles for context, inherently de-noising the data, not discarding poor-quality recordings, and still reaching accuracy levels similar to previous studies. In this way, a new classification method will have succeeded under the worse-case scenario.

Here we present the *Bayesian VMD Method*, which is a process of (1) a detector using proven FFT techniques to find signals in a sound recording, (2) a Variational Mode Decomposition (VMD) algorithm to extract acoustic features from the detected snippet of sound, and (3) a Bayesian classification weighting system to determine which animal most likely produced the detected sound. The VMD step's output (called Intrinsic Mode Functions or IMFs) is represented in the time-frequency plane via the Hilbert spectrum, denoted here as a *VMD-gram*. The sum of the energies from the *VMD-gram* (which is the counterpart to the FFT-based spectrogram) are used to obtain the spectral content. The features that are extracted from that spectral content are then fed into the Bayesian classifier. The VMD feature extractor, as tested on Gg and Lo pulsed calls in this case study, performed better than FFT-based feature extracting

techniques[22,24]. If expanded to other species with similar calls, this could open the door for more efficient tracking for many animal species.

VMD was chosen because it does not require preliminary de-noising steps, which is a common solution when analyzing low SNR data[25–29], making it a more computationally efficient option. For example, PAMGuard offers a multiple-phase process to denoise data for use with its whistle and moan detector with "click removal" being the first phase[30]. This phase removes pulses like mechanical or human-made sounds, but could also remove echolocation and other odontocete clicks. Clicks being interpreted as noises in some scenarios while as signals in others speaks to the difficulty in trying to remove mechanical pulses from data or in differentiating between them and odontocete clicks when the biological signals are the desired signal for analysis. Thanks to the inherent filtering properties of the VMD's sifting process, which is like a multi-band and adaptive generalization of the Wiener filter[31], the denoising capability of VMD magnifies the biological (or man-made) signals' features over the noise, thus increasing the SNR against background noises like electrical bands or distant ships. Ultimately, man-made pulses do not need to be removed at the expense of accidentally removing some biological signals, but rather can be detected anyway and then sorted into a non-biological category later. For these reasons, the *Bayesian VMD Method* uses an FFT style detector similar to that in PAMGuard since it is tried and true and focuses on using VMD as a feature extractor that also intrinsically de-noises the detected segment of an acoustic recording[31].

Classifying detected signals has been attempted in many ways, like with FFT-style algorithms such as short-time Fourier transforms or Wigner-Ville transforms[32], but their performance deteriorates in poor-quality (low SNR) files and is noted to be computationally expensive[33]. A more efficient FFT method using a two-stage classifier with the cepstrum has worked for differentiating between small dolphins, killer whales, pilot whales, sperm whales, and three species of beaked whales[34]. The bandwidths and frequency patterns of these species' clicks and tones are well-documented in the literature and are visibly different in a spectrogram, though. So the species set used to test the cepstrum method contained signals more distinctive from each other than those of Gg and Lo. There has been success using Gaussian Mixture Models (GMMs) to classify between Lo clicks and burst pulses, common dolphin whistles, and bottlenose dolphin whistles[35], but the Lo pulsed signals were likely classified well with the GMM because they were the most different from the other species in the dataset[35] and were not tested against Gg pulsed signals. No study using FFT methods has combined Gg and Lo pulsed signals without the context of whistles, so a VMD method (the *Bayesian VMD Method*) was attempted for classification between the two species. The *Bayesian VMD Method* was tested against Welch's method (an FFT method) to determine how well it performed comparatively on a sparse dataset with some poor-quality files.

While developing the *Bayesian VMD Method*, a visualization tool was produced that turned out to be more helpful than spectrograms in manually analyzing the data. Like the Wigner Plot[36,37] in PAMGuard that makes it easier for a manual analyst to recognize beaked whale calls, the *VMD-gram* developed in this study is presented as a new visualization tool that can be used in addition to or in place of the spectrogram for manual analysis of pulsed signals with hard-to-differentiate peak and notch patterns. The *VMD-gram* term refers to the visualization of VMD intrinsic components in the time-frequency plane[31].

The *Bayesian VMD Method* has two advantages. First, with the *"VMD-gram"*, the peak and notch patterns are more noticeable to the human eye[4] than if they were displayed in a standard FFT spectrogram. The *VMD-gram*, therefore, has cleaner acoustic features extracted from it, providing cleaner input to a classification algorithm. Second, it uses a probability summation to calculate whether the pulsed signals are more likely Lo or Gg based on parameters set by Soldevilla *et al.* (2008)[22]. This likelihood comparison is crucial for deciding the category of an input signal, especially when some peaks or notches are absent. This probability summation is a strength of the *Bayesian VMD Method* insomuch that it gives a measure of the relative strength of the classification like in the BANTER software[8]. Ground-truthing such classification

algorithms for many odontocete species relies heavily on visual analysis by a human analyst since much of the energy in the pulsed signals is above the human hearing range. This combination of an easier-to-see, de-noised visualization tool with extracted acoustic features from VMD fed into a weighted classifier, reflects the complexity of dolphin calling behavior and are the *Bayesian VMD Method*'s contributions to analyzing acoustic data in a timely manner for tracking the effects of climate change on cetaceans.

Results

Proposed Detector and Classifier Accuracy

With 90 audio files, each 4.5 s long, manual analysis counted 174 distinct signals ascribed to Lo or Gg. These 174 signals plus other pulses from noises like a mooring chain, a pinger, and unknowns contained 1730 groups of energy peaks made up of 4815 individual energy peaks. Our FFT detector grouped energy peaks with an inter-pulse interval threshold of 10 ms for all files, but a 100 ms threshold was applied when click trains were known to be in the file. PAMGuard's click detector does not do a similar grouping, so the manual analysis had to include a count of individual energy peaks for us to best compare it to our FFT detector. The settings in PAMGuard's click detector were explored to generate the best detection rate. The best settings were the Ishmael Energy Sum with 20-50 kHz bounds and 10 dB peak detection threshold with a 20 kHz high pass filter. The EMD detector from previous work[38] detected 977 of the 1730 energy peak groups, achieving an accuracy of 56.47%. To improve upon this we switched to an FFT detector based on PAMGuard's method that detected 1542 of the 1730 energy peak groups, scoring an accuracy of 89.13%, while PAMGuard's click detector detected 3876 of the 4815 energy peaks, scoring an accuracy of 80.5% (Table 2). Scoring an accuracy at least as good as PAMGuard while using inspiration from the authors' previously published methods made us confident in continuing to the next step: classifying the detected signals.

Table 2. Accuracy rates for different detectors

	PAMGuard[30]	EMD[39]	Proposed
Accuracy %	80.5	56.47	89.13

The basis of using VMD in a classifier was due to the fact that previous work[38] that used EMD failed to separate Gg from Lo pulsed signals. Both species' pulsed signals resulted in "EMD identities" of the label "[1,2]". For the sake of comparison, though, precision, recall, and accuracy for how the old EMD classifier performed are provided in Table 3. Specifically, 60 of the 84 (71.43%) Gg pulsed signals were correctly classified as the EMD identity [1,2] and 60 of the 90 (66.67%) Lo pulsed signals were also correctly classified as the EMD identity [1,2]. Pulsed signals that were classified as EMD identities other than [1,2] came from low SNR files or were missing key frequency content. Therefore, with 60 correct [1,2] classifications from known Lo signals and 60 correct [1,2] classifications from known Gg signals, the chance of guessing which species the signals were actually from was a coin toss: 50%. Since some signals were false negatives, accuracy would be worse than a coin toss at 34.48% for each species.

The new classifier using a VMD feature extractor was able to separate Gg from Lo pulsed signals and the results achieved 81.03% overall accuracy (for both Lo and Gg) in contrast to 72.99% overall accuracy for the classifier when using FFT feature extractor results (Table 3). As for the performance of the *Bayesian VMD Method* on each dolphin species, it achieved 88.89% recall on Lo and 72.62% recall on Gg as compared to the FFT feature extractor achieving 72.22% and 73.81%, respectively. The *Bayesian VMD Method* achieved 77.6% precision on Lo and 85.91% precision on Gg, as compared to the FFT feature extractor achieving 74.71% and 71.26%, respectively (Table 3). The FFT feature extractor method

slightly outperformed the *Bayesian VMD Method* only for recall with Gg, but otherwise the *Bayesian VMD Method* outperformed the FFT feature extractor method (Table 3). The *Bayesian VMD Method* was better able to identify Lo signals and its accuracy to identify Gg signals is also better.

Table 3. Precision (**P**), recall (**R**), and total accuracy (**A**) of three classifiers for Gg and Lo: the EMD classifier[39], the FFT-based features extractor classifier, and the VMD-based features extracted classifier. *Denotes accuracy for either Lo or Gg classifications separately.

Class	EMD			FFT-based			VMD-based		
	P %	**R %**	**A %**	**P %**	**R %**	**A %**	**P %**	**R %**	**A %**
Gg	50	71.43	34.48*	71.26	**73.81**	72.99	**85.9**	72.62	**81.03**
Lo	50	66.67		74.71	72.22		**77.6**	**88.89**	

The *VMD-gram* Visualization Tool

The *VMD-gram* is the visualization tool that made manual analysis of peak and notch patterns easier to compare to those published in Soldevilla *et al*. (2008). The noise reduction from the FFT spectrogram to the *VMD-gram* can be seen in Fig. 2. Compared to an FFT-gram (a.k.a. a traditional spectrogram), electrical noise bands have largely been eliminated and the energies that are non-stationary (like those produced by animals) are emphasized. The VMD-gram is obtained by applying the Hilbert transform to the VMD output components. The resulting instantaneous energies and frequencies are transformed into a sparse time-frequency matrix[40]. The sparse matrix representation of the *VMD-gram* makes it easier to visualize.

Processing a set of *VMD-grams* is much quicker, by hand, than by squinting at or zooming into spectrograms. For any species identification that relies on frequency banding patterns, the *VMD-gram* accentuates those compared to a spectrogram. The experience of manually analyzing these *VMD-grams* is similar to the Wigner Plot[36] now used in PAMGuard to quickly identify beaked whale upsweep pulses.

Once the *VMD-gram* is made, the acoustic features that can be extracted from it are also largely void of background noise and peaks and notches become more apparent, making features to feed into classification algorithms more apparent. As an image file, it is also applicable as input to machine learning algorithms such as neural networks.

The Good, The Bad, and The Ugly

To demonstrate the capabilities of this new *Bayesian VMD Method*, three examples of different quality audio data files were chosen to represent the variety of results. These three examples are described as "good", "bad", and "ugly" with regard to their SNR and difficulty to be visually classified in spectrograms. It should be noted that the data segments in Fig. 3 are isolated from the full audio files that contain them so that the length of the very short pulses can be better visualized.

The Good

In the "good" example, the less-than-a-millisecond broadband buzz at 1.8 seconds is the only biological signal in the file. The electrical noise bands, like at 10, 12, 14, 20, and 30 kHz are faint, and there is very little background noise from things like ships, singing baleen whales, seals, storms, or ice, so it is a relatively "clean" file. Using visual inspection, like the manual analysis, it is quite clear that there are peaks in the frequencies associated with Lo pulsed signals. One of these characteristics is the strong red bands centered at about 26.6 and 33.7 kHz. The *VMD-gram* merely accentuated these peaks. The

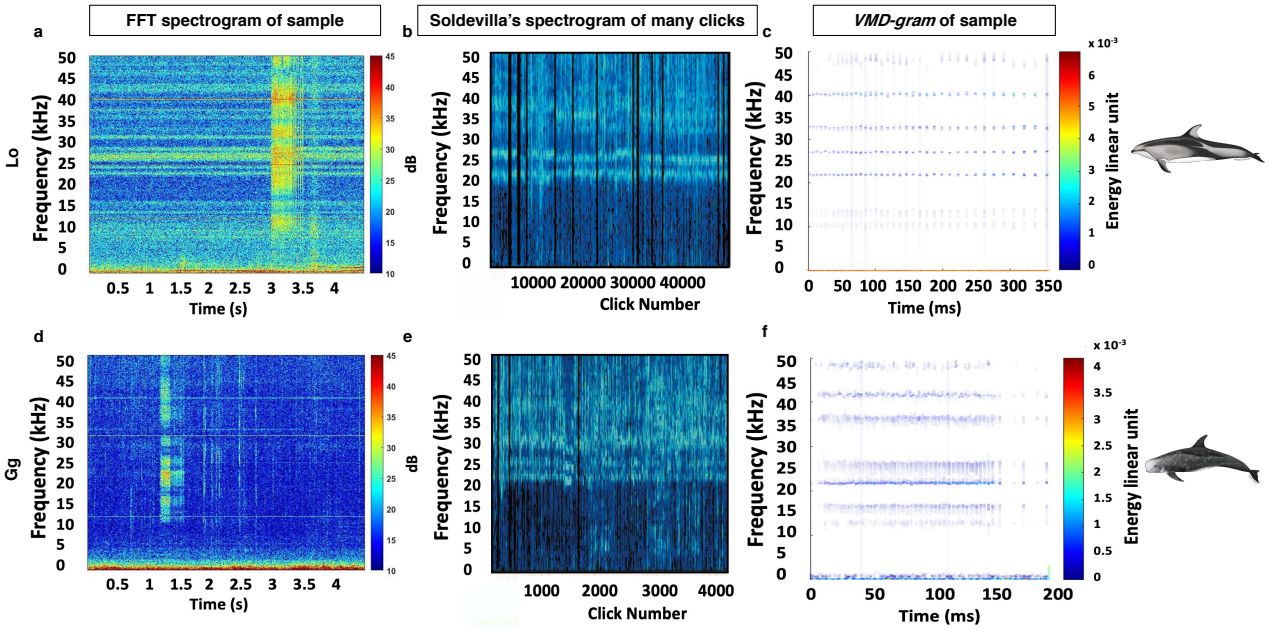

Figure 2. Three visualization tools for delphinid clicks. FFT spectrograms of one sample of a Pacific-white sided dolphin pulsed signal (a) and Risso's dolphin pulsed signal (d) compared to the FFT spectrograms of thousands of clicks from each species by Soldevilla *et al.*, (2008) (b, e) compared to the *VMD-grams* of the segments with pulsed signals (c, f). Ease of visualizing the peak and notch patterns for manual analysis is more evident in the VMD-gram than in the FFT spectrogram.

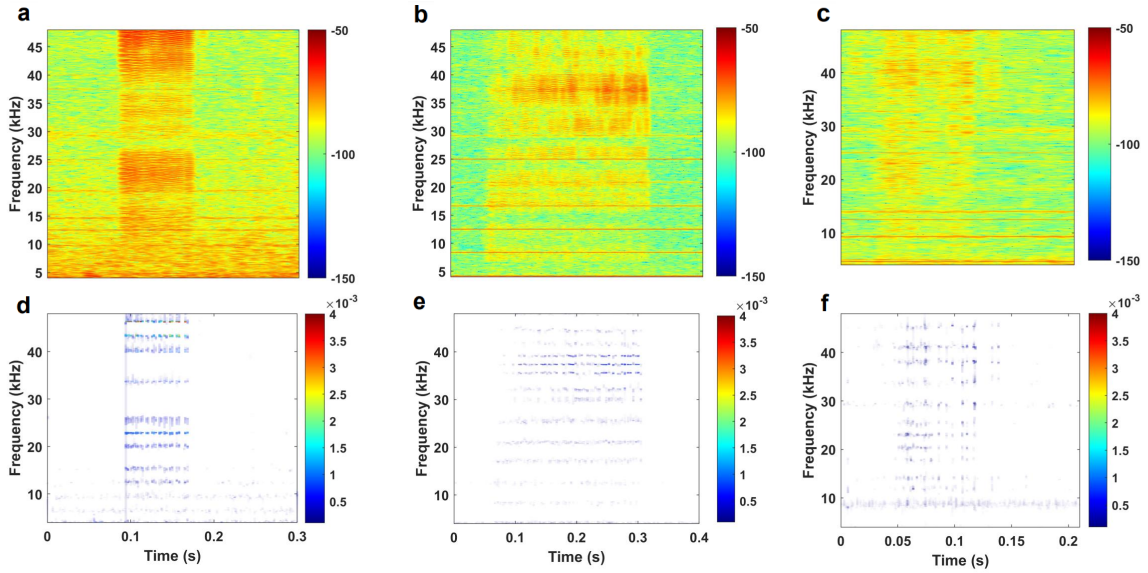

Figure 3. Spectrograms and *VMD-grams* of various data qualities. "Good" (a, d), "bad" (b, e), and "ugly" (c, f) data sample spectrograms (a, b, c) compared to their *VMD-grams* (d, e, f). "Good" means SNR >6 dB and few electrical noise bands; "Bad" means SNR >6 dB and more noise bands or background noise; "Ugly" means SNR <6 dB. Energy bars are in jet color scale where redder is more energy and bluer is less energy.

Bayesian VMD Method classifier agreed with this manual analysis, which means that it should be able to achieve comparable classification results for pulsed signals of similar length and SNR.

The Bad

Before isolating the biological signal present between 3.7 and 3.9 seconds, this "bad" example differentiated itself from the "good" example by having more prominent electrical noise bands (the red horizontal lines every 4.17 kHz) present in the spectrogram. In terms of the biological signals themselves, there are strong peaks around 31, 35.5, and 37 kHz as well as faint peaks around 25.7 and 26.8 kHz. While the peak around 31 kHz is associated with Gg pulsed signals, the 35.5 kHz peak is a strong indicator of Lo pulsed signals[22]. The strongest peak around 37 kHz is indicative of both Gg and Lo pulsed calls[22], so the species classification remains unclear. This signal was difficult to identify visually, but by including species-specific notches in the manual analysis, Lo was chosen as the most likely species for this pulsed signal. The VMD feature extractor disagreed with this, but the weighted values were very similar: 0.608 for Gg versus 0.604 for Lo. The *VMD-gram* showed that peaks around 25 kHz (indicative of Lo) were obscured by electrical noise in the spectrogram. This remains a "bad" example, though, because it is possible the acoustic propagation is different in the Arctic than in the Eastern Tropical Pacific where Soldevilla *et al.* (2008)[22] worked to develop the peak and notch pattern differences. It is also possible that the animals call differently in the two places, meaning the peak and notch patterns could be different in this dataset than in theirs. Regardless of whether the manual classification is correct, the *Bayesian VMD Method* still determined the pulsed call was more likely from one species over another, albeit barely so, using the work from Soldevilla *et al.*[22] as ground truth.

The Ugly

The "ugly" example has the worst combination of factors that make manual analysis using a spectrogram difficult: the overall quality of the signal itself is poor and has low SNR, the electrical noise bands are present, and there is some noise in the lower frequencies of the file. Whereas the "good" and "bad" examples still had at least a 6 dB SNR that would have let them pass the typical quality controls most researchers put in place, the seafoam and yellow colors that are the average in Fig. 3(c,f) makes it almost impossible to see the signal in orange between 0.05 and 0.15 s. In addition to this low SNR, the electrical noise bands, if not removed from the file, could easily become peaks in the *Bayesian VMD Method* that are not related to delphinid pulsed calls at all.

While barely visible over background noise to the eye, there are energy peaks at approximately 38 kHz and in the low 20 kHz. This is not very helpful since both peaks could be evidence for either Gg or Lo. The *Bayesian VMD Method*, however, was able to make more sense of the "ugly" example for the human eye than the spectrogram could. It amplified the peak around 38 kHz and unveiled a peak at about 34 kHz. This latter frequency is associated with Lo pulsed signals[22], allowing for a classification to be made. This shows that the *VMD-gram* is able to uncover information that would normally be hidden by a substantial amount of noise when using a spectrogram. Most acoustic studies institute quality control rules, and this file would fail to pass them, thus getting removed from any dataset that it would be a part of. For sparse datasets, removing files because of poor quality is an impairment to the study, so having a method that could retain files by increasing their visualization quality would be helpful.

Discussion

As climate change occurs, animals that usually occupy temperate habitats shift further towards the poles (a.k.a. northward habitat expansion)[3,10,11], so tracking Lo, Gg, and other delphinid, pulsed-signal-producing species will help address crucial ecological concerns about the restructuring of the Arctic food web. Therefore, we must be able to differentiate between similarly calling species like Lo and Gg to understand how quickly food webs are shifting. Timely analysis of acoustic data will enable conservation efforts to respond more quickly. An algorithm that can eliminate background noise, emphasize frequency content in short signals, and use limited acoustic features to differentiate between species while being computationally efficient would advance the processing of marine mammal signals needed to track habitat expansion of species as the climate changes. Adding a visualization tool to improve manual analysis of pulsed calls would also expedite bioacoustical data processing.

Our *Bayesian VMD Method* and *VMD-gram* meet these advantageous requirements. They were tested on a difficult dataset - one that was sparse, void of whistles, lacked contextual clues beyond 4.5 seconds, and had many low SNR files. These difficulties are the reality for much underwater acoustic data, but future work includes testing this new method on a set of more robust (higher SNR) datasets to determine how well our results generalize. Fig. 4 demonstrates the detection of an Lo signal and the usage of the *VMD-gram* in manual analysis. The promise of the *Bayesian VMD Method* is that it achieved recall, precision, and accuracy values similar to previous work in automated detection and classification while differentiating between two species that produce very similar pulsed signals. The EMD detector from previous work[38] avoided picking most energy peaks generated by noise, but it lacked adaptability and simplicity since multiple tuning parameters had to be manually adjusted for audio files with different levels of background noise. The new detector used in this paper not only automatically estimates the noise floor of each file as an adaptive threshold, but also groups buzzes or clicks based on the inter-click interval of signals of interest (see Fig. 4(a-c)). Lo and Gg were a prime pair of dolphins to test the *Bayesian VMD Method* on because they produce fewer whistles than other species[8] and the peak and notch patterns of

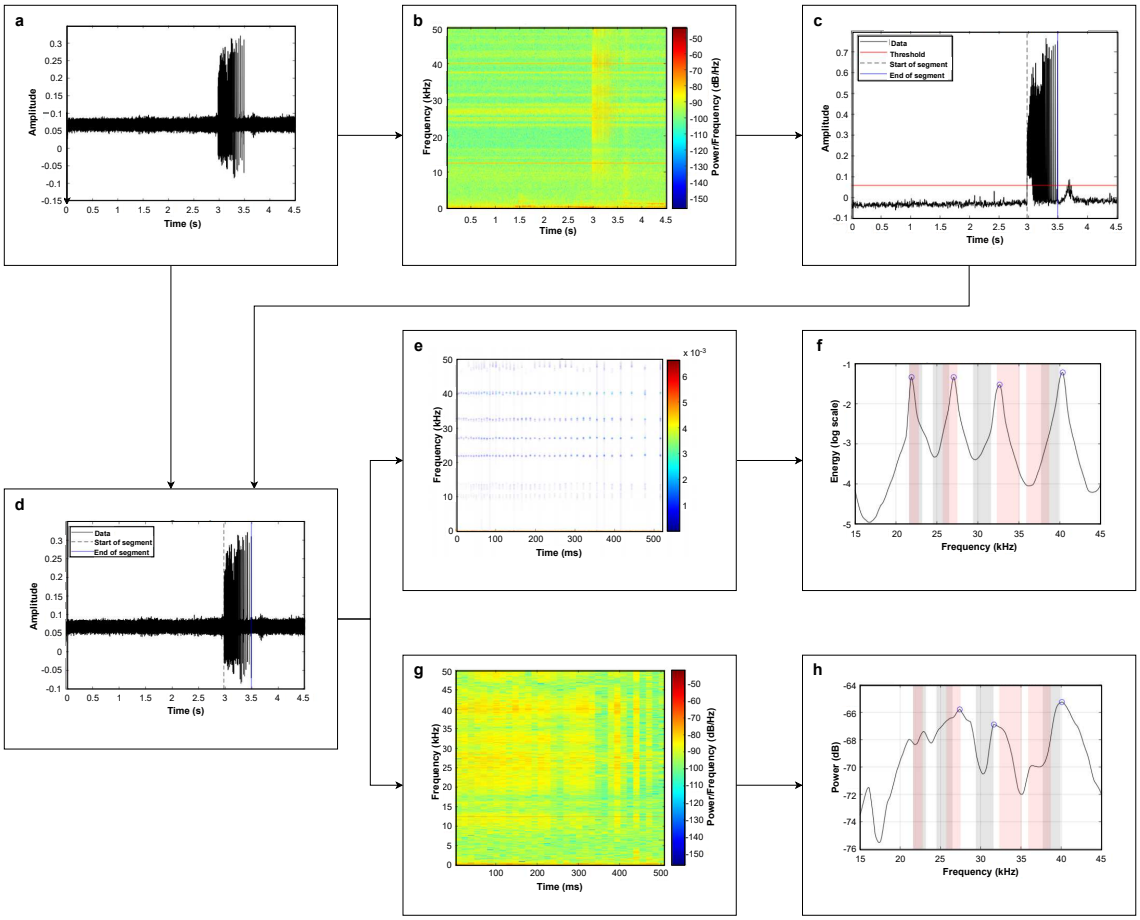

Figure 4. Flowchart demonstrating the usage of the *VMD-gram*. (a) The time-domain of the input signal, (b) the spectrogram of the audio file, (c) application of the threshold and defining time bounds, (d) isolating the significant signal component of the input signal using the time bounds from (c), (e) the *VMD-gram* of the segment, (f) the comparison of the segment to Soldevilla *et al.* 2008, (g) the spectrogram of the segment, and (h) the comparison of the segment to Soldevilla *et al.* (2008). Bars in (f) and (h) indicate the peaks in pulsed signals as documented by Soldevilla *et al.* (2008) for Gg (gray) and Lo (red).

their pulsed calls are not stereotyped across individuals or geography, yet are two important animals to track in the changing Arctic Ocean.

Successful analysis software today perform detection and classification on sparse data because processing is in real-time (like PAMGuard). For sparse data to give reliable real-time results, a few key features in a single event need to be capitalized on and we can not rely on extracting many features in terabytes of data that have been archived for years. Our application of VMD is essentially a distillation process: a small dataset of the good, the bad, and the ugly SNR quality recordings, using a few key features, can still be analyzed equally well between an algorithm and a human analyst. It is not without its needed improvements, though. The *Bayesian VMD Method* may fail when noise (electrical noise, boat noise, etc.) in a recording is stronger than the desired acoustic signals. To improve it, noise with high power could be removed first to get rid of the bulk of background noises. Additionally, the number of IMFs that the signal is decomposed into could be set adaptively depending on the input acoustic file. One difficulty in making the *Bayesian VMD Method* was establishing the best number of IMFs because 9, what we settled on, was

not perfect for every file. Any algorithm or software package that can reliably discern clicks of different species across all the oceans would be a powerful tool in cetacean conservation efforts, and is thus the reason many scientists are focused on advancing such signal processing methods in bioacoustics today. The *Bayesian VMD Method* and *VMD-gram* are two pieces in that pursuit.

Methods

1 Manual Analysis

Data files from previous work[2,41,42] that were manually analyzed (visually validated) and found to contain Lo or Gg pulsed signals were used to test the *Bayesian VMD Method*. No animals were directly involved in this study as all acoustic data were collected passively. Data were collected with passive acoustic listeners (PALs)[43] at a 100 kHz sampling rate with an adaptive duty cycle between 0.75% and 3.75% depending upon whether the PAL's software had detected a signal of interest or not[44]. This created up to twenty-one 4.5 s files per day over each yearlong deployment. Visual validation was done using spectrograms generated by Ulysses software (written by A. Thode; optimized by J. Sarkar). The dataset we used in this study only contained clicks or burst pulses except for two files with faint whistles that were not coincident with pulsed signals, so whistles were not a viable signal for contextual use like in the BANTER[8] study. We used the peak and notch patterns from Soldevilla *et al.* (2008) since they are the best documented pulsed signal characteristics for Gg and Lo and because the geographical range between the ETP groups and the individuals spreading into the Arctic Ocean might overlap.

We used all files that were of good enough quality to manually determine peak and notch patterns to ascribe to a species as well as files from the same day with similar-looking, but poor quality pulsed signals. The poor quality files on their own did not pass quality control guidelines put in place in previous work, so are not considered definitively to be Lo, Gg, or neither. However, they were recorded close in time to good quality files with more certain IDs, so it was likely that the pulsed signals were from the same species. These poor quality files were included in the dataset for VMD accuracy testing to find the point where the *Bayesian VMD Method* failed. If a dataset existed with many good-quality files, then rules set forth by Kowarksi *et al.* (2021)[45] could have been implemented, but a sparse dataset does not have the advantages of "Big Data". Poor quality pulsed signals could have been received from the side of the dolphins' heads or from a distance, so are highly attenuated and variable in both spectral and temporal characteristics[8,46]. It is expected that these poor files would drive down the accuracy of the *Bayesian VMD Method*, but poor quality files are a reality in underwater bioacoustics and manual analysts sift through them regularly, so including them was a better approximation of reality.

For this study, additional manual analysis included visually determining the peak and notch patterns in (1) noisy spectrograms, (2) de-noised spectrograms, and (3) the new *VMD-grams*. If IDs from the three visualizations differed in whether a pulsed signal was "Gg", "Lo", or "too difficult to determine", the ID that was the same for two of the three methods was used as the final ID for the file. These final manual analysis IDs were used as the ground truth set to compare the *Bayesian VMD Method* results against for accuracy and precision/recall calculations. We decided not to use the anecdotally easier *VMD-grams* for the ground truth IDs because using spectrograms for manual analysis has been the historical way to do manual detection and classification work. Note from the results, though, that using the *VMD-gram* IDs as the ground truth improved accuracy. This is because the *VMD-gram* displays the peak and notch patterns more cleanly, thus removing much of the guessing that the human analyst has to do about where the exact peak frequencies appear to be in spectrograms.

2 Detection and Classification System

The *Bayesian VMD Method* is capable of classifying pulsed signals with similar frequency content in poor SNR files from underwater acoustic recordings. The Method consists of two parts. The first part scans the incoming audio data as segments that potentially contain signals of interest by detecting energy peaks. It then uses the start and end of the energy peaks to isolate those areas of interest from non-signal areas of the audio file. The second part classifies the detected signals into separate categories based on their frequency content.

2.1 Detector

The proposed detector uses full audio files that are 4.5 seconds long at a sampling rate of 100 kHz. It then finds audio file segments where potential signals of interest exist.

For a given audio file, denoted by $\hat{x}(n)$, where $n = 1,\ldots,N$, and N is the total number of samples, the Laplacian Differential Operator (LDO) is applied to $\hat{x}(n)$ resulting in an enhanced version of the audio file denoted by $y(n)$, as follows:

$$y(n) = \frac{1}{4}\frac{\partial^2 \hat{x}}{\partial n^2} \tag{1}$$

The LDO enhances the transient signals (edge detection) and filters out the low frequencies (< 10 kHz) which are not needed for Gg and Lo pulsed signal classification. The $y(n)$ is then transformed into a time-frequency representation using Short-time Fourier transform (STFT). The STFT was implemented on 1024 samples with 90% overlap and a 1024-point Hanning window. The magnitude of the STFT matrix $s(n,f)$ is given as $\hat{S}(n,f)$.

$$\hat{S}(n,f) = \begin{bmatrix} s_{11} & \cdots & s_{1p} \\ \vdots & \ddots & \\ s_{m1} & & s_{mp} \end{bmatrix} \tag{2}$$

Next, the dimensionality of matrix $\hat{S}(n,f)$ is reduced from 2-D to 1-D as follows:

$$S_d(n) = \sum_{i=1}^{m} \hat{S}(n,f) \tag{3}$$

The resulting temporal sequence is an accumulated sum of all frequency bins from $s(n,f)$, so scaling is applied, as follows:

$$S_d(n) = \frac{S_d(n)}{max\{S_d(n)\}} \tag{4}$$

After finding $S_d(n)$ from Eq. (4), the mean of $S_d(n)$ is subtracted from it. Then, to determine the boundaries of the acoustic signal, an adaptive threshold is applied. The first step in developing the threshold is to vectorize the matrix $\hat{S}(n,f)$ in column order into a vector called $S_r(n)$:

$$S_r(n) = \overrightarrow{\hat{S}(n,f)} \tag{5}$$

Then, $S_r(n)$ is scaled similar to $S_d(n)$ and is sorted into ascending order, denoted by $\hat{S}_r(n)$. The changing point where the root-mean-square level of the sorted curve $\hat{S}_r(n)$ changes the most is obtained by minimizing Eq. (6)

$$J(k) = \sum_{i=1}^{k-1} \Delta(\hat{S}_{r,i}; \chi([\hat{S}_{r,1} \ldots \hat{S}_{r,k-1}])) + \sum_{i=k}^{N} \Delta(\hat{S}_{r,i}; \chi([\hat{S}_{r,k} \ldots \hat{S}_{r,N}])) \tag{6}$$

where k and N are the index of the changing point and the length of the sorted curve $\hat{S}_r(n)$, respectively, and

$$\sum_{i=u}^{v} \Delta(\hat{S}_{r,i}; \chi([\hat{S}_{r,u} \ldots \hat{S}_{r,v}])) = (u - v + 1) \log \left(\frac{1}{u - v + 1} \sum_{n=u}^{v} \hat{S}_{r,n}{}^2 \right) \tag{7}$$

The threshold, λ, is the value of $\hat{S}_r(k)$ which equals the noise floor estimation, and can be represented as follows:

$$\begin{aligned} \mathcal{H}_0 &: S_d(n) < \lambda \\ \mathcal{H}_1 &: S_d(n) \geq \lambda \end{aligned} \tag{8}$$

where $\mathcal{H}_0$ and $\mathcal{H}_1$ are the hypothesis that the activity was below or above the threshold, respectively. The calculated threshold can vary for each file, thus making it adaptable if ambient noise conditions change between files. The threshold λ is then projected onto the temporal sequence $S_d(n)$ to extract the boundaries of the regions of the acoustic signal that comprised the detected energy peak. The start and end points of each acoustic signal are determined as the first and last points that are greater than λ in amplitude.

The boundaries of the detected segments are scaled by the sampling rate to obtain start and end times which will be used to extract the audio file segments from the original data file in the classification step. Fig. 5 illustrates the layout of the the proposed detector.

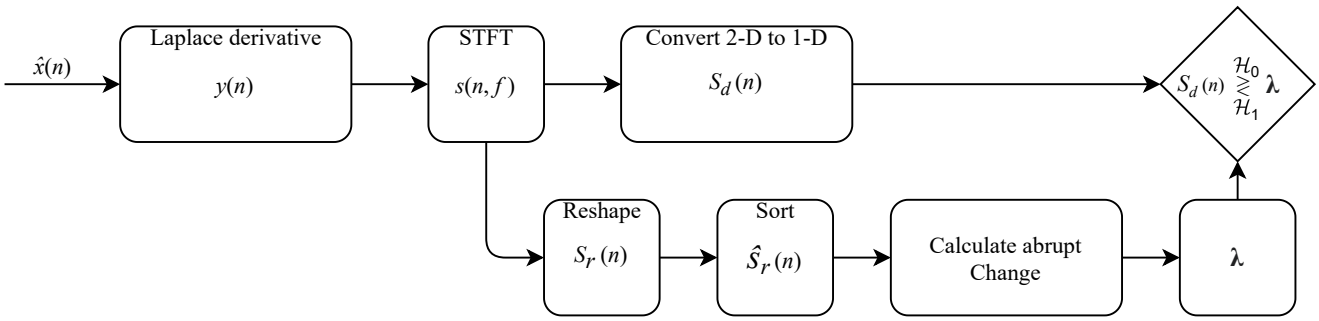

Figure 5. Block diagram of the proposed detector.

2.2 Classifier

Once segments with energy peaks were identified, they were scanned by the team's bioacoustics expert, and any segments confirmed to contain only Gg or Lo signals were sifted out for use in testing the accuracy of the *Bayesian VMD Method* classifier.

In this paper, the metric *weight* was defined for classification purposes. The *weight* for a parameter $\boldsymbol{\theta_i}$ given its measurement $\boldsymbol{y_i}$ is defined as

$$w(\boldsymbol{\theta_i} \mid \boldsymbol{y_i}) = P_{\boldsymbol{\Theta|Y}}(\boldsymbol{\theta_i} \mid \boldsymbol{y_i}) * \boldsymbol{p_i} \tag{9}$$

where, $\boldsymbol{\theta_i}$ is the probability density function (*PDF*) of $\boldsymbol{y_i}$, $\boldsymbol{y_i}$ is one measurement in the measurement vector $\boldsymbol{y}$, $P_{\boldsymbol{\Theta|Y}}(\boldsymbol{\theta_i} \mid \boldsymbol{y_i})$ is the posterior probability of the parameter $\boldsymbol{\theta_i}$ given the measurement $\boldsymbol{y_i}$, and $\boldsymbol{p_i}$ is the scaled prominence value of $\boldsymbol{y_i}$.

When a detected audio file segment is fed into the *Bayesian VMD* classifier, the classification process starts with a feature extraction step. During this step, peak and notch frequencies and their prominence values were obtained from the VMD-Hilbert spectrum of the segment. The prominence measures how much a peak stands out due to its intrinsic height or how much a notch stands out due to its depth and its location relative to surrounding peaks or notches. In general, peaks that are taller and more isolated have a higher "prominence" (p) than peaks that are shorter or surrounded by other peaks.

After VMD, the Hilbert-Huang Transform (HHT) was applied to all IMFs to create a Hilbert spectrum with a frequency resolution of 50 Hz. The Hilbert spectrum is a matrix that contains the instantaneous energy and instantaneous frequency, denoted by $H_s(n, f)$.

$$H_s(n,f) = \begin{bmatrix} h_{11} & \dots & h_{1r} \\ \vdots & \ddots & \\ h_{q1} & & h_{qr} \end{bmatrix} \tag{10}$$

where r is the length of the input segment and q is the number of frequency bins in H_s.

The matrix $H_s(n, f)$ is then converted from a 2-D array to a 1-D spectral representation by summing all instantaneous energy values in each frequency bin, as follows:

$$H_s(f) = \sum_{n=1}^{r} H_s(n,f) \tag{11}$$

The energy summation sequence was converted to a base-10 logarithmic scale and then smoothed by passing through a 17-point median filter and an 11-point moving average filter for the purpose of easily extracting features. All peaks and notches in the sequence whose prominence values exceeded the threshold of 0.5 were located, and their frequency values and prominence values were then stored as extracted features from the input signal (see Fig. 6).

For testing the effectiveness of the VMD feature extractor, a second set of features were extracted from the FFT-based power spectrum using the same input signals with the Welch's algorithm. The FFT-based spectrum was calculated on 2048 samples with 50% overlap and a 2048-point Hanning window with 48.82 Hz frequency resolution. The power spectral density sequence was then converted to dB and went through a 21-point median filter and a 15-point moving average filter. Feature extraction followed the same strategies as in VMD feature extractor except using a prominence threshold of 2 dB.

Next, the measured features, frequencies (Hz) of the peaks and notches (henceforth referred to as "measured peaks and notches"), were matched with the probability distribution functions (PDFs) of peaks and notches (henceforth referred to as "parameter peaks and notches") from Soldevilla *et al.* (2008). The matching between measured and parameter peaks and notches was done in preparation of weight calculations, and it was implemented for both Gg and Lo. There are four Gaussian PDFs for parameter peaks and three for parameter notches for each species in Soldevilla *et al.* (2008) (Table 1). A 95% confidence interval of a Gaussian PDF was used here as a frequency range defined as 1.96 standard deviations to the left and right of its mean value. When measured peaks and notches were matched to parameter peaks and notches, only the peak or notch that fell within a 95% confidence interval were kept. Any peaks or notches outside the 95% confidence intervals were discarded.

Because there are overlaps between the 95% confidence intervals of 22.4 kHz and 25.5 kHz parameter peaks of Gg and between 33.7 kHz and 37.3 kHz parameter peaks of Lo (see Table 1), it is likely that

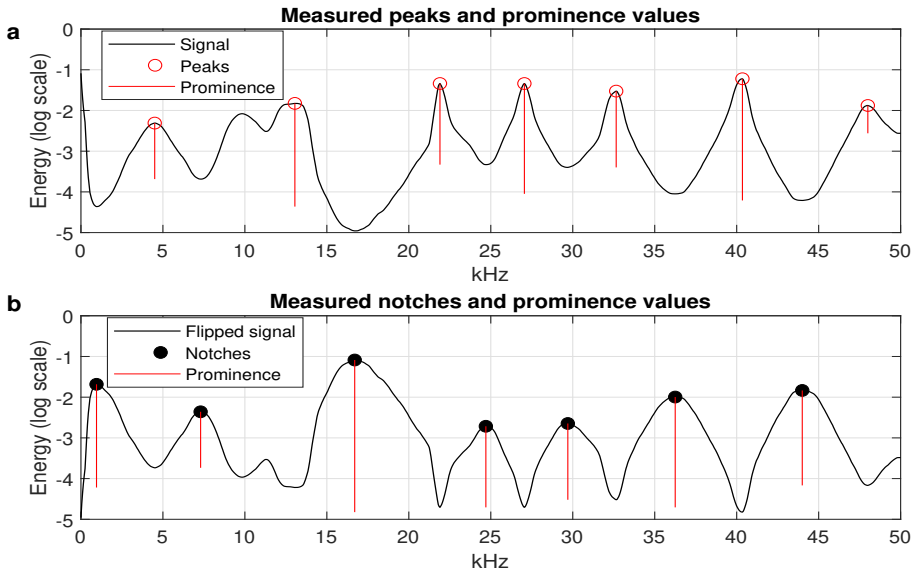

Figure 6. Example of locating peak and notch frequencies and how prominent they are compared to other peaks and notches. The wave form in (a) is the smoothed energy summation sequence from the Hilbert spectrum of the Lo signal in Fig. 2. Subplot (b) is a flipped version of the energy summation sequence for the convenience of extracting notch frequencies and their prominence values. The length of the red line represents the prominence value of a peak or notch.

some measured peaks will fall in the overlapping areas. In this paper, the Maximum A Posterior (MAP) estimation was used to determine which PDF results in the measured peak in an overlapping area. For a measured peak that falls into an overlapping area, two parameter peaks' PDFs are plugged in the MAP estimation equation sequentially, and then the measured peak will be matched with the PDF that maximizes the posterior probability of it given the measured peak.

After the preliminary match, if more than one measured peak or notch remains in any one PDF confidence interval, the measured peak and notch with the highest prominence value is selected as the real measured peak or notch of this PDF, and the redundant ones are discarded. Finally, all remaining peak prominence values and notch prominence values were scaled to be between 0 and 1, respectively.

Once peak and notch matching and selection was finished, Bayesian weights were calculated to select the most likely species. From Bayes's rule, the posterior probability of a parameter given its measurement is proportional to the product of the likelihood function of the measurement given the parameter and the prior probability of the parameter, as shown in Eq. (12).

$$P_{\boldsymbol{\Theta}|\boldsymbol{Y}}(\boldsymbol{\theta_i} \mid \boldsymbol{y_i}) \propto f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y_i} \mid \boldsymbol{\theta_i}) P_{\boldsymbol{\Theta}}(\boldsymbol{\theta_i}) \tag{12}$$

therefore, substitution of the posterior probability in Eq. (9) yields

$$w(\boldsymbol{\theta_i} \mid \boldsymbol{y_i}) = f_{\boldsymbol{Y}|\boldsymbol{\Theta}}(\boldsymbol{y_i} \mid \boldsymbol{\theta_i}) * P_{\boldsymbol{\Theta}}(\boldsymbol{\theta_i}) * \boldsymbol{p_i} \tag{13}$$

With all PDFs and *prior* probabilities from Soldevilla *et al.*(2008), the *weight* value in terms of Gg and Lo given a set of measurements, $\boldsymbol{y}$, was obtained by Eq. (13) and Eq. (14)

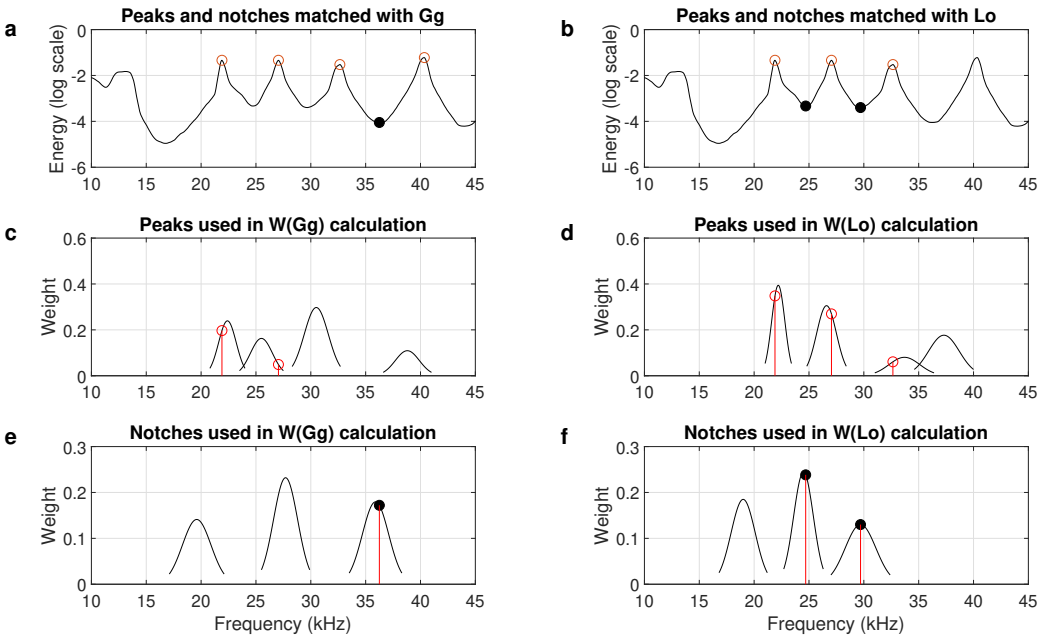

Figure 7. Example of feature matching. The top plots show a set of measured peaks and notches matched with both Gg's PDFs (a) and Lo's PDFs (b) parameter peaks and notches like in Fig. 6 during the feature matching and selection step. Middle plots show how closely to the parameter PDFs that the measured peaks match either Gg (c) or Lo (d) and their weight calculations. The width of each PDF represents its 95% confidence interval, and the ordinate represents the weight value. Subplots (e) and (f) show the same weight calculations for notches. The final weight value is the summation of all weight values of peaks and notches matched with Gg or Lo.

$$w(Gg \mid \boldsymbol{y}) = \sum_{\forall i} w(\boldsymbol{\theta_i} \mid \boldsymbol{y_i}) \qquad w(Lo \mid \boldsymbol{y}) = \sum_{\forall j} w(\boldsymbol{\theta_j} \mid \boldsymbol{y_j}) \tag{14}$$

Where $\boldsymbol{y_i}$ and $\boldsymbol{y_j}$ are the remaining measured peaks and notches that were matched with Gg's PDFs and Lo's PDFs after the matching and matching step. The feature matching and selection results and the weight calculation process are shown in Fig. 7.

The last step was a comparison between weight values in terms of Gg and Lo. If $w(Lo \mid \boldsymbol{y}) > w(Gg \mid \boldsymbol{y})$, the signal was labeled an Lo signal; otherwise, it was labeled a Gg signal. The classifier is illustrated in Fig. 8. The weight values are significant to three digits because weights are normally smaller than 1.000 and three significant digits was sufficient for comparing all calculated weight values for these audio files. In the case that the weight comparison is equal to three significant digits (even though this never happened in these 174 signals), the Bayesian VMD algorithm will automatically classify the input as a Gg signal given that the highest precision (85.91%) by the *Bayesian VMD Method* was achieved on Gg.

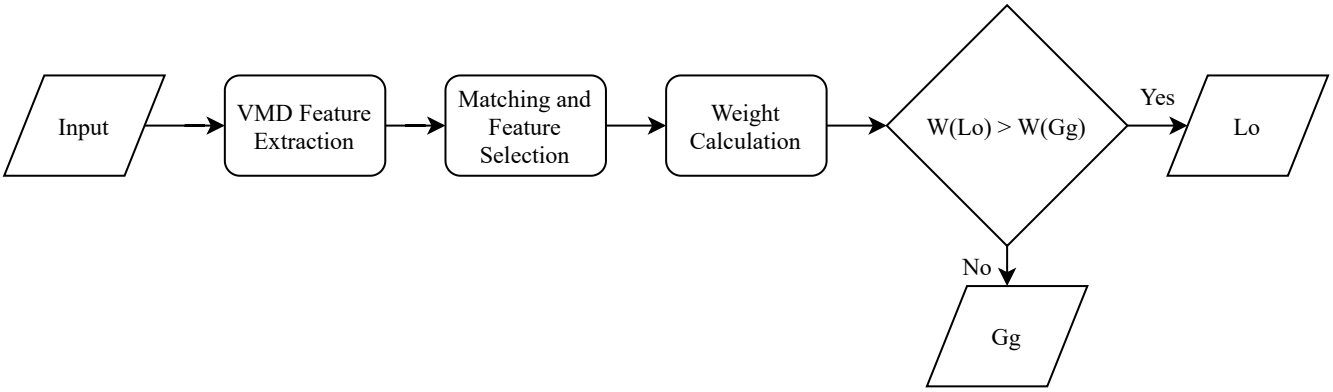

Figure 8. Block diagram of the *Bayesian VMD Method* classifier.

References

1. Würsig, B., Reeves, R. R. & Ortega-Ortiz, J. Global climate change and marine mammals. In *Marine mammals*, 589–608 (Springer, 2002).
2. Seger, K. D. & Miksis-Olds, J. L. Acoustic documentation of temperate odontocetes in the bering and chukchi seas. *Mar. Mammal Sci.* **35**, 1099–1111 (2019).
3. Spies, I. *et al.* Genetic evidence of a northward range expansion in the eastern bering sea stock of pacific cod. *Evol. applications* **13**, 362–375 (2020).
4. Al-Badrawi, M. H., Liang, Y., Kirsch, N. J. & Seger, K. D. Visualization, detection and classification of risso's and pacific white-sided dolphins using an empirical mode decomposition-based process. *The J. Acoust. Soc. Am.* **148**, 2766–2766 (2020).
5. Oswald, J. N., Rankin, S. & Barlow, J. To whistle or not to whistle? geographic variation in the whistling behavior of small odontocetes. *Aquatic Mamm.* **34**, 288–302 (2008).
6. Rankin, S., Oswald, J., Barlow, J. & Lammers, M. Patterned burst-pulse vocalizations of the northern right whale dolphin, lissodelphis borealis. *The J. Acoust. Soc. Am.* **121**, 1213–1218 (2007).

7. Au, W. W., Lammers, M. O. & Yin, S. Acoustics of dusky dolphins (lagenorhynchus obscurus). In *The dusky dolphin*, 75–97 (Elsevier, 2010).

8. Rankin, S. *et al.* Acoustic classification of dolphins in the california current using whistles, echolocation clicks, and burst pulses. *Mar. Mammal Sci.* **33**, 520–540 (2017).

9. Mellinger, D. & Barlow, J. Future directions for acoustic marine mammal surveys: stock assessment and habitat use: report of a workshop held in la jolla, california, 20-22 november 2002. (2003).

10. D'Andrea, L. *et al.* Climate change, anthropogenic disturbance and the northward range expansion of lactuca serriola (asteraceae). *J. Biogeogr.* **36**, 1573–1587 (2009).

11. Carrillo, C., Barbosa, A., Valera, F., Barrientos, R. & Moreno, E. Northward expansion of a desert bird: effects of climate change? *Ibis* **149**, 166–169 (2007).

12. Johnson, J. The effect of cold pool variability on zooplankton dynamics of the eastern bering sea shelf. (2020).

13. Zhang, L. & Delworth, T. L. Simulated response of the pacific decadal oscillation to climate change. *J. Clim.* **29**, 5999–6018 (2016).

14. Comiso, J. C. & Hall, D. K. Climate trends in the arctic as observed from space. *Wiley Interdiscip. Rev. Clim. Chang.* **5**, 389–409 (2014).

15. Cockcroft, V., Haschick, S. & KLAGES, N. W. The diet of risso's dolphin, grampus griseus (cuvier, 1812), from the east coast of south africa. *Zeitschrift für Säugetierkunde* **58**, 286–293 (1993).

16. Kajimura, H., Fiscus, C. H. & Stroud, R. K. Food of the pacific white-sided dolphin, lagenorhynchus obliquidens, dall's porpoise, phocoenoides dalli, and northern fur seal, callorhinus ursinus, off california and washington, with appendices of size and food of dall's porpoise from alaskan waters. (1980).

17. Morton, A. Occurrence, photo-identification and prey of pacific white-sided dolphins (lagenorhyncus obliquidens) in the broughton archipelago, canada 1984–1998. *Mar. Mammal Sci.* **16**, 80–93 (2000).

18. Arkhipkin, A. I. *et al.* World squid fisheries. *Rev. Fish. Sci. & Aquac.* **23**, 92–252 (2015).

19. Trites, A. W., Christensen, V. & Pauly, D. Competition between fisheries and marine mammals for prey and primary production in the pacific ocean. *J. Northwest Atl. Fish. Sci.* **22**, 173–187 (1997).

20. Fahner, M. A. *Acoustic properties of echolocation signals by captive Pacific white-sided dolphins (Lagenorhynchus obliquidens)*. Ph.D. thesis, Western Illinois University (1999).

21. Philips, J. D., Nachtigall, P. E., Au, W. W., Pawloski, J. L. & Roitblat, H. L. Echolocation in the risso's dolphin, grampus griseus. *The J. Acoust. Soc. Am.* **113**, 605–616 (2003).

22. Soldevilla, M. S. *et al.* Classification of risso's and pacific white-sided dolphins using spectral properties of echolocation clicks. *The J. Acoust. Soc. Am.* **124**, 609–624 (2008).

23. Henderson, E. E., Hildebrand, J. A. & Smith, M. H. Classification of behavior using vocalizations of pacific white-sided dolphins (lagenorhynchus obliquidens). *The J. Acoust. Soc. Am.* **130**, 557–567 (2011).

24. Gillespie, D. *et al.* Pamguard: Semiautomated, open source software for real-time acoustic detection and localisation of cetaceans. *J. Acoust. Soc. Am.* **30**, 54–62 (2008).

25. Roch, M. A. *et al.* Classification of echolocation clicks from odontocetes in the southern california bight. *The J. Acoust. Soc. Am.* **129**, 467–475 (2011).

26. Mallawaarachchi, A., Ong, S., Chitre, M. & Taylor, E. Spectrogram denoising and automated extraction of the fundamental frequency variation of dolphin whistles. *The J. Acoust. Soc. Am.* **124**, 1159–1170 (2008).

27. Patil, B., Shastri, R. & Das, A. Wavelet denoising with ica for the segmentation of bio-acoustic sources in a noisy underwater environment. In *2014 International Conference on Communication and Signal Processing*, 472–475 (IEEE, 2014).

28. Ingale, C. B. & Lokhande, S. S. Habitat impact on echo-location characteristics of irrawaddy dolphins from chilika lake and sunderbans. *Int. J. Sci. Res.* **4**, 2249–2252 (2015).

29. Sugimatsu, H. *et al.* Advanced technique for automatic detection and discrimination of a click train with short interclick intervals from the clicks of ganges river dolphins (platanista gangetica gangetica) recorded by a passive acoustic monitoring system using hydrophone arrays. *Mar. Technol. Soc. J.* **48**, 167–181 (2014).

30. Gillespie, D., White, P., Caillat, M. & Gordon, J. Development and implementation of automatic classification of odontocetes within pamguard. In *Workshop on Detection, Classification, Localization, and Density Estimation of Marine Mammals using Passive Acoustics Timberline Lodge, Mt. Hood, Oregon, USA* (2011).

31. Dragomiretskiy, K. & Zosso, D. Variational mode decomposition. *IEEE transactions on signal processing* **62**, 531–544 (2013).

32. Le Bras, R. & Sucic, V. Individual blue whale recognition. wigner-ville time-frequency analysis and preparation for a kaggle contest. In *CTBTO 2013 S&T Conference* (2013).

33. Gillespie, D. & Caillat, M. Statistical classification of odontocete clicks. *Can. Acoust.* **36**, 20–26 (2008).

34. Martin, B., Mouy, X., Gaudet, B. & Kowarski, K. Differentiating marine mammal clicks using time-series properties. *The J. Acoust. Soc. Am.* **146**, 2886–2886 (2019).

35. Roch, M. A., Soldevilla, M. S., Burtenshaw, J. C., Henderson, E. E. & Hildebrand, J. A. Gaussian mixture model classification of odontocetes in the southern california bight and the gulf of california. *The J. Acoust. Soc. Am.* **121**, 1737–1748 (2007).

36. Griffiths, E. T. & Barlow, J. Cetacean acoustic detections from free-floating vertical hydrophone arrays in the southern california current. *The J. Acoust. Soc. Am.* **140**, EL399–EL404 (2016).

37. Jacobson, E. K., Yack, T. M. & Barlow, J. Evaluation of an automated acoustic beaked whale detection algorithm using multiple validation and assessment methods. (2013).

38. Seger, K. D., Al-Badrawi, M. H., Miksis-Olds, J. L., Kirsch, N. J. & Lyons, A. P. An empirical mode decomposition-based detection and classification approach for marine mammal vocal signals. *The J. Acoust. Soc. Am.* **144**, 3181–3190 (2018).

39. Seger, K. D., Al-Badrawi, M. H., Miksis-Olds, J. L., Kirsch, N. J. & Lyons, A. P. An empirical mode decomposition-based detection and classification approach for marine mammal vocal signals. *The J. Acoust. Soc. Am.* **144**, 3181–3190 (2018).

40. Huang, N. E. *Hilbert-Huang transform and its applications*, vol. 16 (World Scientific, 2014).

41. Seger, K. D., Miksis-Olds, J. & Martin, B. A preliminary acoustical survey of echolocating marine mammals in the bering sea. In *Proceedings of Meetings on Acoustics 172ASA*, vol. 29, 010005 (Acoustical Society of America, 2016).

42. Seger, K. D. & Miksis-Olds, J. L. A decade of marine mammal acoustical presence and habitat preference in the bering sea. *Polar Biol.* **43**, 1549–1569 (2020).

43. Nystuen, J. A. Temporal sampling requirements for automatic rain gauges. *J. Atmospheric Ocean. Technol.* **15**, 1253–1260 (1998).

44. Denes, S. L., Miksis-Olds, J. L., Mellinger, D. K. & Nystuen, J. A. Assessing the cross platform performance of marine mammal indicators between two collocated acoustic recorders. *Ecol. informatics* **21**, 74–80 (2014).

45. Kowarski, K. A. & Moors-Murphy, H. A review of big data analysis methods for baleen whale passive acoustic monitoring. *Mar. Mammal Sci.* **37**, 652–673 (2021).

46. Au, W. W. *The sonar of dolphins* (Springer Science & Business Media, 1993).

Acknowledgements (not compulsory)

Thank you to Dr. Jennifer Miksis-Olds for continuing to let us use the Arctic Corridor dataset. All NOAA and other employees and scientists who collaborated in the EcoFOCI program by deploying and recovering the acoustic recorders for a decade have facilitated collection of an invaluable dataset. This work was funded by the Office of Naval Research award number N000142012620.

Author contributions statement

M.B. conceived the experiment, conducted the experiment, contributed to writing code for the experiment, analyzed the results, assisted in writing the manuscript, and made figures. K.S. conceived the experiment, did manual analysis of files for the experiment, managed the team, assisted in writing the manuscript, and made figures. Y.L. conducted the experiment, wrote the automation code for the experiment, analyzed the results, assisted in writing the manuscript, and made figures. C.F. analyzed the results, assisted in writing the manuscript, and made figures. N.K. advised students during the experiment, is listed as PI on the project this experiment is a part of, analyzed the results, and assisted in writing the manuscript.

All authors reviewed the manuscript.

Additional information

The author(s) declare no competing interests.