JASA ARTICLE

# Multi-target 2D tracking method for singing humpback whales using vector sensors

Ludovic Tenorio-Hallé,[1,a)] Aaron M. Thode,[1] Marc O. Lammers,[2] Alexander S. Conrad,[3] and Katherine H. Kim[3]

[1]Marine Physical Laboratory, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093-0238, USA

[2]Hawaiian Islands Humpback Whale National Marine Sanctuary, 726 S. Kihei Rd, Kihei, Hawaii 96753, USA

[3]Greeneridge Sciences, Inc., 5266 Hollister Avenue, Suite 107, Santa Barbara, California 93111, USA

**ABSTRACT:**

Acoustic vector sensors allow estimating the direction of travel of an acoustic wave at a single point by measuring both acoustic pressure and particle motion on orthogonal axes. In a two-dimensional plane, the location of an acoustic source can thus be determined by triangulation using the estimated azimuths from at least two vector sensors. However, when tracking multiple acoustic sources simultaneously, it becomes challenging to identify and link sequences of azimuthal measurements between sensors to their respective sources. This work illustrates how two-dimensional vector sensors, deployed off the coast of western Maui, can be used to generate azimuthal tracks from individual humpback whales singing simultaneously. Incorporating acoustic transport velocity estimates into the processing generates high-quality azimuthal tracks that can be linked between sensors by cross-correlating features of their respective azigrams, a particular time-frequency representation of sound directionality. Once the correct azimuthal track associations have been made between instruments, subsequent localization and tracking in latitude and longitude of simultaneous whales can be achieved using a minimum of two vector sensors. Two-dimensional tracks and positional uncertainties of six singing whales are presented, along with swimming speed estimates derived from a high-quality track. *© 2022 Acoustical Society of America.* https://doi.org/10.1121/10.0009165

(Received 30 September 2021; accepted 6 December 2021; published online 6 January 2022)

[Editor: Klaus Lucke] Pages: 126–137

## I. INTRODUCTION

Each winter, humpback whales (*Megaptera novaeangliae*) congregate in their respective breeding grounds in tropical waters around the globe. At these gatherings, males produce long sequences of structured vocalizations referred to as songs (Payne and McVay, 1971; Au *et al.*, 2006). Song is believed to play an important function in the mating system of humpback whales by signaling to females and mediating male–male interactions (Darling *et al.*, 2006; Herman, 2017; Cholewiak *et al.*, 2018). Despite having been studied since the early 1970s, many aspects of humpback whale singing remain poorly understood and the exact way in which song mediates whale interactions is still unclear.

While the recent advancements of bioacoustic tags have allowed documenting fine-scale acoustic behavior of specific individuals in many marine mammal species (Schmidt *et al.*, 2010; Goldbogen *et al.*, 2014; Stimpert *et al.*, 2020), tagging remains logistically expensive, yields low sample sizes, and does not always allow attribution of detected calls to the tagged whale. Passive acoustic localization and tracking is a complementary approach to tagging as it is non-invasive and allows addressing ongoing research questions on a larger scale (Noad *et al.*, 2004; Schmidt *et al.*, 2010; Stanistreet *et al.*, 2013; Helble *et al.*, 2015; Helble *et al.*, 2016; Guazzo *et al.*, 2017; Henderson *et al.*, 2018). If few whales are present at a time, a network of widely spaced hydrophones can be used to track their position using time-of-arrival techniques on calls that are detected on at least three sensors (Schau and Robinson, 1987; Spiesberger, 2001). However, when large numbers of singing humpback whales are present, such as on their breeding grounds off Hawaii (Au *et al.*, 2000), their songs dominate the ambient noise field. Because the overlap in time and in frequency hinders signal extraction from individual whales, conventional localization techniques (e.g., time-domain cross correlation between sensors) become impractical.

Developed during the second half of the 20th century, acoustic vector sensors were originally used in U. S. Navy operations for detecting and localizing submarines, primarily through their use of directional frequency analysis recording (DIFAR) sonobuoys (Holler, 2014). In recent years, they have been developed into commercial recording packages and are used for a variety of passive acoustic monitoring applications (Greene *et al.*, 2004; Raghukumar *et al.*, 2020). In addition to measuring acoustic pressure like conventional hydrophones, vector sensors also measure particle motion, allowing them to estimate the dominant direction of travel of acoustic energy from a single point (D'Spain *et al.*, 1991; D'Spain *et al.*, 2006; Martin *et al.*, 2016). This ability permits triangulation of acoustic sources using multiple vector sensors. In addition to requiring a minimum of only two

[a)]Electronic mail: ltenorio@ucsd.edu

sensors, this approach is also highly advantageous when compared to time-of-arrival techniques required by conventional hydrophones, as triangulation does not require precise time-synchronization between independent autonomous sensor packages. However, while triangulation is relatively straightforward to execute and automate in the presence of few sources producing stereotyped signals (Greene *et al.*, 2004; Thode *et al.*, 2012; Thode *et al.*, 2021), it becomes challenging to apply when trying to track multiple non-stereotyped sources, such as singing humpback whales.

Data from vector sensors can be processed in a variety of ways, including additive beamforming of the pressure and velocity channels (McDonald, 2004), or by multiplicative processing, where the pressure and velocity channels are multiplied together to model the acoustic intensity (D'Spain *et al.*, 1991). The speed of the latter approach makes it possible to display the dominant directionality of an ambient noise field as a function of time and frequency, a display that has been exploited many times and has recently been nicknamed an "azigram" when applied to biological studies (Thode *et al.*, 2019). One of the key advantages of the azigram versus a conventional spectrogram is that it allows individual sources to be distinguished based on their location. Another key vector sensor metric, the acoustic transport velocity, also describes the directional properties of an acoustic field. This quantity, which will be defined in Sec. II, gives insight into the spatial distribution of the acoustic sources generating an ambient acoustic field. It allows differentiation between directional sources (such as a whale or a boat) and diffuse background ambient sound that arises from numerous simultaneous sources from multiple directions (D'Spain *et al.*, 1991; Schiffrer and Stanzial, 1994).

In this study, we exploit the directional capabilities of vector sensors to track multiple singing humpback whales simultaneously, despite substantial overlap in their songs. We demonstrate this method on data collected off the coast of western Maui during the late breeding season in 2020, using modified DIFAR vector sensors (which measure horizontal particle velocity in two orthogonal directions) that enable two-dimensional (latitude and longitude) localization and tracking of whales. Section II presents the theory related to vector sensors relevant to this work, including the key metrics of dominant directionality and transport velocity. Section III gives an overview of the instruments, the deployment, and the demonstration dataset. Section IV details the tracking method, along with examples and results illustrating the different steps of the algorithm. Finally, Sec. V presents tracking results for six whales and derives the track-derived swimming speed for one singer.

## II. VECTOR SENSOR THEORY

Different quantities can be used to describe an acoustic field. In underwater acoustics, the most prevalent metric is acoustic pressure, which is easily measured underwater with a conventional hydrophone. This scalar quantity represents the amount of compression between particles in the medium and is also the quantity measured by both terrestrial and marine mammal ears (Popper *et al.*, 2000). However, measurements of acoustic pressure alone are insufficient for uniquely characterizing an acoustic wave traveling through a point. Also, required are measurements of particle motion, a vector that quantifies the movement of the particles as sound travels through the medium, and can be expressed in terms of displacement, velocity, or acceleration. Vector sensors are designed to measure both acoustic pressure and particle velocity along two or three orthogonal axes.

The instantaneous acoustic intensity along a given axis $k$ is defined as

$$I_k = pv_k, \tag{1}$$

where $p$ and $v_k$ are the time series of acoustic pressure and particle velocity along axis $k$. If the acoustic field is comprised by a single plane wave arriving from a distant, dominant, and spatially compact source, then the magnitude of the particle velocity is proportional to and in phase with the pressure. Equation (1) then reduces to a form where the squared pressure alone yields the intensity magnitude. However, since vector sensors measure pressure and particle motion independently, they provide direct measurements of the true underlying acoustic intensity, even in circumstances where the acoustic field is not dominated by a single plane wave.

The frequency-domain acoustic intensity $S_k$ can be estimated at time-frequency bin $(T,f)$ as

$$S_k(T,f) = \langle P(T,f)V_k^*(T,f)\rangle \equiv C_k(T,f) + iQ_k(T,f), \tag{2}$$

where P and $V_k$ are short-time fast Fourier transforms (FFTs) of $p$ and $v_k$, respectively (Mann *et al.*, 1987). The symbol $*$ denotes the complex conjugate of a complex number, and $\langle\ \rangle$ represents the ensemble average of a statistical quantity. If a time series can be considered to be statistically ergodic over a given time interval, this ensemble average can be obtained from time-averaging consecutive FFTs (D'Spain *et al.*, 1991). In practice, ambient acoustic fields are often highly nonstationary, but a short enough time interval can typically be found where the ergodicity assumption is valid. In Eq. (2), $C_k$ and $Q_k$ are defined as the active and reactive acoustic intensities, respectively, and they comprise the in-phase and in-quadrature components of the pressure and particle velocity. The active intensity $C_k$ comprises the portion of the field where pressure and particle velocity are in phase and thus, are transporting acoustic energy through the measurement point. The reactive intensity $Q_k$ comprises the portion of the field where pressure and particle velocity are 90° out of phase and arises whenever a spatial gradient exists in the acoustic pressure (Mann *et al.*, 1987). For the rest of this paper, we ignore the reactive component of intensity, and use the active component to define two directional metrics: the dominant azimuth and the normalized transport velocity (NTV).

In the case of a two-dimensional vector sensor that measures particle velocity along the $x$ and $y$ axis, the dominant azimuth from which acoustic energy is arriving, $\varphi$, is defined as

$$\varphi(T,f) = \tan^{-1}\frac{C_x(T,f)}{C_y(T,f)}, \tag{3}$$

where $\varphi$ is expressed in geographical terms: increasing clockwise and starting from the $y$ axis. The dominant azimuth can then be displayed as a function of both time and frequency as an image nicknamed an "azigram" (Thode *et al.*, 2019).

It is important to note that Eq. (3) estimates only the *dominant* azimuth since acoustic energy may be arriving from different azimuths simultaneously at the measurement point. Equation (3) effectively represents an estimate of the "center of mass" of the transported energy but provides no information about its angular distribution around the sensor. The normalized transport velocity (NTV) is a quantity that provides this second order information about the acoustic field. For the same two-dimensional vector sensor assumed for Eq. (3), the NTV is defined by the ratio between the active intensity and the energy density of the field

$$U(T,f) = \frac{2\rho_0 c\left\langle\left[C_x^2(T,f)+C_y^2(T,f)\right]^{1/2}\right\rangle}{\rho_0^2c^2\left\langle\left|V_x(T,f)\right|^2+\left|V_y(T,f)\right|^2\right\rangle+\left\langle\left|P(T,f)\right|^2\right\rangle}, \tag{4}$$

where $\rho_0$ and $c$ are the density and sound speed in the medium, respectively (Mann *et al.*, 1987; D'Spain *et al.*, 1991). Equation (4) is normalized such that the NTV lies between 0 and 1. Although ideally, the NTV should be computed using particle velocity measurements along all three spatial axes, when measuring low-frequency sound in a shallow-water acoustic waveguide only a small fraction of the total acoustic energy is transported vertically (along the z axis) into the ocean floor. Under these circumstances, a relatively accurate NTV can be obtained on a two-dimensional sensor using only particle velocity measurements along the horizontal axes. A NTV close to 1 implies that most of the acoustic energy traveling through the measurement point is clustered around the dominant azimuth. Such would be the case for a single azimuthally compact source, such as a whale or a ship whose signal-to-noise ratio (SNR) is high. By contrast, a NTV of 0 indicates that no net acoustic energy is being transported through the measurement point, which implies either no acoustic energy is present at all, or equal amounts of energy are being propagated from opposite directions, as is the case for a standing wave. Thus, low transport velocity occurs in the presence of ambient fields that are either isotropic or azimuthally symmetric.

## III. ILLUSTRATIVE DATASET: MAUI 2020

A DASAR (directional autonomous seafloor acoustic recorder) model "C" is an autonomous underwater recording package equipped with a DIFAR vector sensor, which is itself composed of an omnidirectional pressure sensor (149 dB re 1 $\mu$Pa/V at 100 Hz sensitivity) and two particle motion sensors capable of measuring the $x$ and $y$ components of particle velocity (Greene *et al.*, 2004; Thode *et al.*, 2012). The signals measured on each of the three channels were sampled at 1 kHz with sensors that have a maximum measurable acoustic frequency of 450 Hz.

The sensitivity of the directional channels, when expressed in terms of plane wave acoustic pressure (–243.5 dB re m/s equates to 0 dB re 1 $\mu$Pa), is 146 dB re 1 $\mu$Pa/V at 100 Hz. The sensitivity of all channels increases by +6 dB/octave (e.g., the sensitivity of the omnidirectional channel is 143 dB re 1 $\mu$Pa/V at 200 Hz), since the channel inputs are differentiated before being recorded. These values were measured from two DASARs calibrated at the U.S. Navy's underwater acoustic test facility TRANSDEC in San Diego in 2008. A finite impulse response (FIR) equalization filter was applied to recorded data to recover the original spectrum.

Between March and July 2020, three DASARs labeled A, B, and C were deployed along the south facing coast of western Maui, capturing the last couple of months of the humpback whale breeding season. The instruments were spaced by approximately 3 km in a line running from the northwest (DASAR A) to the southeast (DASAR C) as shown in Fig. 1 at depths of approximately 20 m (Google). The DASARs were lowered to the ocean floor from a small vessel using a rope, and thus, the orientation of the package on the ocean floor could not be controlled and had to be measured acoustically. Using the same calibration technique as Thode *et al.* (2021), the small vessel was driven clockwise and counterclockwise around each DASAR after its deployment. From the global positioning system (GPS) position of the boat and the associated estimated acoustic azimuths, the clock offset between the GPS and the sensor data could be inverted, along with the seafloor orientation of the sensors' particle velocity axes. Additionally, this procedure was used to estimate a 7.61° median uncertainty for dominant azimuth estimates. The details of the orientation calibration are discussed in the Appendix.

In this study, 24 h of data starting from midnight on April 18, 2020, are presented and analyzed. This time window, which has fewer whales present than earlier in the breeding season, was chosen because individual tracks are easily visually distinguishable.

## IV. TRACKING ALGORITHM

### A. Time-frequency representation of directional metrics

As shown in Eqs. (3) and (4), both the dominant azimuth $\varphi$ and NTV can be associated with each time-frequency bin $(T,f)$ of a spectrogram, allowing these quantities to be displayed as an image. In the dominant azimuth representation–the azigram–the color of each pixel is associated with a given geographical azimuth. In the NTV

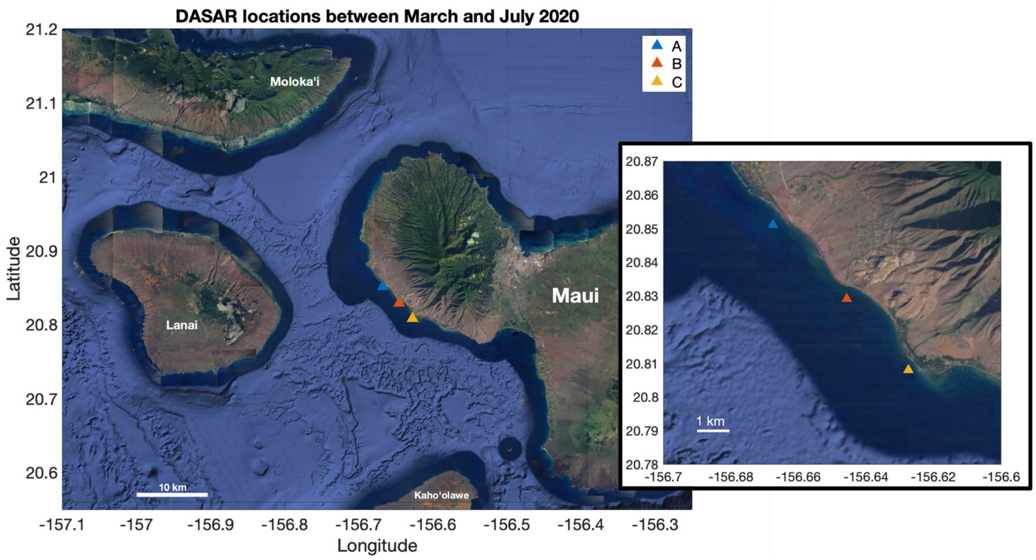

FIG. 1. (Color online) Satellite image indicating the position of DASARs A, B and C, deployed between March and July 2020 off the western coast of Maui. (Reprinted from source: Google).

representation, the color of each pixel corresponds to a value between 0 and 1. Figure 2 shows the spectrogram, azigram, and NTV of 30 s of data from DASAR B, starting at 2:24 UTC-10 on April 18, 2020. While the spectrogram suggests the presence of multiple humpback whales singing simultaneously, it does not allow straightforward association of song units to individual whales [Fig. 2(a)]. The azigram display, however, reveals distinct individual whales based on their color/azimuth [Fig. 2(b)]. In this plot, the color scale has been restricted between 100° and 350° azimuth, an arrival sector that points away from shore (see Fig. 1). The NTV time-frequency representation [Fig. 2(c)] shows that whale calls have high NTV values, as would be expected from a spatially compact acoustic source.

### B. Identifying azimuthal tracks over long intervals

The number of singing whales and their azimuths can be estimated at any given time from the statistical distribution of $\varphi$. Let $h_\theta(\Delta T_h)$ be defined as a histogram that counts the number of observations of $\varphi(T,f)$ that fall within azimuthal bin of center $\theta$ and width $d\theta$ within a time interval $\Delta T_h$. Thus, $h_\theta$ estimates the distribution of azimuths measured across all time-frequency bins in the azigram. Note

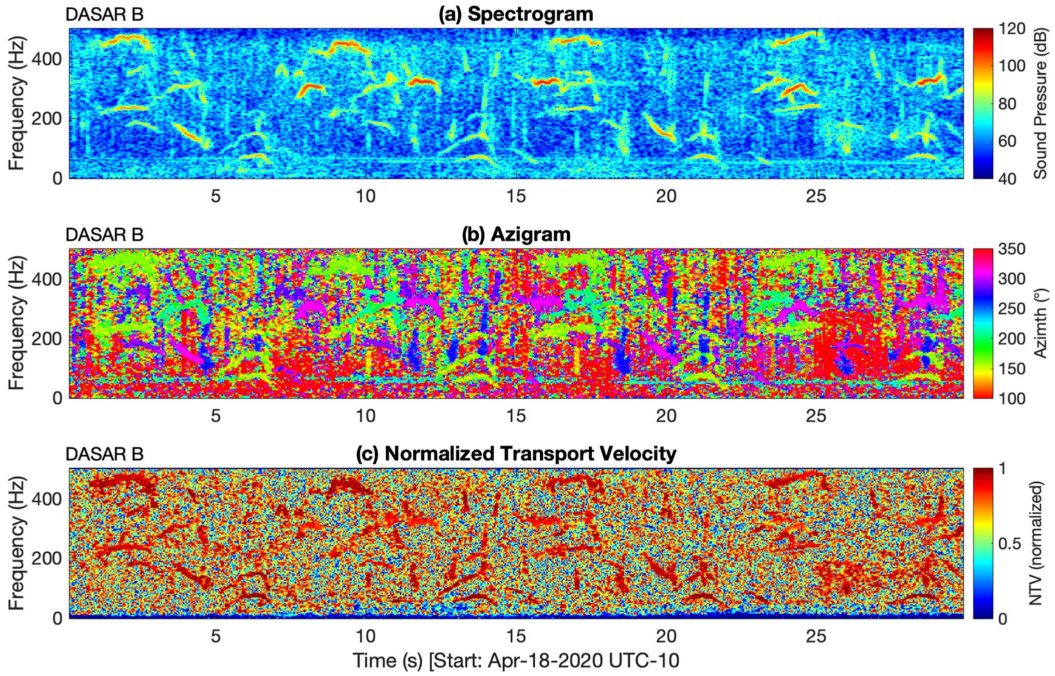

FIG. 2. (Color online) Spectrogram: (a) azigram, (b) normalized transport velocity, (c) over a 30 s time window starting at 2:24 UTC-10 on DASAR B. The color scale for (a) is in terms of power spectral density (dB re $\mu$Pa$^2$/Hz), while the color scale for (b) is in terms of geographic azimuth relative to geographic north. All three subplots are computed using window and FFT lengths of 256 samples with 90% overlap and no time-averaging of the FFTs.

that the histogram time window $\Delta T_h$ should be long enough for the azigram to include whale calls from all currently singing whales, and short enough for any shifts in an animal's azimuth to be negligible. For identifying humpback whale songs in this dataset, a time window $\Delta T_h = 60\,\mathrm{s}$ was sufficient. To minimize contributions to $h_\theta$ from diffuse background noise and other non-directional sources, a NTV threshold can be applied so that any observation $\varphi(T,f)$ associated with a NTV below a value $\gamma_U$ is discarded from the histogram. For example, the histogram computed from Fig. 2(b) would only be computed with azimuthal values whose NTV is above $\gamma_U$ in Fig. 2(c). The resulting filtered histogram $H_\theta(\Delta T_h)$ emphasizes azimuths that are associated with highly directional compact sources, such as whales and boats. $H_\theta$ is normalized by its maximum value so that the bin associated with the most likely azimuth is always scaled to 1. Figure 3 displays sample histograms, before and after applying the NTV threshold and normalization, illustrating how this filtering enhances the azimuths associated with four distinct whales.

$H_\theta$ can be plotted as a function of both time and azimuth by stacking histograms into an image which will be referred to as "azimuthal histogram display" (AHD). This representation allows identification of continuous tracks from discrete sources, such as whales and boats, over longer time periods. These "azimuthal tracks" are not to be confused with the final 2D localized tracks which correspond to a sequence of positions (latitude and longitude) for a given whale. Figure 4 displays the AHDs for all three DASARs over a 24 h time window starting at midnight on April 18, 2020. Also, shown in Fig. 4 is the unfiltered AHD from DASAR C, illustrating how the NTV filtering and normalization improves the quality of the azimuthal tracks. Whales and small boats are easily distinguishable by the slope (azimuthal rate) of their azimuthal tracks. Because they are generally louder and thus typically more distant from the sensors, the azimuthal rate of whales is low, on the order of several degrees per hour. By contrast, small boats generally travel faster and need to be closer to a DASAR to be detected, thus, displaying higher azimuthal rates and producing tracks that are much shorter (on the order of minutes) and steeper (nearly vertical lines in Fig. 4). The AHDs in Fig. 4 also indicate that whale song only arrives from between approximately 100° and 350° (i.e., away from shore), while boat tracks can arrive from any direction, including from the direction of the coastline.

To localize multiple individual whales over time, azimuthal tracks from the same whale need to be linked on at least two DASARs. The deployment configuration is such that if a whale is only detected on two of the three DASARs, as it is sometimes the case, DASAR B (the middle instrument) will always be involved. The practical problem to solve here is thus, to determine which, if any, of the azimuthal tracks from DASAR B are associated with tracks from DASARs A and/or C.

### C. Manual tracing of azimuthal tracks

Tracing azimuthal tracks is a problem that should be possible to automate, but for the purposes of the current study, tracks are traced manually from the AHDs of all three DASARs between midnight and 3:00 UTC-10 on April 18, 2020. Let the $n$th azimuthal track on DASAR $\alpha$ be denoted $\Theta_{n\alpha}(t)$. Figure 5 shows both the AHDs and resulting traced tracks. These sample tracks have been labeled such that tracks on different DASARs that share the same number are associated with the same whale (Sec. IV D explains how traces from the same whale are identified between DASARs). The tracks show six distinct whales; five are

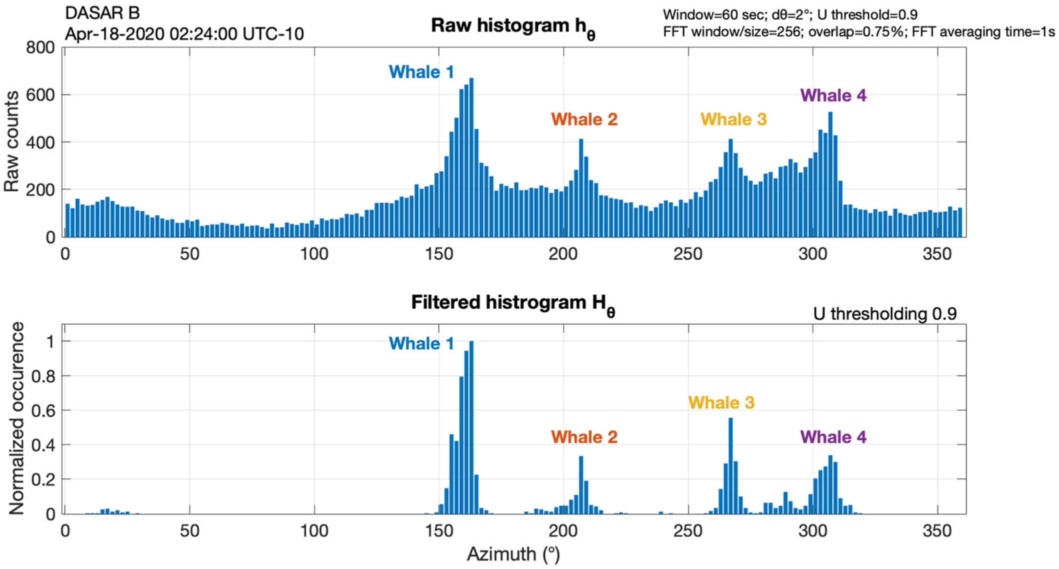

FIG. 3. (Color online) Raw histogram $h_\theta$ and filtered histogram $H_\theta$ at 2:24 UTC-10. These histograms estimate the distribution of azimuths over a time window $\Delta T_h = 60\,\mathrm{s}$ using bin width $d\theta = 2°$, and illustrate how NTV thresholding ($\gamma_U = 0.9$) enhances the azimuthal peaks associated with (at least) four distinct humpback whales. The azigrams used to compute these histograms have window and FFT lengths of 256 samples with 75% overlap, and 1 s time-averaging of the FFTs.

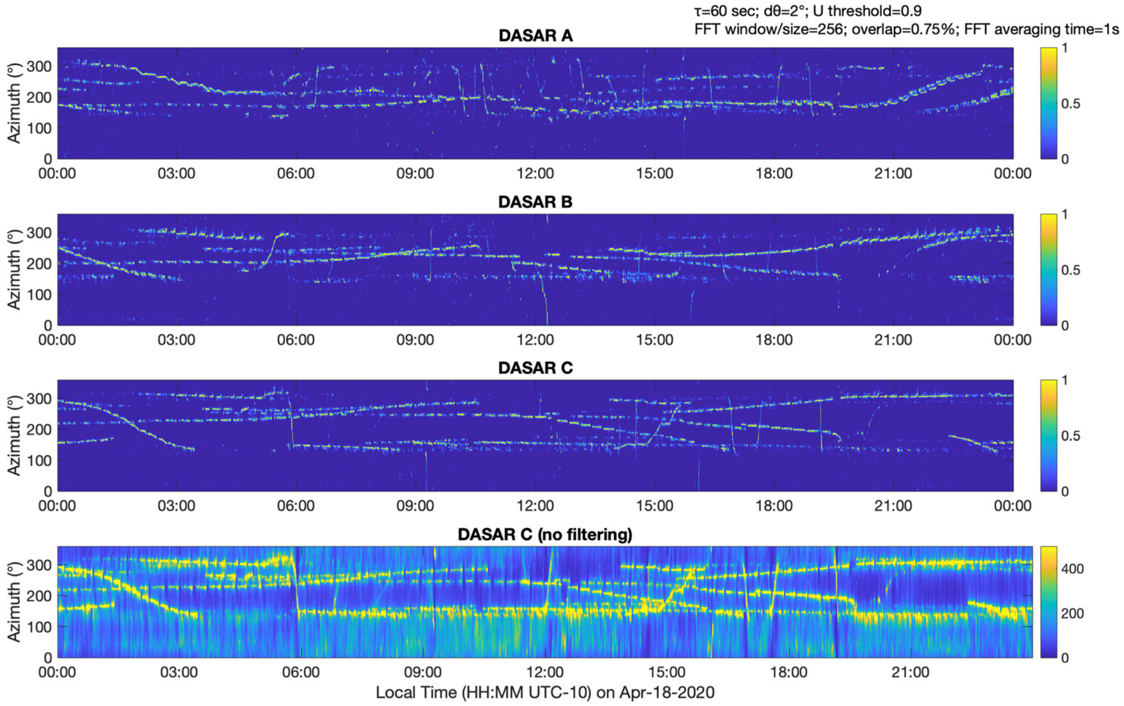

FIG. 4. (Color online) Azimuthal histogram displays for DASARs A, B, and C on April 18, 2020. These plots reveal azimuthal tracks of both whales (longer and smoother) and boats (quasi-vertical lines). The bottom plot shows the unfiltered AHD for DASAR C, illustrating how the filtering improves the visibility of the azimuthal tracks. The parameters used here are the same as those used to compute the histograms in Fig. 3.

detected on all three DASARs, while one is detected on DASARs B and C only (whale 6). The time resolution of the azimuthal tracks is 1 min.

### D. Azigram thresholding

Image thresholding can be applied to any azigram to create a binary image based on an azimuthal sector of width $d\varphi$, thus, allowing whale calls arriving from a specific direction to be isolated on different DASARs. This process, defined here as "azigram thresholding", works as long as any two singing individuals are separated by at least an angle $d\varphi/2$. To enhance the thresholded image, a 2D 3 × 3-pixel median filter (Huang *et al.*, 1979) is applied to remove speckle components of the image.

For any given time and DASAR, the azimuth associated with track $\Theta_{n\alpha}(t)$ can be used to threshold the corresponding azigram on sensor $\alpha$ and isolate the song units of whale $n$. Figure 6 demonstrates examples of four contemporary whale songs extracted from DASARs B and C. The 30 s time window presented here corresponds to the first half of the histograms shown in Fig. 3, and thus, shows the same four whales.

### E. Matching azimuthal tracks between DASARs

Let $B_\alpha(T,f)$ and $B_\beta(T,f)$ be two binary images covering a same time window of length $\Delta T_r$, obtained from applying azigram thresholding to DASARs $\alpha$ and $\beta$, respectively. The azimuthal sector used to produce the images may differ between sensors (e.g., Fig. 6). The similarity of the two images can be quantified by taking the maximum value of the cross correlation between $B_\alpha$ and $B_\beta$ along time, expressed as

$$R = \max_{\tau}\left\{\sum_{f}\sum_{T} B_\alpha(T,f)B_\beta(T+\tau,f)\right\}, \tag{5}$$

where $\tau$ is the cross correlation time delay. $R$ can be normalized into a "cross correlation score" as

$$\overline{R} = \frac{100}{\max\{[P_\alpha,\, P_\beta]\}}R, \tag{6}$$

where $P_\alpha$ and $P_\beta$ are the total number of positive pixels shared by $B_\alpha$ and $B_\beta$, respectively. Equation (6) normalizes the cross correlation score between any two images to lie between 0 and 100. Cross correlating binary images is conceptually similar to "spectrogram correlation" methods used to detect stereotyped baleen whale calls (Mellinger and Clark, 2000).

For any time window that reports azimuthal tracks on two DASARs, the likelihood of these tracks being related can be assessed by computing their cross correlation. The time window used for the cross correlation $\Delta T_r$ should be long enough to include songs from all the whales being tracked, and short enough for their azimuths to remain constant to within the azimuthal sector $d\varphi$. For comparing humpback whale calls in this dataset, a time window of $\Delta T_r = \Delta T_h = 60$ s and an azimuthal sector of $d\varphi = 15°$ (which corresponds to the azimuthal uncertainty derived in the Appendix) was sufficient. By computing the median score of each cross-track combination across the portion of the two tracks that overlap temporally, the likelihood of any two azimuthal tracks being from the same whale can be identified. As an example, Fig. 7 shows the cross correlation scores between $\Theta_{1C}$ and all

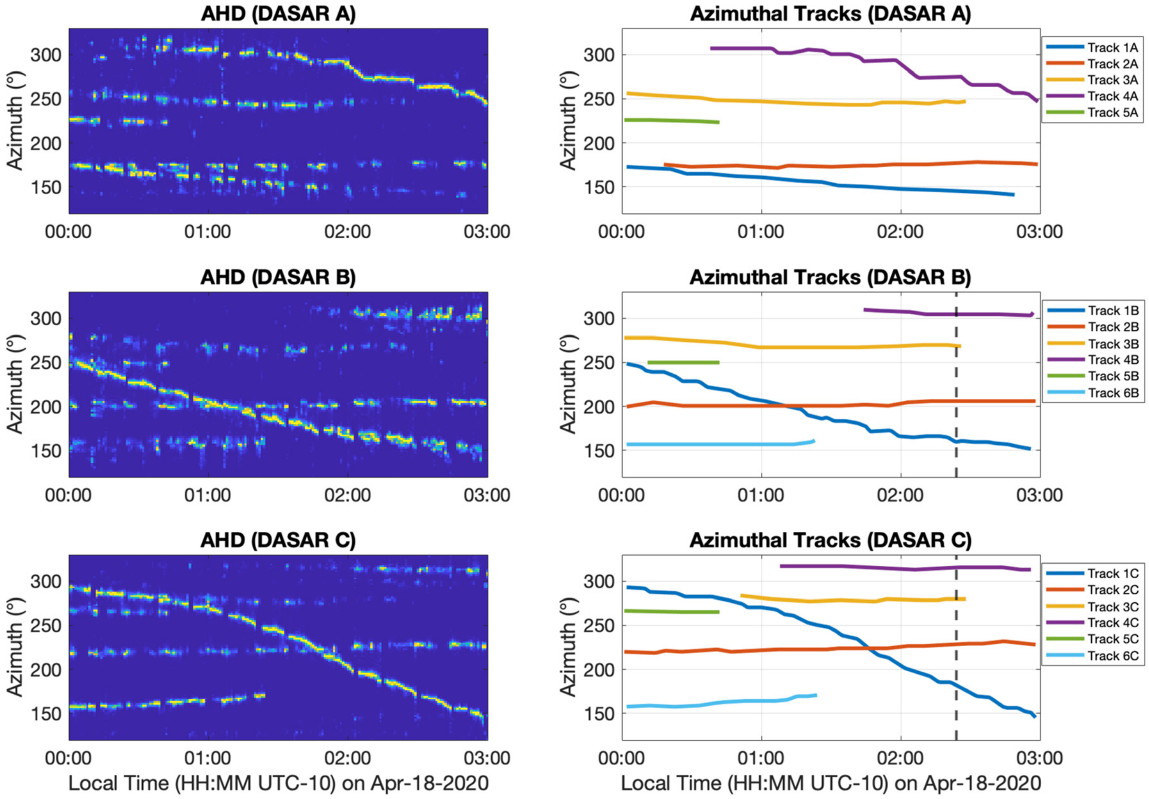

FIG. 5. (Color online) AHDs and manually selected azimuthal tracks/traces from DASARs A, B, and C between midnight and 3:00. The azimuthal traces shown here have been labeled such that tracks sharing the same number are associated with the same whale. The time resolution of these tracks is 1 min.

six tracks from DASAR B. The median scores clearly indicate that $\Theta_{1B}$ is the most likely track associated with the reference track $\Theta_{1C}$. The regular troughs in the cross correlation score occur when the whale stops singing while it surfaces to breathe, approximately every 15 min.

Following this procedure, the median scores for all combinations of tracks between DASAR B and DASARs A and C can be used to create confusion matrices, as illustrated in Fig. 8. The correct associations (along the diagonal of the confusion matrices) consistently produce the highest median scores. Note that when comparing track $\Theta_{6B}$ to tracks from DASAR A (bottom row of left panel in Fig. 8), all median scores are low, which is, expected since whale 6 is not detected on DASAR A (i.e., $\Theta_{6B}$ has no match on DASAR A). Based on this analysis, a score above 15 is associated with a correct match between tracks.

## V. LOCALIZATION AND TRACKING RESULTS

The whales whose azimuthal tracks were extracted and matched in Sec. IV are used to demonstrate the 2D tracking results. The 2D localization method used in this study relies

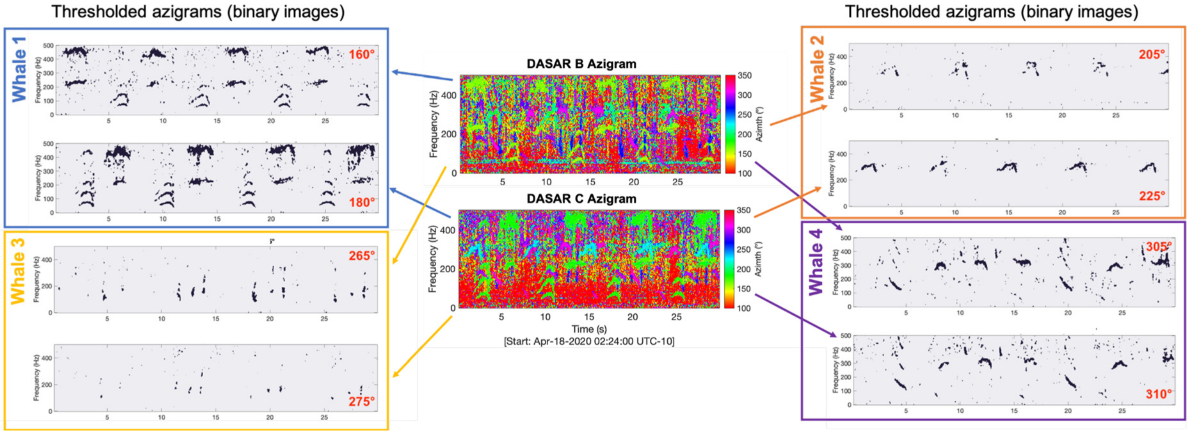

FIG. 6. (Color online) Azigrams and associated binary images (thresholded azigrams) obtained by applying azigram thresholding to DASARs B and C to isolate calls from four whales. The azigrams were computed using the same parameters as used for Fig. 2. The center of the azimuthal sectors of width $d\varphi = 15°$ is displayed in red on each binary image. The 30 s time window presented here corresponds to the first half of the histograms shown in Fig. 3, and thus, shows the same four whales.

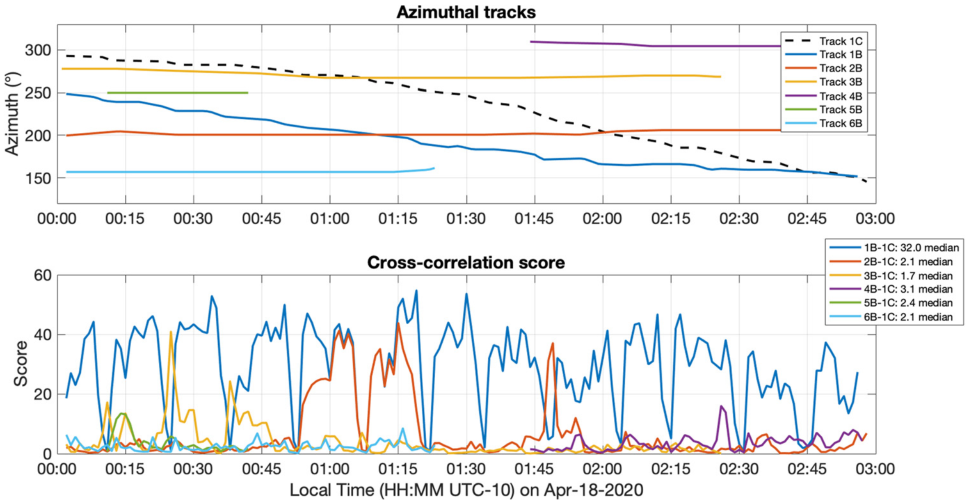

FIG. 7. (Color online) (Top) Reference azimuthal track $\Theta_{1C}$ (dashed black line) compared to all six tracks from DASAR B (solid lines). These are the same manually traced tracks shown in Fig. 5. (bottom) Cross correlation scores [Eq. (6)]for tracks on DASAR B, when compared with $\Theta_{1C}$. The median scores of each comparison, which are displayed in the bottom legend, suggest that $\Theta_{1B}$ is the best match with reference track $\Theta_{1C}$, as it has the highest median score. The cross correlation time window and azimuthal sector width used here are $\Delta T_r = 60$ s and $d\varphi = 15°$, respectively.

on triangulation from at least two DASAR azimuthal measurements and allows estimating a localization precision if at least three measurements are available. This approach is based on (Lenth, 1981) and has already been used to track bowhead whales in several studies involving DASARs (Greene *et al.*, 2004; Blackwell *et al.*, 2007; Thode *et al.*, 2012). Here, a 2D "localized track" refers to a sequence of latitudes and longitudes derived from the linked azimuthal tracks of a same whale between distinct vector sensors. Individual localizations are therefore not linked to specific song units, but rather have the same 1 min temporal resolution as the azimuthal tracks extracted from the AHDs.

Up to three DASARs are available in this dataset. Thus, at any given time, the number of available azimuth measurements $N$ (from the azimuthal tracks) for a whale lies between 0 and 3, $N \in [0, 3]$. A 2D whale track starts whenever $N \geq 2$ and ends if $N < 2$. Over time periods where $N = 3$, the localization uncertainty is also computed for each localization. Figure 9 shows the 2D localized tracks of all six whales with a localization interval of 1 min. Using the method from Greene *et al.* (2004), the 90% confidence ellipse is computed and displayed every 10 min over time periods whenever $N = 3$. The location uncertainty can be computed for portions of all tracks with the exception of whale 6, which is only detected on DASARs B and C.

The track from whale 1 shows particularly good resolution and shows the animal traveling towards the southeast. The direction of travel can be inferred from the azimuthal

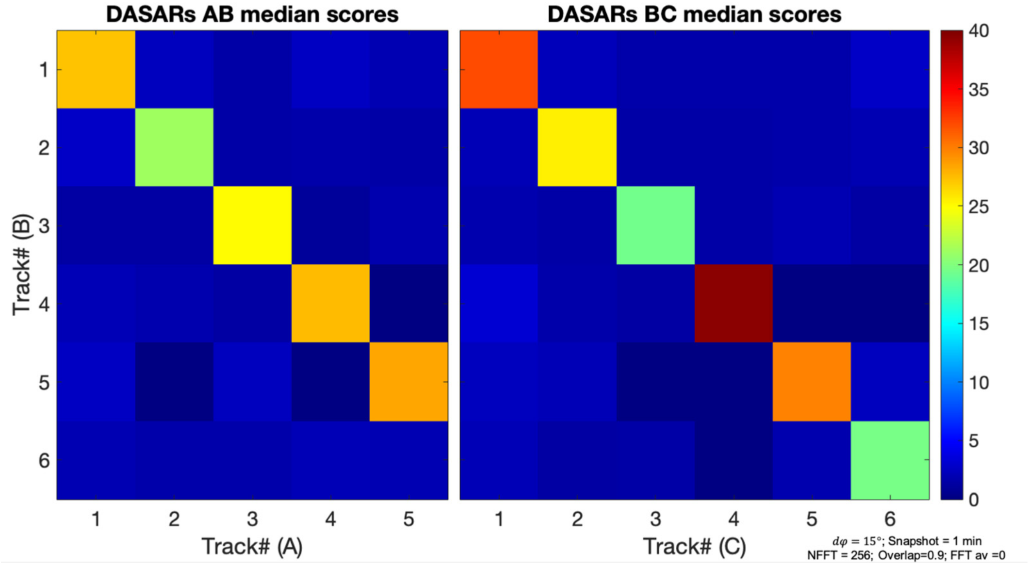

FIG. 8. (Color online) Confusion matrices for all combinations of azimuthal tracks between DASAR B and DASARs A and C. Each grid point in these matrices represents a track combination and is computed as the median value of their normalized cross correlation scores [Eq. (6)]. The correct associations of azimuthal tracks along the diagonal consistently show the highest scores. The parameters used here are the same as those from Fig. 7.

tracks in Fig. 6. During the first part of the track, the whale is directly in front of DASAR B (about 2.5 km off the coast) and is detected on all three instruments. As the whale travels southeast, the range uncertainties grow larger, since the triangulated bearings cross at shallower angles. Eventually, at 2:28, when the whale is about 10 km away from DASAR A, it is only detected on DASARs B and C, and no additional localization uncertainties can be estimated. Using the first hour of the localized 2D track from whale 1 (when the animal is directly in front of DASAR B), the swimming speed can be derived and displayed in Fig. 10. The reported average swim speed of approximately 2 km/h is consistent with previous studies of singing whales (Frankel *et al.*, 1995; Noad and Cato, 2007). The azigram cross correlation score [Eq. (6)] between the tracks from whale 1 on DASARs B and C is also displayed in Fig. 10 to show when the whale is singing (high score) and when it is surfacing to breathe (low score). Comparing the swim speed to the dive cycle reveals that the two appear to be correlated, with the whale displaying the largest swimming speed immediately after beginning a new dive.

## VI. DISCUSSION

### A. Localization performance and limitations

Because of the intrinsic uncertainty in the dominant azimuth estimates, the performance of the 2D localization technique is highly dependent on the relative location of the source to the vector sensors and the deployment geometry itself. For this specific configuration of the DASARs arranged linearly along the coast, whale positions that are a few kilometers to the southwest of DASAR B produce the best results because (1) whales are detected on all three DASARs, and (2) the azimuthal beams from all three DASARs intersect at steep angles. This is the case for approximately the first hour of the 2D localized track from whale 1 in Fig. 9. As a whale moves farther offshore, the SNR of its songs decreases along with the precision of the azimuth estimates and thus, the localization precision. Similarly, whenever whales lie along the line connecting sensors of this nearly linear array, the location uncertainties also become larger since the triangulated azimuths intersect at shallower angles. Furthermore, whales that are present near the ends of the linear array are often detected on only the two closer DASARs. While the coastline deployment has practical advantages in terms of logistics (e.g., shallow water depth and proximity to a harbor), the quality of the 2D localizations could be greatly improved by placing one or multiple DASARs farther away from the coast in deep water, or on the coast of Lanai, to define a triangular array perimeter.

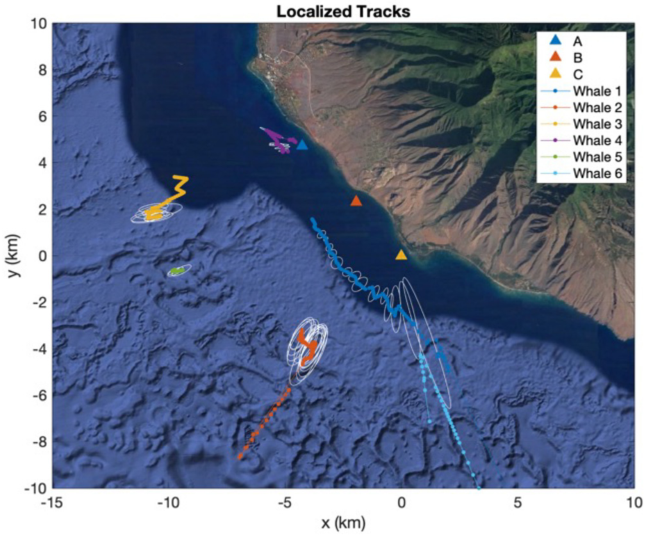

FIG. 9. (Color online) 2D localized tracks of the six whales whose azimuthal tracks were extracted and matched in Sec. IV. The whale 2D location estimates are computed every minute whenever at least two azimuthal measurements are available ($N \geq 2$). A 90% confidence ellipse is plotted every 10 min over time periods where three DASARs are available for the localization ($N = 3$).

### B. Azimuthal track matching

The method used to compare and match azimuthal tracks across DASARs relies on cross correlation of binary images obtained from azigram thresholding. While the example presented here is idealized (each azimuthal track from DASAR B has a unique matching track on at least one of the other two DASARs), identifying each matching track is straightforward (Fig. 8). The main potential weakness of this approach is its inability to differentiate between two whales that have the same azimuth on a given DASAR. Taking Fig. 7 as an example, the local cross correlation around 1:10 UTC-10 shows high scores for both tracks $\Theta_{1B}$ and $\Theta_{2B}$, which intersect at this time. Alternatively, two matching azimuthal tracks can produce low cross correlation scores at times when the whale stops singing while it surfaces. These troughs, which can be seen approximately every 15 min in both the cross correlation scores (to track 1B in Fig. 7) and the AHDs (Fig. 4), occasionally cause cross correlation scores with the wrong azimuthal track to become relatively high (as seen from the two cross correlation spikes for track $\Theta_{3B}$, around 00:24 and 00:40 UTC-10). Despite these potential limiting factors, when considering the entire azimuthal tracks and their median scores, the track from the correct whale (track $\Theta_{1B}$) clearly produces the highest median score. Therefore, this technique can successfully link two tracks from a same whale as long as (1) it is singing throughout the majority of the azimuthal track, and (2) the majority of its tracks are azimuthally unique on each DASAR.

An alternative method for matching azimuthal tracks from the same whale on multiple DASARs would be to attempt localizing every combination of tracks, and making a decision based on the uncertainty of the resulting 2D localized tracks, an approach dubbed "localize-before-detect". Indeed, the localization for the correct set of azimuthal tracks should produce physical 2D locations with lower uncertainties than potential localizations resulting from the

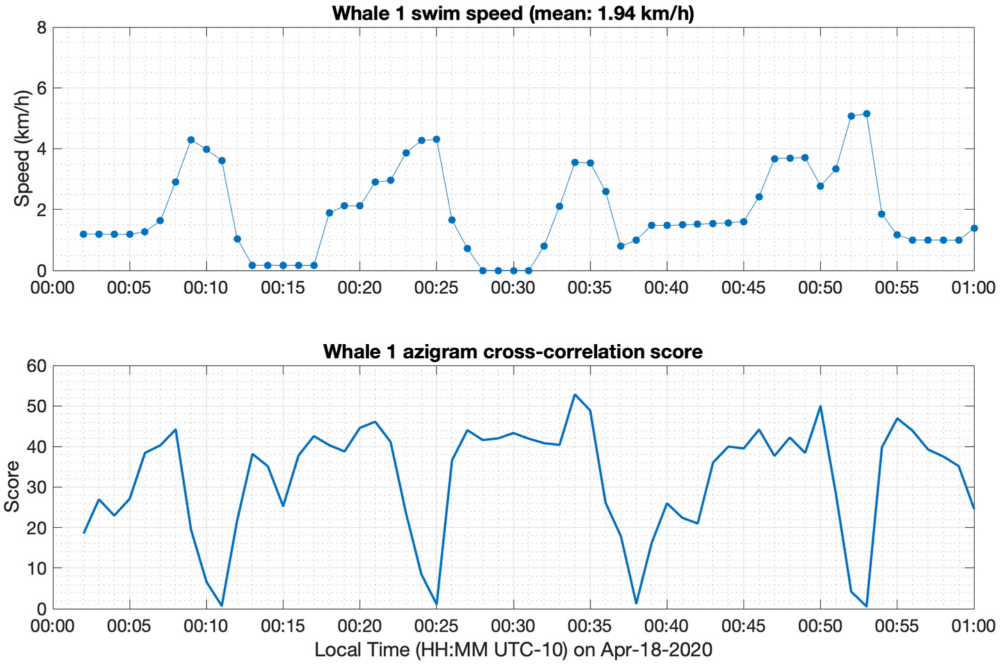

FIG. 10. (Color online) Swim speed of whale 1 derived from the first hour of its 2D localized track, when the whale is in front of DASAR B (top) and azigram cross correlation score between the tracks from whale 1 on DASARs B and C (bottom). Here, the cross correlation score indicates when the whale is singing (high score) and when it is surfacing to breathe (low score). Comparing the two plots shows that the swimming speed of the whale evolves in a consistent fashion between surfacings.

wrong combination of azimuthal tracks. However, this approach requires azimuthal tracks to be present on all three DASARs in order to compute the location uncertainties. While this approach may be faster and effective in accurately matching azimuthal tracks in conditions where whales have clearly distinguishable trajectories, the azigram cross correlation technique presented here is more robust as it exploits additional features in the time-frequency domain to link azimuthal tracks.

## C. Automation and real-time implementation

To process the entire dataset in a way that is both efficient and systematic, further automation of two key steps of the tracking algorithm are required: (1) extracting azimuthal traces from the AHDs, and (2) making association decisions between azimuthal tracks from different DASARs based on the median score confusion matrices (Fig. 8).

In this study, the matching problem was simplified by having a unique track for each whale on each DASAR. In practice, depending on how the azimuthal tracks are obtained, the track from a single whale may be split into multiple segments. In this case, rather than tracks being matched one-to-one, split tracks from one DASAR may need to be attributed to a single track from another DASAR. Alternatively, an azimuthal track from one DASAR may have no match on the other two (orphan track). Automating the whale tracking method would thus require implementing a decision-making algorithm (e.g., using thresholding) to determine whether azimuthal tracks are related. The azimuthal track tracing is thus closely related to the subsequent track matching process.

Several “multiple target tracking” (MTT) techniques exist that should allow automatic extraction of individual azimuthal traces from AHDs, including probability hypothesis density (PHD) filtering (Gruden and White, 2020) and graph-based approaches (Vo *et al.*, 2010; Meyer *et al.*, 2018). These extraction techniques might also be implemented in real-time.

## D. Processing parameters and potential applications

The parameters used here were chosen specifically for successfully identifying and matching humpback whale songs. The histogram and azigram cross correlation time windows, $\Delta T_h$ and $\Delta T_r$, respectively, have similar requirements in that they both should be long enough to include song units from all currently singing whales. However, they need to be short enough to account for both azimuthal changes of the source and the measurement uncertainty, with respect to the histogram bin width $d\theta$ and azimuthal sector $d\varphi$. The time window of 60 s used here for both the AHDs and azigram cross correlation is suited for the rate at which the azimuth of whales’ changes with time. In principle, this same approach could be used to track boats. To do so, the duration of the time windows would have to be reduced to account for the rapid changes in azimuth.

Note that the FFT averaging (1 s) was only used when computing the histograms. This allows reducing the variance of the azimuth estimates, which in turn, makes the azimuthal tracks in the AHDs sharper. Because FFT averaging affects the shape of individual song units in the azigram image, the cross correlation produces better results when no FFT averaging is applied.

The high NTV threshold used here to filter the AHDs (0.9) is designed to enhance the contributions from compact sources, such as whales and boats. Alternatively, the diffuse ambient noise field could be examined by removing samples of $\varphi(T, \mathrm{f})$ that have a high NTV and generating AHDs that only show contributions from low NTV samples. This approach would allow studying the directional characteristics of the diffuse ambient noise field over time. For example, the bottom subplot in Fig. 4 indicates that much of the diffuse ambient noise arises from the shoreline.

In the subset of data presented in this study, the number of singing whales is such that individual whales can be distinguished. Earlier in the breeding season, however, the number of whales is much higher, and distinguishing individual tracks in the AHDs becomes difficult. While this makes tracking individual whales less feasible, the overwhelming amount of song becomes an opportunity to treat humpback whale song as a diffuse noise source (Seger *et al.*, 2016). Instead of tracking individual whales, the DASARs could be used to localize the "center of mass" of a singing region. Furthermore, using longer time windows to compute the NTV, the azimuthal distribution of whales could be assessed to determine if whales are clustered or widely distributed.

## VII. CONCLUSION

This study presents a method for multi-target 2D tracking of continuous acoustic sources using vector sensors. The technique, which relies on vector sensors' ability to measure directional quantities of the acoustic field, is demonstrated on simultaneously singing humpback whales off western Maui using three sensors. The extraction of azimuthal tracks from individual whales on each sensor is possible using a histogram representation of the azigrams (AHDs), to which a transport velocity threshold is applied to enhance their quality. Subsequently, the azimuthal tracks are compared across vector sensors by cross correlating thresholded azigrams, allowing tracks from the same whale to be linked between sensors. Once the azimuthal tracks are correctly matched, individual whales can be localized and tracked in 2D (latitude and longitude) using triangulation. A position uncertainty can be estimated whenever three azimuthal measurements are available. The method was demonstrated by tracking the position of six singing whales over 3 h. The derived swimming speed of one whale track showed that the swimming speed of the whale is related to its dive cycle.

## ACKNOWLEDGMENTS

The authors would like to acknowledge Russell Hansen from the Pride of Maui, as well as Lee James and Peter Colombo from Ultimate Whale Watch, who provided the vessels and assisted in the deployment and recovery of the instruments off Maui. Valuable suggestions and reviews to the manuscript were also provided by Jay Barlow, William Hodgkiss, Peter Shearer, William Coles, Florian Meyer, and Pina Gruden.

## APPENDIX

Figure 11 shows the results of the optimized boat calibration, along with the residual difference between the acoustically derived and GPS-derived azimuths derived

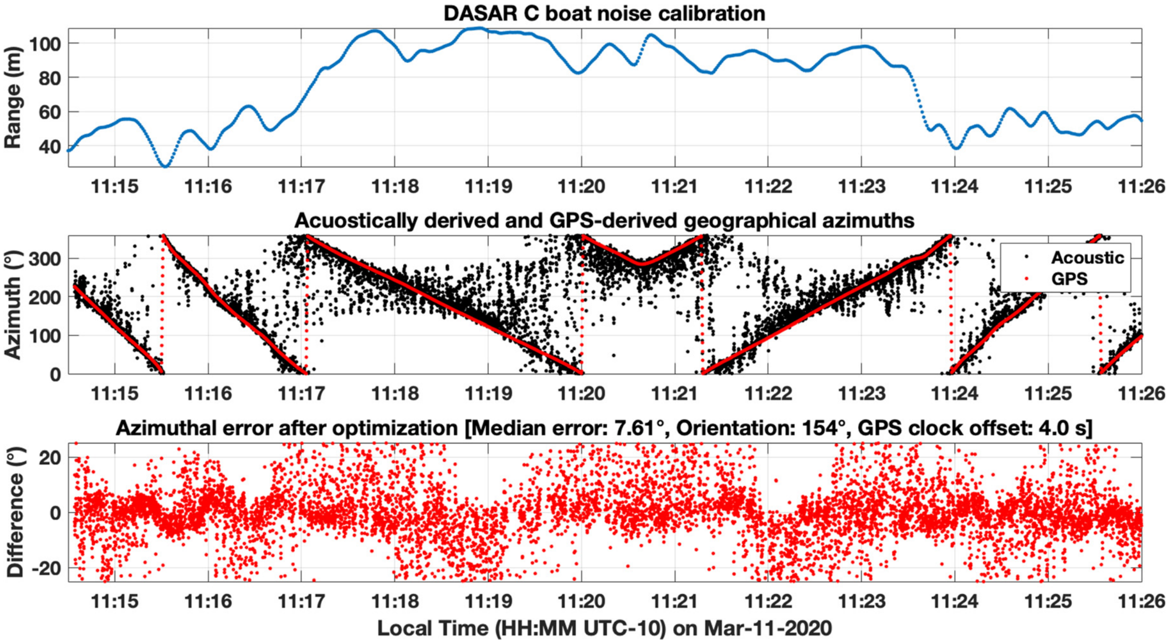

FIG. 11. (Color online) Boat noise calibration for DASAR C. GPS-derived range of boat (top), acoustically derived and GPS-derived geographic azimuths (middle), and azimuthal difference, after optimization (bottom). The acoustically derived azimuths were estimated between 350 and 400 Hz. The optimization results show that the median error of the DASARs is 7.61°. For DASAR C, the x-velocity axis is 154° clockwise relative to true north and the recorder timing offset from GPS is 4 s.

from boat noise. Only the narrow bandwidth between 350 and 400 Hz was available for boat noise calibrations due to the strong interference of overlapping humpback song. The median absolute difference between the GPS and acoustic estimates is 7.61°, and so the uncertainty of the acoustic bearings was set to ±7.5° (azimuthal sector width, $d\varphi = 15°$) in the azigram thresholding (Sec. IV D).

Au, W., Pack, A., Lammers, M., Herman, L., Deakos, M., and Andrews, K. (**2006**). "Acoustic properties of humpback whale songs," J. Acoust. Soc. Am. **120**, 1103–1110.

Au, W. W. L., Mobley, J., Burgess, W. C., Lammers, M. O., and Nachtigall, P. E. (**2000**). "Seasonal and diurnal trends of chorusing humpback whales wintering in waters off western Maui," Mar. Mamm. Sci. **16**, 530–544.

Blackwell, S. B., Richardson, W. J., Greene, C. R., and Streever, B. (**2007**). "Bowhead whale (Balaena mysticetus) migration and calling behaviour in the Alaskan Beaufort sea, Autumn 2001-04: An acoustic localization study," Arct. **60**, 255–270.

Cholewiak, D. M., Cerchio, S., Jacobsen, J. K., Urban-R, J., and Clark, C. W. (**2018**). "Songbird dynamics under the sea: Acoustic interactions between humpback whales suggest song mediates male interactions," Royal Soc. Open Sci. **5**, 171298.

D'Spain, G. L., Hodgkiss, W. S., and Edmonds, G. L. (**1991**). "Energetics of the deep ocean's infrasonic sound field," J. Acoust. Soc. Am. **89**, 1134–1158.

D'Spain, G. L., Luby, J. C., Wilson, G. R., and Gramann, R. A. (**2006**). "Vector sensors and vector sensor line arrays: Comments on optimal array gain and detection," J. Acoust. Soc. Am. **120**, 171–185.

Darling, J. D., Jones, M. E., and Nicklin, C. P. (**2006**). "Humpback whale songs: Do they organize males during the breeding season?," Behaviour **143**, 1051–1101.

Frankel, A. S., Clark, C. W., Herman, L. M., and Gabriele, C. M. (**1995**). "Spatial distribution, habitat utilization, and social interactions of humpback whales, Megaptera novaeangliae, off Hawaii, determined using acoustic and visual techniques," Can. J. Zool. (Revue Canadienne De Zoologie) **73**, 1134–1146.

Goldbogen, J. A., Stimpert, A. K., DeRuiter, S. L., Calambokidis, J., Friedlaender, A. S., Schorr, G. S., Moretti, D. J., Tyack, P. L., and Southall, B. L. (**2014**). "Using accelerometers to determine the calling behavior of tagged baleen whales," J. Exp. Biol. **217**, 2449–2455.

Google (Last viewed 04/19/2021).

Greene, C. R., McLennan, M. W., Norman, R. G., McDonald, T. L., Jakubczak, R. S., and Richardson, W. J. (**2004**). "Directional frequency and recording (DIFAR) sensors in seafloor recorders to locate calling bowhead whales during their fall migration," J. Acoust. Soc. Am. **116**, 799–813.

Gruden, P., and White, P. R. (**2020**). "Automated extraction of dolphin whistles-A sequential Monte Carlo probability hypothesis density approach," J. Acoust. Soc. Am. **148**, 3014–3026.

Guazzo, R. A., Helble, T. A., D'Spain, G. L., Weller, D. W., Wiggins, S. M., and Hildebrand, J. A. (**2017**). "Migratory behavior of eastern North Pacific gray whales tracked using a hydrophone array," PloS One **12**, e0185585.

Helble, T. A., Henderson, E. E., Ierley, G. R., and Martin, S. W. (**2016**). "Swim track kinematics and calling behavior attributed to Bryde's whales on the Navy's Pacific Missile Range Facility," J. Acoust. Soc. Am. **140**, 4170–4177.

Helble, T. A., Ierley, G. R., D'Spain, G. L., and Martin, S. W. (**2015**). "Automated acoustic localization and call association for vocalizing humpback whales on the Navy's Pacific Missile Range Facility," J. Acoust. Soc. Am. **137**, 11–21.

Henderson, E. E., Helble, T. A., Ierley, G. R., and Martin, S. W. (**2018**). "Identifying behavioral states and habitat use of acoustically tracked humpback whales in Hawaii," Mar. Mamm. Sci. **34**, 1–17.

Herman, L. M. (**2017**). "The multiple functions of male song within the humpback whale (Megaptera novaeangliae) mating system: Review, evaluation, and synthesis," Biol. Rev. **92**, 1795–1818.

Holler, R. A. (**2014**). *The Evolution of the Sonobuoy from World War II to the Cold War*,“ (Navmar Applied Sciences Corp, Warminster PA).

Huang, T., Yang, G., and Tang, G. (**1979**). "A fast two-dimensional median filtering algorithm," IEEE Trans. Acoust. Speech Signal Process. **27**, 13–18.

Lenth, R. V. (**1981**). "On finding the source of a signal," Technometrics **23**, 149–154.

Mann, J. A., Tichy, J., and Romano, A. J. (**1987**). "Instantaneous and time-averaged energy-transfer in acoustic fields," J. Acoust. Soc. Am. **82**, 17–30.

Martin, B., Zeddies, D. G., Gaudet, B., and Richard, J. (**2016**). "Evaluation of three sensor types for particle motion measurement," Eff. Noise Aquatic Life II **875**, 679–686; avaiable at https://link.springer.com/chapter/10.1007/978-1-4939-2981-8_82.

McDonald, M. A. (**2004**). "DIFAR hydrophone usage in whale research," Can. Acoust. **32**, 155–160.

Mellinger, D. K., and Clark, C. W. (**2000**). "Recognizing transient low-frequency whale sounds by spectrogram correlation," J. Acoust. Soc. Am. **107**, 3518–3529.

Meyer, F., Kropfreiter, T., Williams, J. L., Lau, R. A., Hlawatsch, F., Braca, P., and Win, M. Z. (**2018**). "Message passing algorithms for scalable multitarget tracking," Proc. IEEE **106**, 221–259.

Noad, M. J., and Cato, D. H. (**2007**). "Swimming speeds of singing and non-singing humpback whales during migration," Mar. Mamm. Sci. **23**, 481–495.

Noad, M. J., Cato, D. H., and Stokes, M. D. (**2004**). "Acoustic tracking of humpback whales: Measuring interactions with the acoustic environment," in Proc. Acoust. 353–358.

Payne, R. S., and McVay, S. (**1971**). "Songs of humpback whales," Science **173**, 585–597.

Popper, A. N., Fay, R. R., and Au, W. (**2000**). *Hearing by Whales and Dolphins* (Springer Science & Business Media, New York) Vol. 12.

Raghukumar, K., Chang, G., Spada, F., and Jones, C. (**2020**). "A vector sensor-based acoustic characterization system for marine renewable energy," J. Mar. Sci. Eng. **8**, 187.

Schau, H. C., and Robinson, A. Z. (**1987**). "Passive source localization employing intersecting spherical surfaces from time-of-arrival differences," IEEE Trans. Acoust. Speech Signal Process. **35**, 1223–1225.

Schiffrer, G., and Stanzial, D. (**1994**). "Energetic properties of acoustic fields," J. Acoust. Soc. Am. **96**, 3645–3653.

Schmidt, V., Weber, T. C., Wiley, D. N., and Johnson, M. P. (**2010**). "Underwater tracking of humpback whales (Megaptera novaeangliae) with high-frequency pingers and acoustic recording tags," IEEE J. Oceanic Eng. **35**, 821–836.

Seger, K. D., Thode, A. M., Urban-R, J., Martinez-Loustalot, P., Jimenez-Lopez, M. E., and Lopez-Arzate, D. (**2016**). "Humpback whale-generated ambient noise levels provide insight into singers' spatial densities," J. Acoust. Soc. Am. **140**, 1581–1597.

Spiesberger, J. L. (**2001**). "Hyperbolic location errors due to insufficient numbers of receivers," J. Acoust. Soc. Am. **109**, 3076–3079.

Stanistreet, J. E., Risch, D., and Van Parijs, S. M. (**2013**). "Passive acoustic tracking of singing humpback whales (Megaptera novaeangliae) on a Northwest Atlantic feeding ground," PloS One **8**, e61263.

Stimpert, A. K., Lammers, M. O., Pack, A. A., and Au, W. W. L. (**2020**). "Variations in received levels on a sound and movement tag on a singing humpback whale: Implications for caller identification," J. Acoust. Soc. Am. **147**, 3684–3690.

Thode, A. M., Conrad, A. S., Ozanich, E., King, R., Freeman, S. E., Freeman, L. A., Zgliczynski, B., Gerstoft, P., and Kim, K. H. (**2021**). "Automated two-dimensional localization of underwater acoustic transient impulses using vector sensor image processing (vector sensor localization)," J. Acoust. Soc. Am. **149**, 770–787.

Thode, A. M., Kim, K. H., Blackwell, S. B., Greene, C. R., Nations, C. S., McDonald, T. L., and Macrander, A. M. (**2012**). "Automated detection and localization of bowhead whale sounds in the presence of seismic airgun surveys," J. Acoust. Soc. Am. **131**, 3726–3747.

Thode, A. M., Sakai, T., Michalec, J., Rankin, S., Soldevilla, M. S., Martin, B., and Kim, K. H. (**2019**). "Displaying bioacoustic directional information from sonobuoys using 'azigrams,'" J. Acoust. Soc. Am. **146**, 95–102.

Vo, B. N., Vo, B. T., Pham, N. T., and Suter, D. (**2010**). "Joint detection and estimation of multiple objects from image observations," IEEE Trans. Signal Process. **58**, 5129–5141.