JASA ARTICLE

Bayesian detection and tracking of odontocetes in 3-D from their echolocation clicks

Junsu Jang,[a] Florian Meyer, Eric R. Snyder, Sean M. Wiggins, Simone Baumann-Pickering, and John A. Hildebrand

Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093, USA

ABSTRACT:

Localization and tracking of marine animals can reveal key insights into their behaviors underwater that would otherwise remain unexplored. A promising nonintrusive approach to obtaining location information of marine animals is to process their bioacoustic signals, which are passively recorded using multiple hydrophones. In this paper, a data processing chain that automatically detects and tracks multiple odontocetes (toothed whales) in three dimensions (3-D) from their echolocation clicks recorded with volumetric hydrophone arrays is proposed. First, the time-difference-of-arrival (TDOA) measurements are extracted with a generalized cross-correlation that whitens the received acoustic signals based on the instrument noise statistics. Subsequently, odontocetes are tracked in the TDOA domain using a graph-based multi-target tracking (MTT) method to reject false TDOA measurements and close gaps of missed detections. The resulting TDOA estimates are then used by another graph-based MTT stage that estimates odontocete tracks in 3-D. The tracking capability of the proposed data processing chain is demonstrated on real acoustic data provided by two volumetric hydrophone arrays that recorded echolocation clicks from Cuvier's beaked whales (*Ziphius cavirostris*). Simulation results show that the presented MTT method using 3-D can outperform an existing approach that relies on manual annotation. © *2023 Acoustical Society of America*.

https://doi.org/10.1121/10.0017888

(Received 22 October 2022; revised 4 April 2023; accepted 5 April 2023; published online 2 May 2023)

[Editor: Haiqiang Niu]]

Pages: 2690–2705

I. INTRODUCTION

Passive acoustic monitoring (PAM) is a nonintrusive and efficient approach for studying and monitoring acoustically active animals, especially species that are challenging to observe visually. PAM enables detection, localization, and tracking of those animals and is, therefore, well-suited for studying their abundance,[1,2] behavior,[3] and response to anthropogenic activities.[4,5] With continuously increasing human activities in the ocean,[6–8] consistent monitoring and assessment of the population, behavior, phenology, and physiology of marine animals are necessary to make informed conservation plans and management policies.[9] Of particular monitoring interest are cetaceans (whales) because they are apex predators and environment sentinels.[10] Information on their density and geographic location can help us understand complex environmental changes, e.g., those caused by anthropogenic disturbances.

Cetaceans are known to produce various types of sounds for communication, navigation, and foraging.[11] They are divided into two suborders: *Odontoceti* (toothed whales or odontocetes) and *Mysticeti* (baleen whales or mysticetes). Odontocetes predominantly use high-frequency whistles or burst pulses to communicate,[12] whereas mysticetes produce low-frequency tonal calls, which when used in a pattern are considered songs.[13] To locate prey and relevant features of the environment, odontocetes also frequently emit echolocation clicks,[11] which are intense and directional short pulses underwater. As the echolocation clicks are impulsive and braodband in frequency,[14] they are promising signals for researchers to process to localize and track the echolocating odontocetes with PAM.

Various PAM technologies that are suitable for studying whales have been developed. Promising sensing approaches for PAM include towed arrays,[15] fiber optic cables,[16] mobile hydrophone recorders,[17,18] and bottom-mounted hydrophone arrays.[3,19] Emerging accessible and inexpensive PAM technologies are expected to provide acoustic datasets that are orders of magnitude larger than datasets provided by conventional technologies.[20] Thus, establishing effective algorithmic solutions for data processing, data management, data analysis, and performance evaluation is crucial.

This work focuses on the acoustic data recorded by volumetric hydrophone arrays that can provide location information of echolocating odontocetes in a three-dimensional Euclidean space. Here, human operators are typically required to manually inspect acoustic data, make decisions on the presence of whales, and select promising measurements.[3,21,22] Fully automated tracking of acoustically active whales from their recorded acoustic data involves numerous algorithmic challenges. In particular, typically, there are false positive detections due to noise from the environment and the instrument itself. Furthermore, there are missed detections as a result of signal aspect dependence with

[a]Electronic mail: jujang@ucsd.edu

 0001-4966/2023/153(5)/2690/16/$30.00

https://doi.org/10.1121/10.0017888

respect to the receivers and signal masking by background noise. Hence, it is necessary to solve a data association problem for the automated tracking of acoustically active whales across multiple data snapshots. Data association is complicated in scenarios where multiple acoustically active whales and other acoustic sources are present. In this work, a novel data processing chain that automatically detects and tracks odontocetes in three dimensions (3-D) from their echolocation clicks is proposed, and the tracking of Cuvier's beaked whales (*Ziphius cavirostris*) near the coast of California is demonstrated.

A. State-of-the-art

An established method to detect acoustic signals produced by whales is to first cross-correlate the time series of acoustic measurements provided by a pair of hydrophones and then apply a detection criterion to the peaks of the resulting cross-correlation signal. In addition, time-difference-of-arrival (TDOA) measurements can be extracted using the detected peaks from the cross-correlation signal. Although the conventional cross-correlation[23] is well suited for signals with a high signal-to-noise ratio (SNR), the generalized cross-correlation (GCC)[24] is typically used. The GCC performs frequency weighting to suppress noise, improving the detection performance.

If the TDOA measurements are extracted from four hydrophone pairs and the positions of the hydrophones are known, in principle, the three-dimensional location of a single whale can be computed by solving a nonlinear optimization problem.[11] Multiple hydrophones typically form a rigid volumetric hydrophone array to facilitate deployment and data processing, recording, and storage. For geometric reasons, just location information in bearing can be provided for whales in the far field of the array. Thus, two or more arrays are deployed for three-dimensional acoustic source localization.

In particular, in Refs. 3 and 25, a sequence of three-dimensional locations of echolocating beaked whales is estimated from the recordings of two high-frequency acoustic recording packages (HARPs), which host tetrahedron-shaped hydrophone arrays. For each snapshot of the acoustic signal, potential direction-of-arrivals (DOAs) relative to each HARP are computed from the TDOA measurements.[11] Each whale is localized in 3-D using the least squares method from manually selected DOAs that are likely generated from that whale. The sequence of each whale's three-dimensional location forms a track.

Multi-target tracking (MTT) methods are sequential Bayesian estimation techniques that automatically infer the number and states of multiple targets from sequences of measurements provided by one or multiple sensors, i.e., without the involvement of human operators. In the context of whale tracking, the states of interest can be the three-dimensional locations of the whales, the TDOAs of whales for a particular hydrophone pair,[15] or the frequency of narrow-band whale whistles in the spectrogram.[26] The MTT methods can succeed in tracking scenarios with false positive detections, missed detections, and data association uncertainty between measurements and targets. Traditional MTT methods, such as the joint probabilistic data association filter[27] and multiple hypothesis tracker (MHT),[28] model measurements and target states as random vectors. Newer methods, such as the probabilistic hypothesis density (PHD) filter[29] and multi-Bernoulli filter,[30,31] are derived in the formalism of random finite set.[30] Recently, a graph-based MTT method that is highly scalable in the number of targets, measurements, and sensors has been proposed.[32] This approach uses particle-based computations to perform operations that cannot be evaluated in closed form due to nonlinearities in the system model.[33] Graph-based MTT methods for localization and tracking from TDOA measurements in two dimensions have been introduced in Refs. 34 and 35.

Some MTT methods have been successfully employed for whale tracking. In Ref. 36, the MHT is applied to tracking beaked whales in 3-D. Here, potential three-dimensional locations of beaked whales are preprocessed from TDOA measurements, which are associated across hydrophone pairs based on click characteristics.[37] The preprocessed three-dimensional locations are used as measurements for a MHT that determines the number of beaked whale tracks, performs data association of three-dimensional locations with tracks, and runs a Kalman filter for each track. The suboptimum computing of three-dimensional locations is necessary because the MHT is limited to linear measurement models or mildly nonlinear measurement models. Further inherent challenges of the MHT are its computational complexity and memory requirements.[38] In addition, a Gaussian mixture probabilistic hypothesis density (GM-PHD) filter[39] for the tracking of whales in the TDOA domain is introduced in Ref. 15. Echolocation clicks and whistles of false killer whales are exploited to compute TDOA measurements for MTT. The data were acquired during line-transect surveys with towed hydrophone arrays. The introduced method extends the original GM-PHD filter by updating the existing and new whale tracks separately and incorporating amplitude information to support the initialization of new whale tracks and to better reject false positive detections. However, the GM-PHD filter also relies on linear measurement models or mildly nonlinear measurement models and is, therefore, limited to tracking in the TDOA domain.

Various methods for the localization and tracking of whales from their acoustic signals have been developed. A common approach for the three-dimensional localization of whales is a grid-search,[3,22,40–42] wherein an ambiguity surface is generated on a three-dimensional grid of potential whale locations by comparing the expected modeled TDOA measurements with the actual measurements. In Refs. 41 and 42, multiple whales are localized by determining local maxima of the ambiguity surface using the Simplex[43] or Metropolis-Hastings[44] algorithm. The method in Ref. 41 iteratively finds whale locations in 3-D by selecting the maximum peak that corresponds to the best match of the TDOA measurements, removing the TDOA measurements

corresponding to this maximum peak, and selecting the maximum peak that corresponds to the best match of the remaining TDOA measurements. In Ref. 42, Kalman filtering is used to estimate the whale tracks. In the case of simultaneously present whales, multiple Kalman filters run in parallel, and the Hungarian algorithm[45] is used for associating local maxima of the ambiguity surface to Kalman filters. Alternative approaches perform three-dimensional localization of whales by first estimating the DOAs of acoustic signals from hydrophone arrays.[17,25,46] The method in Ref. 17 subsequently tracks individual whales using nonsequential estimation based on Gibbs sampling.[47]

Common challenges of TDOA-based localizing and tracking are finding the correct combination of measurements across hydrophone pairs that correspond to the same acoustic source and initializing whale tracks accordingly. Typically, there are multiple TDOA measurements per hydrophone pair due to the presence of noise, echoes, and simultaneous vocalization of multiple whales. This challenge is often referred to as multisensor data association.[33,48] To address this, existing techniques either employ human operators to select and combine measurements manually,[3] rely on the local maxima corresponding to incorrectly matched TDOA measurements being significantly lower than those corresponding to true whale locations,[41] compute potential whale locations in a brute-force manner, i.e., based on all of the possible combinations of TDOA measurements,[42] or weigh grid points based on the number of TDOA measurements that are consistent with the corresponding potential whale location.[37] All of the existing methods for TDOA-based localizing and tracking of whales in 3-D either rely on human operators or heuristics to combine TDOA measurements and initialize the whale tracks.

B. Contributions and notation

The fundamental problem addressed in this paper is establishing an algorithmic solution for the tracking of echolocating odontocetes in 3-D. The goal is to develop a data processing method that fully automatically determines the number of odontocetes in the environment and estimates their tracks in 3-D from acoustic measurements. The proposed method will make it possible to (i) study deep-diving echolocating odontocetes more objectively and efficiently compared to approaches that rely on human operators and (ii) reveal key insights on behaviors of odontocetes underwater that otherwise would remain unexplored.

The proposed data processing chain extracts TDOA measurements of echolocation clicks from the raw acoustic signals using a GCC. An algorithm that uses a variant of the GCC, referred to as a generalized cross-correlation for whitening instrument noise (GCC-WIN), is introduced. This technique aims to suppress the instrument noise that interferes with the echolocation clicks. The peaks of the TDOAs above a certain amplitude threshold are used to estimate the parameters of interest, i.e., locations and velocities of the odontocete in time. Odontocetes are first tracked in the TDOA domain using a MTT method based on the framework of factor graphs and the sum-product algorithm (SPA).[49] A second MTT stage estimates odontocete tracks in 3-D by consistently combining ("fusing") estimated TDOAs of all of the hydrophone pairs provided by the first stage. The increased detection rate of the GCC-WIN algorithm results in a lower probability of missing an echolocation click and, in turn, improved tracking performance. The first tracking stage aims to reject false positive TDOA measurements and resolve longer gaps of missing TDOAs. This first MTT stage significantly improves the performance of estimating odontocete tracks in 3-D as performed by the second MTT stage.

Tracking whales in 3-D from TDOA measurements is further complicated because the underlying measurement model is nonlinear and the state space is high-dimensional. To address this challenge, a SPA that embeds particle flow[50] is used. Here, the particles are actively migrated toward high likelihood regions, making it possible to obtain good target detection and tracking performance in high dimensions.[51] Contrary to existing methods for detecting and tracking whales in 3-D, the proposed signal processing chain systematically reduces the instrument noise and uses a statistical model for multisensor data association and initializing whale tracks. This is expected to improve detection and tracking performance, especially in scenarios with low SNR and a significant number of false positive measurements.

This paper establishes a data processing chain that automatically detects and tracks odontocetes from acoustic measurements of their echolocation clicks. The key contributions of this paper are summarized as follows:

- A GCC that whitens instrument noise is developed. This GCC increases the detection probability of echolocation clicks and reduces the number of false positive detections.
- Two stages of graph-based MTT are established. The two MTT stages reject false positives, perform data association, determine the number of odontocetes, and estimate odontocete tracks in 3-D.
- The capabilities of the proposed data processing chain are demonstrated. For performance evaluation, recordings of echolocation clicks from two Cuvier's beaked whales are considered.

1. Notation

Random variables are displayed in sans serif and upright fonts; their realizations are displayed in serif, italic fonts. Vectors and matrices are denoted by bold lowercase and uppercase letters, respectively. For example, a random variable and its realization are denoted by x and x, respectively, and a random vector and its realization are denoted by $\mathbf{x}$ and $\boldsymbol{x}$, respectively. $\mathbf{x}_{0:k}$ is short for $[\mathbf{x}_0, \ldots, \mathbf{x}_k]^{\mathrm{T}}$. Furthermore, $||\boldsymbol{x}||$ and $\boldsymbol{x}^{\mathrm{T}}$ denote the Euclidean norm and the transpose of vector $\boldsymbol{x}$, respectively; $\propto$ indicates equality up to a normalization factor; $f(\boldsymbol{x})$ denotes the probability density function (PDF) of random vector $\mathbf{x}$, and $f(\boldsymbol{x}|\boldsymbol{y})$ denotes

https://doi.org/10.1121/10.0017888

the conditional PDF of random vector $\mathbf{x}$ conditioned on random vector $\mathbf{y}$. $f_{\mathsf{RV}(\boldsymbol{x})}\mathsf{V}(\boldsymbol{x})$ and $f(\boldsymbol{x}|\boldsymbol{y})$ are short notations for $f\mathbf{x}(\mathbf{x})$ and $f_{\mathsf{RV}(\boldsymbol{x})|\mathsf{V}(\boldsymbol{y})}\mathsf{V}(\boldsymbol{x})|\mathsf{V}(\boldsymbol{y})$ respectively. $|\mathcal{S}|$ denotes the cardinality of set $\mathcal{S}$; $\boldsymbol{I}_n$ denotes the $n \times n$ identity matrix. The operator "*" denotes the complex conjugate, and $j = \sqrt{-1}$ is the imaginary unit. Finally, the acronyms used throughout this paper are summarized in the nomenclature at the end of the paper.

II. GCC AND TDOA MEASUREMENTS

TDOA measurements are typically extracted from pairs of receivers for the localization of an uncooperative source. In three-dimensional space, each TDOA measurement gives rise to a hyperboloid (Fig. 1). With more than two receivers, multiple TDOA measurements can be extracted, and all of the hyperboloids, ideally, intersect in a single point at the source location.

For TDOA measurement extraction, the cross-correlation between the signals from a spatially separated hydrophone pair (s_1,s_2) is computed. The hydrophone pair forms a TDOA sensor $s \in \{1, \ldots, n_s\}$, where n_s is the number of sensors, i.e., number of pairs of receivers. The two received signals from a remote source in the presence of noise are modeled as

$$\begin{aligned} y_{s_1}(t) &= \chi_{s_1}(t) + n_{s_1}(t), \\ y_{s_2}(t) &= \alpha\chi_{s_1}(t+d) + n_{s_2}(t), \end{aligned} \tag{1}$$

where $\chi_{s_1}(t)$, $n_{s_1}(t)$, and $n_{s_2}(t)$ are real, stationary, and ergodic random processes, respectively, α is a scaling factor, and d is the TDOA. For an observation interval, T_g, and a TDOA sensor, s, an estimate of the cross-correlation as a function of time delay, τ, can be obtained as

$$\phi_s(\tau) = \frac{1}{T_g - \tau}\int_{\tau}^{T_g} y_{s_1}(t)y_{s_2}(t-\tau)\,\mathrm{d}t. \tag{2}$$

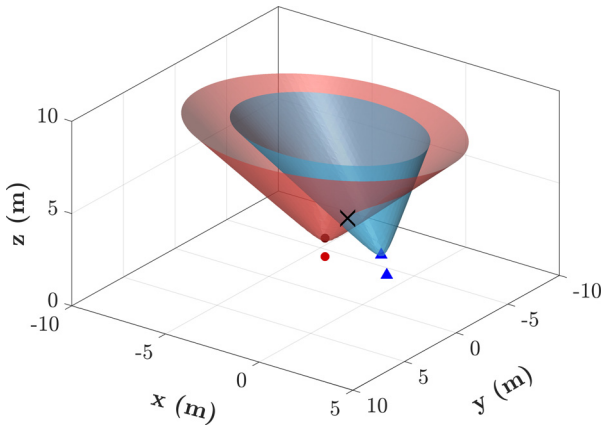

FIG. 1. (Color online) An example of acoustic source localization in 3-D from TDOA measurements. The cross is the location of the acoustic source. Each pair of dots and triangles represents a hydrophone pair that produces a TDOA measurement. Each TDOA measurement gives rise to a hyperboloid. The line resulting from the intersection of the two hyperboloids describes potential source locations. To obtain a unique source location, further hydrophone pairs are needed.

An estimate of the cross-power spectral density (CPSD), $\Phi_s(f)$, between the two signals is computed by taking the Fourier transform of the cross-correlation, i.e.,

$$\Phi_s(f) = \int_{\infty}^{-\infty} \phi_s(\tau)e^{-j2\pi f\tau}\,\mathrm{d}\tau = Y_{s_1}(f)Y^*_{s_2}(f), \tag{3}$$

where $Y_{s_1}(f)$ and $Y_{s_2}(f)$ are the Fourier transforms of $y_{s_1}(t)$ and $y_{s_2}(t)$, respectively, for the observation interval T_g.

The GCC (Ref. 24) is defined as the inverse Fourier transform of the frequency weighted CPSD, i.e.,

$$\hat{\phi}_s(\tau) = \frac{1}{2\pi}\int_{\infty}^{-\infty} \Gamma(f)\Phi_s(f)e^{j2\pi ft}\,\mathrm{d}f, \tag{4}$$

where $\Gamma(f)$ is the frequency weighting factor. Note that the GCC with $\Gamma(f) = H_{s_1}(f)H_{s_2}(f)$ can be interpreted as applying linear filters with frequency responses $H_{s_1}(f)$ and $H_{s_2}(f)$ to the signals $y_{s_1}(t)$ and $y_{s_2}(t)$, respectively, and subsequently performing a conventional cross-correlation.

Frequency weighting is performed according to specific optimization criteria. In applications where the noise power spectral density (PSD) is unknown, popular choices of frequency weighting factors result in the smoothed coherence transform (SCOT)[52] and phase transform (PHAT).[24] The SCOT normalizes the PSDs of the individual signals to unit magnitude, i.e., $\Gamma_{\mathrm{SCOT}}(f) = 1/(R_{s_1,s_1}(f)R_{s_2,s_2}(f))^{1/2}$, where R_{s_1,s_1} denotes the PSD of $y_{s_1}(t)$. Similarly, the PHAT normalizes the CPSD to unit magnitude, i.e., $\Gamma_{\mathrm{PHAT}}(f) = 1/|\Phi_s(f)|$.

In this work, it is assumed that the PSDs, $G_{s_1,s_1}(f)$ and $G_{s_2,s_2}(f)$, of the respective noise, $n_{s_1}(t)$ and $n_{s_2}(t)$, are mainly dominated by the instrument noise. It can be measured or precomputed and is, thus, considered known. The optimal frequency weighting for whitening the instrument noise is given by $\Gamma_{\mathrm{WIN}}(f) = 1/(G_{s_1,s_1}(f)G_{s_2,s_2}(f))^{1/2}$. The resulting GCC-WIN can be interpreted as applying the noise whitening filters $H_{s_1}(f) = 1/(G_{s_1,s_1}(f))^{1/2}$ and $H_{s_2}(f) = 1/(G_{s_2,s_2}(f))^{1/2}$ to the signals $y_{s_1}(t)$ and $y_{s_2}(t)$, respectively, and subsequently performing a conventional cross-correlation. A detailed discussion of the resulting GCC-WIN technique will be presented in Sec. IV.

To enhance the probability of detection, TDOA measurements can be either (i) computed based on a sequence of click trains[36,41] or (ii) obtained by extracting TDOAs of individual echolocation clicks and combining them into a single measurement. In the presence of highly correlated noise, the former approach is unsuitable; hence, the latter approach is considered. It is assumed that the odontocete is stationary over a time interval, T_m, which is longer than T_g, i.e., $T_m > T_g$. (T_m is the duration of the time between the discrete time step k of the considered tracking algorithms.) For each sensor, TDOAs of individual echolocation clicks are computed by finding the peaks of the GCC-WIN that are above a certain threshold, A_{tdoa}. The peaks of multiple observation intervals of length T_g are then accumulated over a time interval of duration T_m. The resulting set of TDOAs,

$z_{k,s}^{(m)}$, where $m_{k,s}$ is the number of measurements at time step k and sensor s and $m \in \{1, \ldots, m_{k,s}\}$, is considered as TDOA measurements of echolocation clicks generated by odontocetes. The TDOA measurements of all of the sensors are used as input for MTT.

III. MTT

A key challenge of MTT from sequences of measurements provided by one or multiple sensors is that the origin of measurements is typically unknown, i.e., it is not clear which target originated which TDOA measurement. This problem is referred to as measurement-origin uncertainty (MOU). Moreover, because the number of targets is also unknown, it has to be estimated directly from the data.

A. MTT with perfect measurement-to-target associations

Assuming that the origin of each measurement is perfectly known, i.e., measurement-to-target associations are either provided by a perfect human operator or a data association algorithm, the MTT problem can be split up into multiple parallel single target tracking problems. Here, a sequential Bayesian estimation or Bayes filter[53,54] is typically employed to estimate the state of each target individually and recursively. Define the target state and its associated measurements for all n_s sensors and at a discrete time step k as random vectors $\mathbf{x}_k$ and $\mathbf{z}_k = [\mathbf{z}_{k,1}^T, \ldots, \mathbf{z}_{k,n_s}^T]^T$, respectively. The target state typically consists of the target's position and motion-related parameters.

The objective is to estimate the target state, $\mathbf{x}_k$, from the available measurements up to time k, $\mathbf{z}_{1:k}$. Given the conditional PDF of the state given the measurements, $f(\boldsymbol{x}_k|\mathbf{z}_{1:k})$, the minimum mean square error (MMSE) estimate of the state of a single target, $\boldsymbol{x}_k$, can be found to be[55]

$$\hat{\boldsymbol{x}}_k^{\mathrm{MMSE}} = \int \boldsymbol{x}_k \, f(\boldsymbol{x}_k|\mathbf{z}_{1:k}) \mathrm{d}\boldsymbol{x}_k. \tag{5}$$

To obtain $f(\boldsymbol{x}_k|\mathbf{z}_{1:k})$, one could naively marginalize the available joint PDF $f(\boldsymbol{x}_{0:k}|\mathbf{z}_{1:k})$. This approach, however, suffers from the curse of dimensionality as the dimension of $\mathbf{x}_{0:k}$ grows with each time step. As a result, the computational complexity of naive marginalization increases exponentially and becomes intractable. The Bayes filter exploits that a first-order Markov process can describe a statistical model of single target tracking to reduce computations. At each time k, a prediction and n_s update steps are performed, and the resulting sequential processing schemes yields a computational complexity that is linear with time k.[53,54] MTT methods are sequential Bayesian estimation methods that also consider MOU and the unknown number of states to be estimated.

B. MTT with MOU and known number of targets

Consider a MTT problem with multiple sensors where the number of targets is known but measurements are subject to MOU. In addition, there are false positives, i.e., measurements that have not been generated by any target, and missed detections, i.e., present targets may not generate a measurement. It is assumed that there are $i \in \{1, \ldots, n_t\}$ targets. At time k, the state of the ith target is denoted as $\mathbf{x}_k^{(i)}$. For future reference, the notation $\mathbf{x}_k = [\mathbf{x}_k^{(1)\mathrm{T}}, \ldots, \mathbf{x}_k^{(n_t)\mathrm{T}}]^T$ is introduced. Each target state evolves independently according to the Markovian state-transition PDF, i.e., $f(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) = \prod_{i=1}^{n_t} f(\boldsymbol{x}_k^{(i)}|\boldsymbol{x}_{k-1}^{(i)})$.

Each sensor $s \in \{1, \ldots, n_s\}$ produces $m_{k,s}$ TDOA measurements $\mathbf{z}_{k,s} = [\mathbf{z}_{k,s}^{(1)\mathrm{T}}, \ldots, \mathbf{z}_{k,s}^{(m_{k,s})\mathrm{T}}]^T$. It is assumed that each measurement either originates from the target or is a false positive, and a target generates at most one measurement at each sensor. Measurement generation of target i at sensor s is modeled by a Bernoulli experiment characterized by the probability of detection, $p_{\mathrm{d}}^{(s)}(\boldsymbol{x}_k^{(i)})$. If the target with state $\boldsymbol{x}_k^{(i)}$ generates a measurement, $\mathbf{z}_{k,s}^{(m)}$, the measurement is distributed according to $f(\mathbf{z}_{k,s}^{(m)}|\boldsymbol{x}_k^{(i)})$. The number of false positives at each time step is Poisson distributed with a mean $\mu_{\mathrm{fp}}^{(s)}$. False positives are independent of the measurements that have originated from the targets and also independent and identically distributed (iid) according to the PDF $f_{\mathrm{fp}}^{(s)}(\mathbf{z}_{k,s}^{(m)})$.

At time k and sensor s, the unknown association between measurements and targets is modeled by the latent random vector, $\mathbf{a}_{k,s} = [\mathrm{a}_{k,s}^{(1)}, \ldots, \mathrm{a}_{k,s}^{(n_t)}]^T$, which is composed of random variables, $\mathrm{a}_{k,s}^{(i)}$, defined as[27]

$$\mathrm{a}_{k,s}^{(i)} = \begin{cases} m \in \{1, \ldots, m_{k,s}\}, & \text{at time } k \text{ and sensor } s, \\ & \text{target } i \text{ generates} \\ & \text{measurement } m, \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

For future reference, the joint vectors $\mathbf{a}_k = [\mathbf{a}_{k,1}^T, \ldots, \mathbf{a}_{k,n_s}^T]^T$ and $\mathbf{z}_k = [\mathbf{z}_{k,1}^T, \ldots, \mathbf{z}_{k,n_s}^T]^T$ are introduced.

The restriction that there can be at most one measurement associated with a target at every time step can be checked by the following indicator function:

$$\psi(\boldsymbol{a}_{k,s}) = \begin{cases} 0, & \exists i, i' \in \{1, \ldots, n_t\} \text{ such that} \\ & i \neq i' \text{ and } a_{k,s}^{(i)} = a_{k,s}^{(i')} \neq 0, \\ 1, & \text{otherwise.} \end{cases} \tag{7}$$

The marginal posterior PDF, $f(\boldsymbol{x}_k^{(i)}|\mathbf{z}_{1:k})$, used for state estimation in Eq. (5), could be found by marginalization, i.e., $f(\boldsymbol{x}_k^{(i)}|\mathbf{z}_{1:k}) = \int \sum_{\boldsymbol{a}_{1:k}} f(\boldsymbol{x}_{0:k}, \boldsymbol{a}_{1:k}|\mathbf{z}_{1:k}) \, \mathrm{d}\boldsymbol{x}_{\sim k}^{(i)}$, where $\boldsymbol{x}_{\sim k}^{(i)}$ is equal to $\boldsymbol{x}_{0:k}$ with state $\boldsymbol{x}_k^{(i)}$ removed. However, the computational complexity of this naive marginalization would, again, be infeasible due to the reasons discussed above. To reduce computational complexity, one can again exploit the fact that the posterior PDF, $f(\boldsymbol{x}_{0:k}, \boldsymbol{a}_{1:k}|\mathbf{z}_{1:k})$, follows a first-order Markov process, i.e.,

https://doi.org/10.1121/10.0017888

$$f(\boldsymbol{x}_{0:k},\boldsymbol{a}_{1:k}|\boldsymbol{z}_{1:k}) \propto f(\boldsymbol{x}_0)\prod_{k'=1}^{k} f(\boldsymbol{x}_{k'}|\boldsymbol{x}_{k'-1}) f(\boldsymbol{z}_{k'},\boldsymbol{a}_{k'}|\boldsymbol{x}_{k'}) = f(\boldsymbol{x}_0)\prod_{k'=1}^{k} f(\boldsymbol{x}_{k'}|\boldsymbol{x}_{k'-1})\prod_{s=1}^{n_s} f(\boldsymbol{z}_{k',s},\boldsymbol{a}_{k',s}|\boldsymbol{x}_{k'}). \tag{8}$$

Here, $f(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$ is the state-transition PDF discussed above, $f(\boldsymbol{x}_0)$ is an arbitrary prior at time $k = 0$, and $f(\boldsymbol{z}_{k,s}, \boldsymbol{a}_{k,s}|\boldsymbol{x}_k)$ is the conditional PDF that models the MOU measurement generation process. The last line of Eq. (8) uses the fact that the measurement generation given the target states is independent across the sensors. Note that $f(\boldsymbol{z}_{k,s}, \boldsymbol{a}_{k,s}|\boldsymbol{x}_k)$ is a function of $p_{\mathrm{d}}^{(s)}(\boldsymbol{x}_k^{(i)})$, $\mu_{\mathrm{fp}}^{(s)}$, $f_{\mathrm{fp}}^{(s)}(\boldsymbol{z}_{k,s}^{(m)})$, and $f(\boldsymbol{z}_{k,s}^{(m)}|\boldsymbol{x}_k^{(i)})$ (see Ref. 27 for details). Based on the factorization of the statistical model in Eq. (8), e.g., a sequential Bayesian estimation approach, which is referred to as probabilistic data association filter,[27] can be developed.

C. MTT with MOU and unknown number of targets

In real-world scenarios, such as the whale tracking problem, the number of targets is time-varying and unknown. To account for this, potential target (PT) states can be introduced.[32] Consider PTs with indexes $j \in \{1, \ldots, j_k\}$, where j_k is the number of PTs at time k. A binary variable, $\mathrm{r}_k^{(j)} \in \{0, 1\}$, indicates the existence of the PT j, where $\mathrm{r}_k^{(j)} = 1$ if and only if PT j exists. The augmented state of PT j is given by $\mathbf{y}_k^{(j)} = [\mathbf{x}_k^{(j)\mathrm{T}}\ \mathrm{r}_k^{(j)}]^{\mathrm{T}}$, where $\mathbf{x}_k^{(j)}$ consists of the target's position and further motion-related parameters. There are two types of PTs:

- *New PTs* represent targets that, for the first time, have generated a measurement. Their states are denoted by $\bar{\mathbf{y}}_{k,s}^{(j)} = [\bar{\mathbf{x}}_{k,s}^{(j)\mathrm{T}}\ \bar{\mathrm{r}}_{k,s}^{(j)}]^{\mathrm{T}}$. At time k and sensor s, a new PT is introduced for each measurement $m \in \{1, \ldots, m_{k,s}\}$; and
- *legacy PTs* represent targets that already have generated at least one measurement at a previous time step $k' < k$ or previous sensor $s' < s$. Their states are denoted by $\underline{\mathbf{y}}_{k,s}^{(j)} = [\underline{\mathbf{x}}_{k,s}^{(j)\mathrm{T}}\ \underline{\mathrm{r}}_{k,s}^{(j)}]^{\mathrm{T}}$.

The vectors that consist of all legacy and new PT states are denoted by $\underline{\mathbf{y}}_k$ and $\bar{\mathbf{y}}_k$, respectively, and the vector that consists of all PT states at time k is represented as $\mathbf{y}_k = [\underline{\mathbf{y}}_k \bar{\mathbf{y}}_k]^{\mathrm{T}}$.

At each time k, the number of targets that, for the first time, have generated a measurement at sensors s are Poisson distributed with mean $\mu_{\mathrm{n}}^{(s)}$. The states of these newly detected targets are iid with the PDF $f_n^{(s)}(\boldsymbol{x}_k^{(j)})$. Newly detected targets are statistically independent of existing targets. A PT j that existed at time $k-1$ continues to exist at time k with survival probability $p_{\mathrm{su}}(\boldsymbol{x}_k^{(j)})$. All of the PTs at time $k-1$ become legacy PTs at time k.

To reduce computational complexity, one can, again, exploit structure in the factorization of the posterior PDF $f(\boldsymbol{y}_{0:k}, \boldsymbol{a}_{1:k}|\boldsymbol{z}_{1:k})$. In particular, using common Markov assumptions,[32] $f(\boldsymbol{y}_{0:k}, \boldsymbol{a}_{1:k}|\boldsymbol{z}_{1:k})$ factorizes according to

$$f(\boldsymbol{y}_{0:k},\boldsymbol{a}_{1:k}|\boldsymbol{z}_{1:k}) \propto f(\boldsymbol{z}_{1:k},\boldsymbol{a}_{1:k},\boldsymbol{y}_{0:k}) = f(\boldsymbol{y}_0)\prod_{k'=1}^{k} f(\boldsymbol{z}_{k'},\boldsymbol{a}_{k'},\boldsymbol{y}_{k'}|\boldsymbol{y}_{k'-1}). \tag{9}$$

Upon explicitly distinguishing between the legacy and new PTs and by exploiting the chain rule for PDFs, the conditional PDF of current measurements, current association variables, and current states given the previous states can be expanded as

$$f(\boldsymbol{z}_k,\boldsymbol{a}_k,\boldsymbol{y}_k|\boldsymbol{y}_{k-1}) = f(\boldsymbol{z}_k,\boldsymbol{a}_k,\underline{\boldsymbol{y}}_k,\bar{\boldsymbol{y}}_k|\boldsymbol{y}_{k-1}) = f(\underline{\boldsymbol{y}}_k|\boldsymbol{y}_{k-1}) f(\boldsymbol{z}_k,\boldsymbol{a}_k,\bar{\boldsymbol{y}}_k|\underline{\boldsymbol{y}}_k,\boldsymbol{y}_{k-1}). \tag{10}$$

Now, one can use the fact that given the legacy PT states at time k, (i) the new PT states at time k are conditionally independent of the previous PT states at time $k-1$ and (ii) the measurements, association variables, and new PTs are statistically independent across the n_{s} sensors, i.e.,

$$f(\boldsymbol{z}_k,\boldsymbol{a}_k,\boldsymbol{y}_k|\boldsymbol{y}_{k-1}) = f(\underline{\boldsymbol{y}}_k|\boldsymbol{y}_{k-1})\, f\left(\boldsymbol{z}_k,\boldsymbol{a}_k,\bar{\boldsymbol{y}}_k|\underline{\boldsymbol{y}}_k\right) = f(\underline{\boldsymbol{y}}_k|\boldsymbol{y}_{k-1})\prod_{s=1}^{n_s} f\left(\boldsymbol{z}_{k,s},\boldsymbol{a}_{k,s},\bar{\boldsymbol{y}}_{k,s}|\underline{\boldsymbol{y}}_{k,s-1}\right), \tag{11}$$

where $\underline{\boldsymbol{y}}_{k,0} = \underline{\boldsymbol{y}}_k$ and $\underline{\boldsymbol{y}}_{k,s} = [\underline{\boldsymbol{y}}_{k,0}^{\mathrm{T}}, \bar{\boldsymbol{y}}_{k,1}^{\mathrm{T}}, \ldots, \bar{\boldsymbol{y}}_{k,s-1}^{\mathrm{T}}]^{\mathrm{T}}$.

In addition, targets move in time independently; therefore, the state-transition PDF of the legacy PTs reads

$$f(\underline{\boldsymbol{y}}_k|\boldsymbol{y}_{k-1}) = \prod_{j=1}^{j_{k-1}} f(\underline{\boldsymbol{y}}_k^{(j)}|\boldsymbol{y}_{k-1}^{(j)}). \tag{12}$$

The state-transition model $f(\underline{\boldsymbol{y}}_k^{(j)}|\boldsymbol{y}_{k-1}^{(j)})$ is a function of the survival probability $p_{\mathrm{su}}(\boldsymbol{x}_k^{(j)})$.

By plugging Eq. (12) into Eq. (11) and, in turn, Eq. (11) into Eq. (9), and making use of the functional form of $f(\boldsymbol{z}_{k,s}, \boldsymbol{a}_{k,s}, \bar{\boldsymbol{y}}_{k,s}|\underline{\boldsymbol{y}}_{k,s-1})$ (see Ref. 32 for details), the joint posterior PDF of $\mathbf{y}_{1:k}$ and $\mathbf{a}_{1:k}$ given an observed $\boldsymbol{z}_{1:k}$ becomes

$$f(\boldsymbol{y}_{1:k},\boldsymbol{a}_{1:k}|\boldsymbol{z}_{1:k}) \propto \prod_{k'=1}^{k}\left(\prod_{j'=1}^{j_{k'-1}} f(\underline{\boldsymbol{y}}_{k'}^{(j')}|\underline{\boldsymbol{y}}_{k'-1}^{(j')})\right)\prod_{s=1}^{n_s}\psi(\boldsymbol{a}_{k',s}) \times\left(\prod_{j=1}^{j_{k',s}} q(\underline{\boldsymbol{x}}_{k',s}^{(j)},\underline{r}_{k',s}^{(j)},a_{k',s}^{(j)};\boldsymbol{z}_{k',s})\right) \times\prod_{m=1}^{m_{k',s}} v(\bar{\boldsymbol{x}}_{k',s}^{(m)},\bar{r}_{k',s}^{(m)},a_{k',s}^{m}). \tag{13}$$

Here, the factors $q(\underline{\boldsymbol{x}}_{k,s}^{(j)},\underline{r}_{k,s}^{(j)},a_{k,s}^{(j)};\boldsymbol{z}_{k,s})$ and $v(\bar{\boldsymbol{x}}_{k,s}^{(m)},\bar{r}_{k,s}^{(m)},a_{k,s}^{m})$ are functions of $p_{\mathrm{d}}^{(s)}(\boldsymbol{x}_k^{(j)})$, $\mu_{\mathrm{fp}}^{(s)}$, $f_{\mathrm{fp}}^{(s)}(\boldsymbol{z}_{k,s}^{(m)})$, and $f(\boldsymbol{z}_{k,s}^{(m)}|\boldsymbol{x}_k^{(j)})$ (see Ref. 32 for details).

For target detection and estimation, the marginal PDFs $f(\boldsymbol{x}_k^{(j)}, r_k^{(j)}|\boldsymbol{z}_{1:k}) = f(\boldsymbol{y}_k^{(j)}|\boldsymbol{z}_{1:k})$ are required. In particular, target detection is performed by introducing a threshold p_{th} that is

compared with the existence probability, $p(r_k^{(j)} = 1|z_{1:k})$, i.e., a PT $j \in \{1, \ldots, j_k\}$ is declared to exist if $p(r_k^{(j)} = 1|z_{1:k}) > p_{\text{th}}$. Note that $p(r_k^{(j)} = 1|z_{1:k}) = \int f(\boldsymbol{x}_k^{(j)}, r_k^{(j)} = 1|z_{1:k}) \mathrm{d}\boldsymbol{x}_k^{(j)}$. For PTs declared to exist, state estimation is performed by computing the MMSE estimate[55] as

$$\hat{\boldsymbol{x}}_k^{(j)} \triangleq \int \boldsymbol{x}_k^{(j)} f\left(\boldsymbol{x}_k^{(j)} | r_k^{(j)} = 1, z_{1:k}\right) \mathrm{d}\boldsymbol{x}_k^{(j)}, \tag{14}$$

where

$$f\left(\boldsymbol{x}_k^{(j)} | r_k^{(j)} = 1, z_{1:k}\right) = \frac{f\left(\boldsymbol{x}_k^{(j)}, r_k^{(j)} = 1 | z_{1:k}\right)}{p\left(r_k^{(j)} = 1 | z_{1:k}\right)}. \tag{15}$$

For direct computation of $f(\boldsymbol{y}_k|z_{1:k})$, one could, again, naively marginalize the available joint PDF $f(\boldsymbol{y}_{1:k}, \boldsymbol{a}_{1:k}|z_{1:k})$. Instead, this marginalization can be performed efficiently by the framework of factor graphs and the SPA. The complete system model and the SPA for MTT can be found in Ref. 32.

The nonlinear measurement model and high-dimensional state space impose further challenges to tracking in 3-D from TDOA measurements; therefore, a SPA that embeds particle flow is employed. To perform the SPA effectively, particles are migrated toward regions of high likelihood based on the solution of a partial differential equation. This makes it possible to obtain good target detection and tracking performance in 3-D.[51]

IV. THE PROPOSED DATA PROCESSING CHAIN

The proposed data processing chain performs two main tasks: (i) signal processing and (ii) parameter estimation (Fig. 2). In the signal processing module, prefilters are first applied to the raw acoustic signal. Then, the GCC-WIN technique is performed to extract time delay peaks with the GCC-WIN amplitudes above a threshold.

The parameter estimation module estimates the locations and velocities of the echolocating odontocetes. The odontocetes are first tracked in the TDOA domain and then in 3-D. In both cases, the tracking algorithm based on the SPA described in Sec. III C is applied. In the TDOA domain, the motion and measurement models are linear. Tracking first in the TDOA domain makes it possible to reduce significantly the number of false positives. A low number of false positives is essential for successful tracking in 3-D. Finally, the output of the TDOA tracker is used as input for odontocete tracking in 3-D.

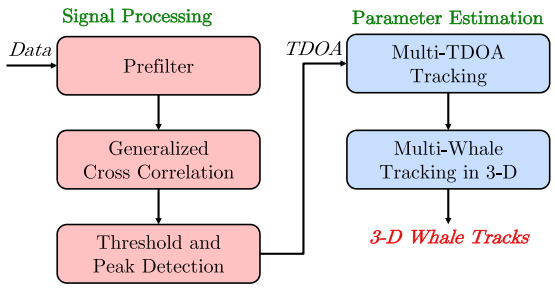

FIG. 2. (Color online) The block diagram of the proposed data processing chain for detecting and tracking odontocetes in 3-D from their echolocation clicks. The TDOAs of echolocation clicks are extracted in the signal processing module. The three-dimensional odontocete tracks are computed in the parameter estimation module. The parameter estimation module performs sequential Bayesian estimation for MTT in two stages.

A. Echolocation click detection and TDOA estimates

The echolocation clicks are characterized by their short pulse length and broad bandwidth. Depending on prior knowledge of the noise, different adaptations of the GCC can be used to detect the echolocation clicks and estimate the TDOA. In this work, it is desirable to maximize the number of detected echolocation clicks to reduce the duration and frequency of data gaps in time, which can significantly hinder the performance of the tracking algorithms. The echolocation clicks are challenging to detect if their SNR is low. The SNR depends on various factors such as the distance between the source and the hydrophones, the animal orientation and its echolocation click beam direction with respect to the receiver, as well as the ambient or system noise.

Each GCC of length T_g corresponds to n_g samples of the discrete time signal and eventually results in a discrete tracking time step, n. Following Eq. (1), the discrete received signals for sensor s, at time n, and with the TDOA h are given by

$$\begin{aligned} y_{s_1}[n] &= x_{s_1}[n] + n_{s_1}[n], \\ y_{s_2}[n] &= \alpha x_{s_1}[n+h] + n_{s_2}[n]. \end{aligned} \tag{16}$$

The discrete Fourier transform pairs of $y_{s_1}[n]$ and $y_{s_2}[n]$ are $Y_{s_1}[l]$ and $Y_{s_2}[l]$, respectively, where l is the discrete frequency. The GCC as a function of discrete time delay, m, is

$$\hat{\phi}_s[m] = \frac{1}{n_g} \sum_{l=0}^{n_g-1} \Gamma[l] \Phi_s[l] e^{j2\pi ml/N}, \tag{17}$$

where $\Phi_s[l] = Y_{s_1}[l]\, Y_{s_2}^*[l]$ is the CPSD estimate of the received signals, and $\Gamma[l]$ is the frequency weighting of the GCC.

In this work, it is assumed that an accurate estimate of the PSD of the instrument noise is available and can be used within the GCC-WIN. In particular, the CPSD estimate, $\Phi_s[l]$, is normalized by the PSD estimates of the known noise. Let $G_{s_1,s_1}[l]$ and $G_{s_2,s_2}[l]$ be the PSD estimates of the noise at the respective receivers. The frequency weighting used in the GCC-WIN then reads $\Gamma_s[l] = 1/(G_{s_1,s_1}[l]\, G_{s_2,s_2}[l])^{1/2}$. In case the statistics of the noise produced by the instrument are time-varying but periodic, as is the case for the HARP, a sequence of time-varying noise PSD estimates is extracted from precomputed spectrograms of the noise signal, i.e., from portions of signals without echolocation clicks (see the example in Fig. 3). A concrete implementation of this procedure for acoustic data provided by the HARP is presented in Sec. VI B 1.

https://doi.org/10.1121/10.0017888

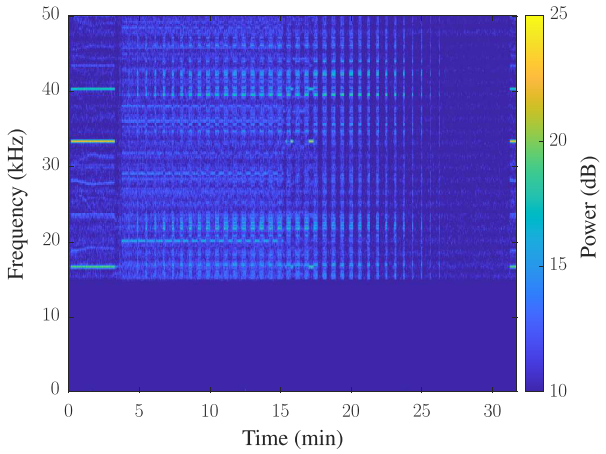

FIG. 3. (Color online) An example spectrogram of the instrument noise that repeats periodically, recorded by a hydrophone on a HARP on July 1, 2018. Each column of the spectrogram serves as an estimated noise PSD for the GCC-WIN. For each hydrophone, an individual spectrogram is extracted.

Finally, as described in Sec. II, TDOAs extracted from individual echolocation clicks are accumulated over a longer interval, T_{m}, to increase the probability of detection. Accumulating TDOAs, however, can lead to multiple TDOAs corresponding to one odontocete per time step k due to noise. Hence, a clustering of measurements is necessary to enforce the assumption that each odontocete generates at most one measurement. The following clustering procedure led to the best tracking results: First, large clusters are formed by finding TDOAs that are n_{c} samples apart and grouping. Then, within each large cluster, TDOAs are further separated into smaller clusters based on the local minima of the GCC-WIN amplitudes. Finally, the TDOAs within each small cluster are weighted based on their amplitudes and merged to generate a single TDOA measurement per small cluster.

B. MTT of echolocating odontocetes

In this section, details on how the MTT framework presented in Sec. III C is applied to the odontocete tracking are provided. In particular, the statistical models of the two MTT stages are discussed. The first stage operates in the TDOA domain and is applied to each hydrophone pair, i.e., a TDOA sensor, in parallel. The second stage operates in a three-dimensional Cartesian coordinate system and fuses the results provided by the first stage.

1. MTT in the TDOA domain

For the tracking in the TDOA domain, at time step k and sensor s, the state of the odontocete j is given by $\mathbf{d}_{k,s}^{(j)} = [\mathrm{d}_{k,s}^{(j)}\ \dot{\mathrm{d}}_{k,s}^{(j)}]^{\mathrm{T}}$, where $\mathrm{d}_{k,s}^{(j)}$ is the true TDOA of the echolocation clicks emitted by the odontocete, and $\dot{\mathrm{d}}_{k,s}^{(j)}$ is its rate of change. A nearly constant velocity motion model[53] is considered, i.e.,

$$\mathbf{d}_{k,s}^{(j)} = \begin{bmatrix} 1 & T_{\mathrm{m}} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mathrm{d}_{k-1,s}^{(j)} \\ \dot{\mathrm{d}}_{k-1,s}^{(j)} \end{bmatrix} + \mathbf{u}_{k,s}^{(j)}, \tag{18}$$

where the driving noise, $\mathbf{u}_{k,s}^{(j)} \in \mathbb{R}^2$, is a zero-mean multivariate Gaussian random vector with covariance matrix,

$$\mathbf{\Sigma}_{\mathbf{u}} = \begin{bmatrix} \frac{T_{\mathrm{m}}^3}{3} & \frac{T_{\mathrm{m}}^2}{2} \\ \frac{T_{\mathrm{m}}^2}{2} & T_{\mathrm{m}} \end{bmatrix} \sigma_{\mathbf{u}}^2, \tag{19}$$

and driving noise standard deviation (STD) $\sigma_{\mathbf{u}}$.

Let the TDOA measurement $\mathrm{z}_{k,s}^{(m)}$ be originated from the odontocete with index j at sensor s. The measurement model then reads

$$\mathrm{z}_{k,s}^{(m)} = \mathrm{d}_{k,s}^{(j)} + \mathrm{v}_{k,s}^{(m)}. \tag{20}$$

Here, $\mathrm{v}_{k,s}^{(m)}$ is a zero-mean Gaussian measurement noise with STD σ_{v}. The PDF that characterizes false positives, $f_{\mathrm{fp}}^{(s)}(z_{k,s}^{(m)})$, is uniform on the interval $[-T_s^{\max}\ T_s^{\max}]$, where $T_s^{\max}$ is the maximum time delay that can be measured by sensor s, i.e.,

$$T_s^{\max} = \|\boldsymbol{q}_{s_1} - \boldsymbol{q}_{s_2}\|/c, \tag{21}$$

where $q_{s_1} \in \mathbb{R}^3$ and $q_{s_2} \in \mathbb{R}^3$ are the positions of the hydrophone pair (s_1,s_2) that defines sensor s, and c is the sound velocity.

In the considered linear MTT problem, there is MOU, and the number of odontocetes is unknown and time-varying. Thus, the SPA-based MTT method described in Sec. III C is employed. The results of this first stage are sets of TDOA estimates $\hat{d}_{k,s}^{(j)}$, $j \in \{1,\ldots,j_{k,s}\}$ for each time step k and each sensor s. These TDOA estimates are then used as measurements in the three-dimensional MTT tracking stage. In what follows, the aforementioned set of TDOA estimates are denoted as $\hat{d}_{k,s}^{(m)}$, $j \in \{1,\ldots,m_{k,s}\}$ to indicate that they are now used as measurements in the second MTT stage. For simplicity, it is assumed that the measurements provided by different TDOA sensors are statistically independent of each other. Note, however, that dependencies among the measurements provided by the TDOA sensors do exist. This is because a TDOA sensor consists of a pair of hydrophones, and a hydrophone is involved in multiple TDOA sensors. Thus, the acoustic data of a single hydrophone is used in the measurements of multiple TDOA sensors.

2. MTT in 3-D

With the TDOA estimates available across all of the sensors, odontocetes are tracked in 3-D. The three-dimensional tracking method performs multisensor data association and track initialization. At time k, the state of the odontocete j is

denoted as $\mathbf{p}_k^{(j)} = [\mathrm{p}_{k,x}^{(j)} \mathrm{p}_{k,y}^{(j)} \mathrm{p}_{k,z}^{(j)} \dot{\mathrm{p}}_{k,x}^{(j)} \dot{\mathrm{p}}_{k,y}^{(j)} \dot{\mathrm{p}}_{k,z}^{(j)}]^{\mathrm{T}}$, where $\mathrm{p}_{k,x}^{(j)}$, $\mathrm{p}_{k,y}^{(j)}$, and $\mathrm{p}_{k,z}^{(j)}$ are the position of the odontocete in a 3-D Cartesian coordinate system and $\dot{\mathrm{p}}_{k,x}^{(j)}$, $\dot{\mathrm{p}}_{k,y}^{(j)}$, and $\dot{\mathrm{p}}_{k,z}^{(j)}$ are the respective velocities.

The motion model follows the nearly constant velocity model such that[53]

$$\mathbf{p}_k^{(j)} = \begin{bmatrix} \boldsymbol{I}_3 & T_m \boldsymbol{I}_3 \\ \mathbf{0}_3 & \boldsymbol{I}_3 \end{bmatrix} \mathbf{p}_{k-1}^{(j)} + \begin{bmatrix} \frac{T_m^2}{2} \boldsymbol{I}_3 \\ T_m \boldsymbol{I}_3 \end{bmatrix} \mathbf{w}_k^{(j)}, \tag{22}$$

where $\mathbf{w}_k^{(j)} \in \mathbb{R}^3$ is a zero-mean Gaussian driving noise with covariance $\boldsymbol{I}_3 \sigma_{\mathbf{w}}^2$ and driving noise STD $\sigma_{\mathbf{w}}$.

Suppose that the odontocete j generated the TDOA $\hat{\mathrm{d}}_{k,s}^{(m)}$ at time k and sensor s. The corresponding TDOA model is given by

$$\hat{\mathrm{d}}_{k,s}^{(m)} = \left(\| \boldsymbol{p}_k^{(j)} - \boldsymbol{q}_{s_1} \| - \| \boldsymbol{p}_k^{(j)} - \boldsymbol{q}_{s_2} \| \right) / c + \mathrm{b}_{k,s}^{(m)}, \tag{23}$$

where $\mathrm{b}_{k,s}^{(m)}$ is a zero-mean Gaussian measurement noise with STD σ_{b}. The PDF that characterizes false positives, $f_{\mathrm{fp}}^{(s)}(z_{k,s}^{(m)})$, is again uniform on the interval $[-T_s^{\max}\ T_s^{\max}]$.

Note that the nonlinear measurement model in Eq. (23) is underdetermined, i.e., the position of the odontocete is three-dimensional while the TDOA is only one-dimensional. Therefore, for estimating the three-dimensional position, TDOAs provided by multiple sensors have to be fused. In particular, for a fixed TDOA, $\hat{d}_{k,s}^{(m)}$, Eq. (23) describes potential odontocete locations on a hyperboloid. If there were no MOU and no noise, one could estimate the three-dimensional odontocete location by computing the intersection of multiple hyperboloids provided by multiple TDOA sensors. However, due to the presence of MOU and unknown number of odontocetes, reliable state estimation can only be performed sequentially. In particular, the MTT approach reviewed in Sec. III C is used again. Because of the nonlinear measurement model [Eq. (23)], a particle-based implementation of this MTT approach is considered. However, as the conventional particle filter suffers from weight degeneracy resulting from the curse of dimensionality[56] and is, thus, only suitable for low-dimensional state spaces, the particle flow variant recently proposed in Ref. 51 is used.

Particle flow techniques are attractive methods that address the weight degeneracy issues.[50] For each odontocete j and time k, particles are migrated iteratively to the posterior distribution. Particle motion follows a stochastic process described by an ordinary differential equation that expresses a Bayesian update step. The flow equations and particles after the flow describe a proposal distribution that can be used for importance sampling.[57] By embedding particle flow into SPA-based MTT, the weight degeneracy is avoided, and MTT in 3-D from TDOA measurements can be performed with a reasonable number of particles and computational complexity.[51]

V. SIMULATION

Intrinsic challenges of tracking marine animals with PAM are the lack of ground truth tracks and the fact that the manual tracks are imperfect because the data annotation process is subjective. Therefore, the three-dimensional tracking performance of different benchmark approaches from simulations are compared to motivate the proposed 3-D tracking approach.

Four sets of 200 Monte Carlo simulations of odontocete tracks are generated and tracked using three different approaches. Each set of Monte Carlo simulations has an increasing number of simultaneously present odontocetes, ranging from one to four. The odontocetes' starting positions are placed uniformly on a circle of radius 1000 m on an xy-plane and at a depth of 1000 m. TDOA measurements are generated based on the same array geometry described in Sec. VI A and from the model described in Eq. (23). False positives and missed detections are also generated based on the model discussed in Sec. VI B 2. The same hyperparameters from Table I are used to generate the simulated data. Each simulation is 85 discrete time steps long, and the time step has a duration of 7 s. An odontocete is present for 50 time steps. In simulations with multiple odontocetes, an odontocete is introduced every ten steps.

Three tracking approaches are simulated by: (i) an approach based on nonsequential tracking[25] (NST), (ii) a single Bernoulli tracker[58] (SBT), and (iii) the proposed MTT approach described in Sec. IV B 2. In what follows, the approaches (i), (ii), and (iii) will be referred to as (NST), (SBT), and (MTT), respectively. (NST) computes odontocete positions by combining the DOAs of the echolocation clicks of the odontocete computed at each array but does not filter the results, i.e., odontocete positions are computed at each discrete time step individually. Moreover, the localization and DOA computations are performed within the least squares framework. (SBT) is a recursive Bayesian filter that can detect and track a single target in the presence of missed detections and false positives. (NST) and (SBT) cannot track multiple targets in an automated way; thus, (NST) and (SBT) are provided with the correct data association solution and applied to each odontocete track individually. (NST) and (SBT) only use the true measurements corresponding to a single odontocete subject to missed detections.

TABLE I. Hyperparameters used for the tracking of multiple Cuvier's beaked whales in the TDOA domain and in 3-D.

Hyperparameters	TDOA domain	3-D
Detection probability, p_{d}	0.80	0.80
Survival probability, p_{su}	0.90	0.99
Mean number of false positives, μ_{fp}	10	1
Mean number of whale birth	1.0×10^{-4}	1
Measurement noise STD, σ_{v} and σ_{b}	1.0×10^{-5}	3.0×10^{-5}
Driving noise STD, $\sigma_{\mathbf{u}}$ and $\sigma_{\mathbf{w}}$	1.5×10^{-7}	1.0×10^{-2}
Number of particles	30 000	100 000
Minimum track length	20	5

https://doi.org/10.1121/10.0017888

If there is a missed detection, (NST) performs interpolation to obtain a measurement for each odontocete at each TDOA sensor. (SBT) relies on the same statistical model and parameters as (MTT). On the other hand, (MTT) does not know the correct association and, therefore, performs data association automatically. It is provided with the TDOA measurements generated for all present odontocetes following the model for missed detections and false positives discussed in Sec. IV B 2. It is worth noting that the correct data association solution is only available if TDOA measurements are synthetically generated. In a real-world scenario, the correct data association solution is unavailable and can only be approximated by a human operator.

Each track's root mean square error (RMSE) is computed for each approach mentioned above (Fig. 4). (MTT) and (SBT) sometimes do not detect the odontocete or have a significant error due to false positives or missed detections. In that case, the RMSE is set to an error value of 110 m, which is approximately twice the average RMSE related to (NST). Based on the simulation results, (NST) has the highest RMSE as it does not perform any filtering. On the other hand, (SBT) yields the lowest RMSE because it relies on the correct data association solution and makes use of a statistical model for measurement noise, missed detection, and false positives. (MTT) yields a RMSE between those from (NST) and (SBT). (MTT) makes use of the same statistical model as does (SBT) but also has to solve the data association problem; hence, it is expected to perform worse than (SBT). However, since (MTT) performs filtering, it still performs better than (NST). Furthermore, the RMSE of (MTT) increases with the number of odontocetes because the presence of multiple odontocetes makes the data association problem more challenging. Recall that (NST) and (SBT) require manual annotations in reality where the correct data association solution is unavailable, whereas (MTT) performs data association in an automated manner. Thus, (MTT) is well-suited for tracking multiple odontocetes automatically.

Sometimes a different number of odontocete tracks are generated by (MTT). Either a single track is broken into two tracks, or extra tracks are generated from false positives. The percentage of the number of simulations in which extra numbers of tracks are generated are 0%, 1%, 11%, and 19% for scenarios with one odontocete, two odontocetes, three odontocetes, and four odontocetes, respectively. Additional tracks can, indeed, be formed with the real data but in the final stage, either a human operator or an algorithm would join broken tracks or prune unlikely tracks.

In Fig. 5, the average RMSE over 800 tracks is shown as a function of the time step for the scenario with four odontocetes. The RMSE of the initialization phase of (MTT) and (SBT) are high, but after a few steps, they converge to a smaller error value compared to that of (NST). The real data tracks are also longer than 50 time steps; therefore, the total error is expected to be even smaller for longer tracks. The simulation outcome shows that the proposed approach can generate trustworthy odontocete tracks.

VI. REAL DATA APPLICATION

The tracking capability of the proposed data processing chain is demonstrated on acoustic datasets containing echolocation clicks from Cuvier's beaked whales. In the signal processing step, their echolocation clicks are detected, and the corresponding TDOA measurements are computed using the GCC-WIN algorithm described in Sec. IV A. In the parameter estimation step, the whales are tracked first in the TDOA domain and then in the three-dimensional domain using the implementation of the MTT framework described in Sec. IV B. The three-dimensional tracking results are compared to the tracks generated from hand-annotated DOA measurements following the approach in Ref. 25.

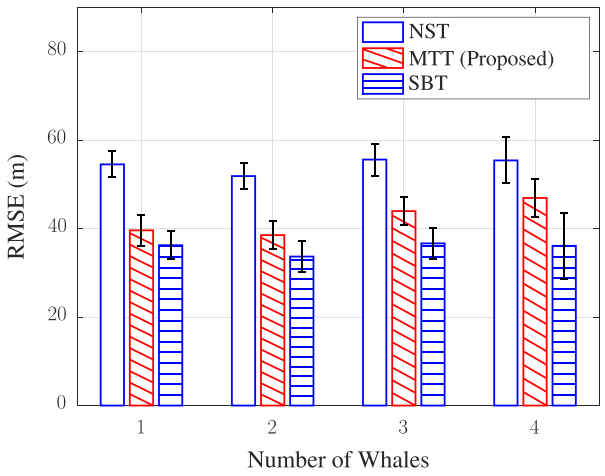

FIG. 4. (Color online) The average tracking RMSE across 200 Monte Carlo runs versus number of simultaneously present odontocetes for three different three-dimensional tracking approaches. The error bars denote the 75th percentile of the measured RMSE. While the correct data association solution is available to (NST) and (SBT), it is unknown to the proposed (MTT) method. In real data processing, the correct association solution is unavailable and can only be approximated by a human operator.

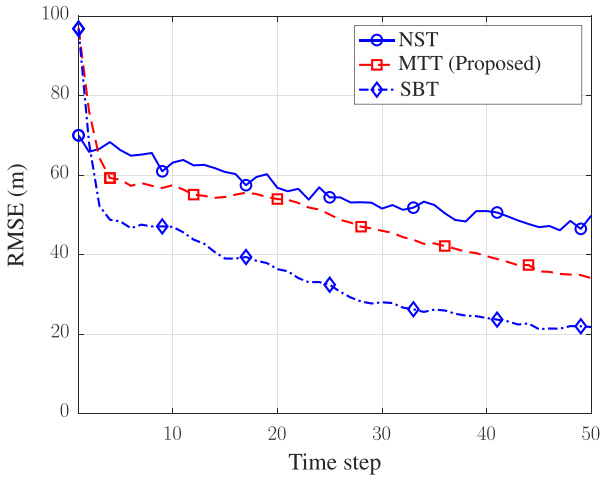

FIG. 5. (Color online) The average RMSE versus time for three different three-dimensional tracking approaches. Two hundred Monte Carlo runs of a scenario with four odontocetes are considered. (SBT) and (MTT) perform sequential processing and, thus, lead to reduced RMSE values over time.

A. Data

This study uses acoustic signals measured on two HARPs,[59] each of which is equipped with four hydrophones. The sampling frequency is 100 kHz; the corresponding Nyquist frequency is 50 kHz, and the TDOA resolution is 10 μs. The hydrophones on the HARP are 1 m apart and arranged in a tetrahedral shape (see Fig. 2 in Ref. 3) to form a small-aperture array. They were deployed off of the coast of California (32° 39′ 31.4″ N, 119° 28′ 37.6″ W) at a depth of ∼1330 m, and the arrays were approximately 1 km apart. The arrays located east and west are referred to as HARP EE and HARP EW, respectively, and they recorded from March 15 to July 18, 2018.

Five encounters of Cuvier's beaked whales are processed; in this paper, the results from processing the data measured on June 11 and July 1 in 2018 are presented. The encounter durations on those two dates are 52 min long and 20 min long, respectively. These encounters were detected using the long-term spectral average (LTSA) from the MATLAB-based program *Triton*.[59]

Because the hydrophones on each HARP share the same clock, they are time-synchronized. The two arrays, however, are not synchronized; therefore, the TDOA measurements are extracted only from the hydrophone pairs on the same arrays. Nonetheless, as long as the echolocation clicks are measured on both HARPs within a time window in which the Cuvier's beaked whale can be considered stationary, precise synchronization is unnecessary.[17,36]

In addition, the considered water depth is far below the thermocline with minimal change in the sound velocity as a function of depth. Hence, an acoustic wave propagation with spherical spreading in an isovelocity medium is assumed. The sound velocity is estimated to be 1490 m s^{-1}.

Different species of vocalizing marine animals could be present simultaneously near the PAM instruments, and their bioacoustic signals could interfere with one another. In such a case, the proposed data processing chain would be extended with classifying algorithms to discern among the species. However, there was no interference from other marine animals in the datasets used; hence, the classification step was not required. When inspected manually, the echolocation clicks followed the characteristics of those from the Cuvier's beaked whales described in Ref. 14.

B. Implementation

1. Signal processing

The GCC-WIN technique is used to detect the echolocation clicks and estimate the corresponding TDOAs. Three noise sources are identified: a pulse signal from the copresent acoustic Doppler current profiler (ADCP), an instrument noise on HARP highly correlated among hydrophone measurements, and an ambient noise (Fig. 6). Note that the instrument noise is harmonic and broadband thereby giving rise to multiple high amplitude peaks at wrong time delay locations when these signals are cross-correlated.

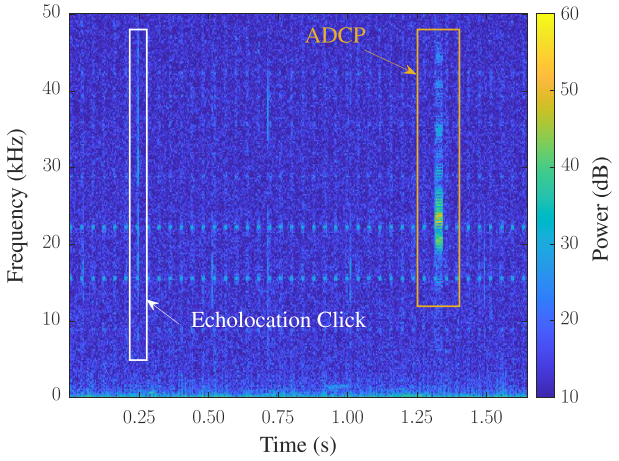

FIG. 6. (Color online) An example spectrogram of acoustic data from a single hydrophone with echolocation clicks and noise. The identified noise sources are the ADCP, instrument noise from the data storage system, and undersea environment. The echolocation clicks are broadband signals whose center frequency is approximately 36 kHz. The ADCP signal is also a broadband signal with a center frequency near 25 kHz. The instrument noise is a harmonic narrowband signal that is repeated every 31 s. Its spectrogram is displayed in Fig. 3.

This would hinder the tracker's performance significantly. The ambient noise is dominant below 2 kHz, whereas the instrument noise is the major source of noise above 2 kHz. The echolocation clicks of a Cuvier's beaked whale are known to have center frequency at 35.9 kHz and a −10 dB bandwidth of 10.9 kHz.[14] Hence, the signals above 15 kHz are considered for this study.

The signal is first prefiltered with a high pass filter based on a Parks-McClellan optimal finite impulse response filter design,[60] which is applied with a zero-phase shifting digital filter.[61] The stop band frequency and passband frequency are 13 kHz and 15 kHz, respectively. Then, the ADCP signal is identified and removed. A nearby ADCP generated a short dominant pulse recurring approximately every 54 s, which completely flooded the acoustic measurements. Its center frequency is at 75 kHz, but as the Nyquist frequency of the instrument is 50 kHz, it is aliased and present at 25 kHz. The ADCP signal was too powerful that it was not fully suppressed by the anti-aliasing filter onboard. It is, nonetheless, readily identified and nulled based on its high energy and center frequency characteristics.

To perform the GCC-WIN, a model of the noise in the form of a PSD is required. With the considered HARP instrument, the PSD of the noise is time-varying but periodic with a period of approximately 31 s. The spectrogram of the noise model is referred to as the noise PSD template (see an example in Fig. 3). The noise PSD template is estimated for each hydrophone individually from 20 min long signals before the appearance of echolocation clicks in the data. This segment is chosen manually by inspecting the LTSA. The 20 min long signal is further divided into shorter segments that are 31.65 s long. Their spectrograms are then averaged to estimate the noise PSD template. All of the spectrograms are generated with the fast Fourier transform

https://doi.org/10.1121/10.0017888

(FFT) length of $N = 512$ samples, i.e., $T_g = 5.12$ ms (50% overlap, Hamming window[62]). Separate noise PSD templates are estimated for different days of data because the noise statistics vary throughout the deployment. The main source of instrument noise is believed to be the mechanical noise of the hard disks from the data storage system.

Before performing the GCC-WIN, a critical step is to align the noise PSD template with the instrument noise of the signals of interest. This alignment step is based on the averaged noise template in the time domain, which is also computed for each hydrophone individually from the 20 min long signals used for estimating the noise PSD template. First, any pulse-like components (e.g., echolocation clicks) that could interfere with the alignment process are smoothed out. Smoothing is performed by applying a moving average filter to the signal of interest and replacing components with amplitudes that are at least twice as large as the STD of the moving averaged signals with the mean amplitude. Then, the smoothed signal is cross-correlated with the noise template in the time domain to identify the locations of the peaks. These locations are the time steps at which the noise PSD template is aligned with the noise in the signal of interest. The peaks of the cross-correlation are expected to repeat approximately 31 s, which is equal to the period of the noise PSD.

Once the noise template and signals are aligned, the GCC-WIN technique is performed as described in Sec. IV A. The spectrogram of the signal of interest is similarly generated using segments of $N = 512$ samples (50% overlap, Hamming window). The resultant peaks from the GCC-WIN, whose amplitudes are more significant than $A_{tdoa} = 0.15$, are extracted and saved along with their amplitude information. Note that because a low amplitude threshold is applied to maximize the detection rate, the instrument and white noise in the background can give rise to multiple false positive TDOA measurements. When the GCC-WIN peak amplitude is more than ten, i.e., a direct echolocation click is detected, any pulse-like signals within the next 40 ms are ignored as they are likely to be multipath signals.

The TDOA measurements are merged and binned with a longer discrete time step length of $T_m = 7$ s and clustering distance $n_c = 2$ samples, following the method described in Sec. IV A. Given that inter-click-intervals (ICIs) of Cuvier's beaked whales range between 0.3 and 0.9 s,[3] the probability of detection is increased by inspecting over a longer time window. This also helps the tracker to be robust against the irregular nature of the ICIs as the Cuvier's beaked whale is diving.[3] Moreover, using a step length of 7 s instead of a step length of 5.12 ms, which is the step length of performing one GCC-WIN, reduces the processing time by a factor of approximately 1300. As the average speed of the Cuvier's beaked whale is 1.2 m s^{-1},[3] it would have moved approximately 8.4 m within 7 s. Assuming that the Cuvier's beaked whales are primarily hundreds of meters away from the arrays, the corresponding TDOA measurements are unlikely to change significantly during this period.

2. Parameter estimation

Details for obtaining the tracks of the Cuvier's beaked whales are provided. Because the TDOA is computed between a pair of hydrophones, there are $\binom{4}{2} = 6$ TDOA sensors per array and $n_s = 12$ sensors total. As described earlier, the TDOA measurements are accumulated over $T_m = 7$ s. The Cartesian coordinate system follows the east, north, and up (ENU) convention, where the x, y, and z axes are positive along the ENU directions, respectively. The origin is between the two arrays and at the sea surface. The hyperparameters used for TDOA tracking and three-dimensional tracking are summarized in Table I. For TDOA tracking, the birth distribution is chosen to be uniformly distributed between the minimum and maximum possible TDOAs of a hydrophone pair. For three-dimensional tracking, the birth distribution is chosen to be uniformly distributed on the three-dimensional region of interest. A final pruning step is used to remove extra and unreasonable tracks. The average swim speed (Euclidean norm of the estimated velocity) of a Cuvier's beaked whale is 1.2 m s^{-1},[1,17] and its horizontal speed could range from 1 to 3 m s^{-1}.[3] Consequently, if the median speed of the whale track is more significant than 2 m s^{-1} or the track length is shorter than five time steps, the track is discarded. Furthermore, at every time step, if its estimated speed is faster than 3.5 m s^{-1}, the state at that time step is ignored.

C. Results

The tracks in the TDOA and three-dimensional domains using the datasets from two different dates are presented. The results using the proposed method are compared to the tracking results from hand-annotated data using the framework proposed in Ref. 25, where the detected TDOAs are used to compute the azimuth and elevation angles of the Cuvier's beaked whale relative to each array. They are tracked manually in the azimuth and elevation angles domain, whose tracks are fused between two hydrophones to estimate their three-dimensional locations.

Two Cuvier's beaked whales are identified and tracked from the data recorded on June 11, 2018. While echolocation clicks from only one Cuvier's beaked whale are detected from the TDOA data of HARP EE, those from two whales are detected from the TDOA data of HARP EW (Fig. 7). Based on the three-dimensional tracks (Figs. 8 and 9), the Cuvier's beaked whale that first appeared in the data is far from HARP EE, resulting in the lack of its detected echolocation clicks on HARP EE. To verify further, the corresponding TDOA measurements from the first Cuvier's beaked whale three-dimensional track are computed using the TDOA model, and they are in accordance with the tracked TDOA. The manual tracking method could not generate the track for the first Cuvier's beaked whale because it requires that the DOAs exist on both arrays simultaneously.

In the data collected on July 1, 2018, two Cuvier's beaked whales are observed again (Figs. 10–12). Their diving

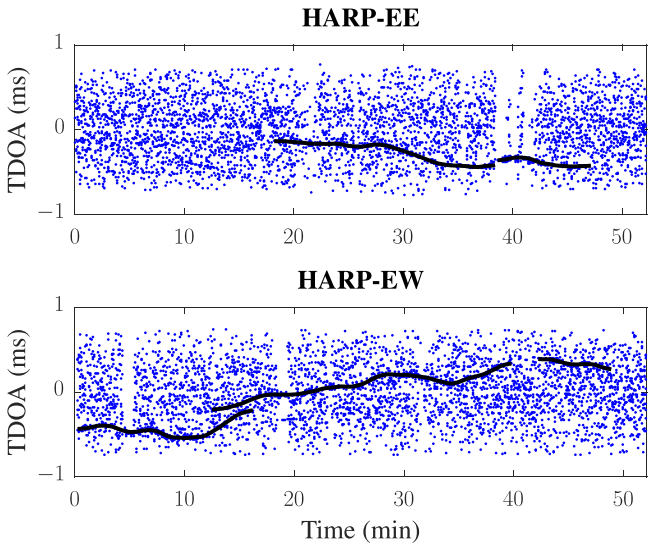

FIG. 7. (Color online) Example TDOA measurements (dots) and corresponding TDOA tracking results (solid lines) from one TDOA sensor of HARP EE (top) and one TDOA sensor of HARP EW (bottom) are shown. Two Cuvier's beaked whales are observed at HARP EW while only one is observed at HARP EE. As can be seen in Fig. 8, both Cuvier's beaked whales are closer to HARP EW. The considered acoustic data were collected on June 11, 2018.

behaviors are detected at the initial depths of approximately 450 m. The two tracks start close to each other, but the tracker is capable of separating them. More echolocation clicks are detected using the GCC-WIN algorithm. As a result, the proposed method yields longer three-dimensional tracks (approximately by 5 min) than the tracks from the hand-annotated data. To verify the reliability of the results, the three-dimensional tracker is run again with the TDOA data in the reverse time steps. As it is easier for the tracking algorithm to identify two

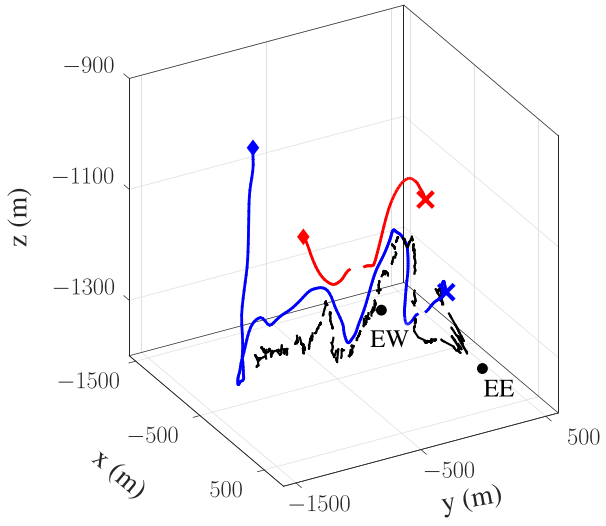

FIG. 8. (Color online) A comparison between a track generated from the hand-annotated data (dashed line) and estimated tracks provided by the proposed MTT approach (solid lines) in 3-D. With MTT, an additional track is extracted from the TDOA data of HARP EW, and the diving behavior of a Cuvier's beaked whale can be explored. The diamond and the cross indicate each track's start and end, respectively. The data from June 11, 2018, are considered.

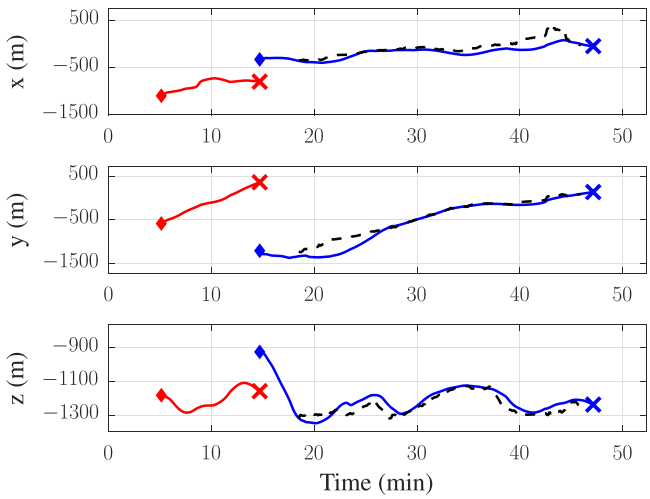

FIG. 9. (Color online) A comparison between the track generated from the hand-annotated data (dashed line) and estimated tracks provided by the proposed MTT approach (solid lines). Each axis of the three-dimensional domain is shown individually. The data from June 11, 2018, are considered.

Cuvier's beaked whales if they are spatially distinct in the beginning, the tracks from the reverse order would be more reliable in this scenario. The reverse order tracking results are indeed similar to the tracking presented here, verifying the correctness of the proposed tracker.

VII. DISCUSSION

A data processing chain for automatically detecting and tracking multiple odontocetes from their echolocation clicks is developed. It successfully tracks multiple Cuvier's beaked whales in 3-D from their echolocation click recordings made on a pair of volumetric hydrophone arrays. No human

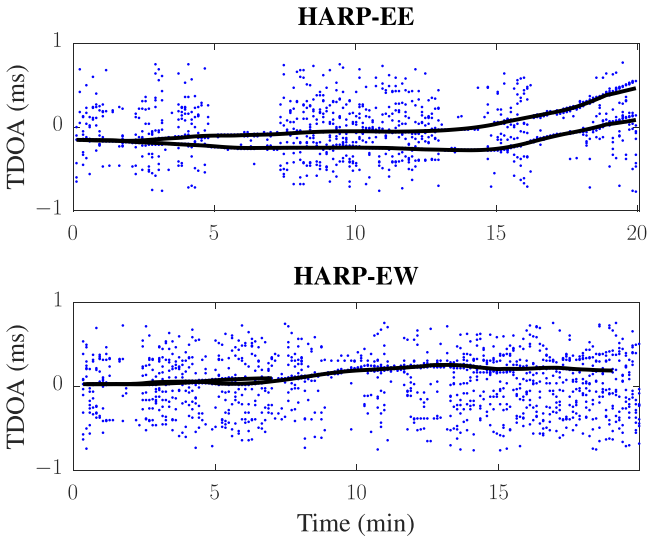

FIG. 10. (Color online) Example TDOA measurements (dots) and corresponding TDOA tracking results (solid lines) from one TDOA sensor of HARP EE (top) and one TDOA sensor of HARP EW (bottom) are shown. One of the Cuvier's beaked whales is not successfully tracked toward the end by HARP EW because of long gaps of missing TDOA. The considered acoustic data were collected on July 1, 2018.

https://doi.org/10.1121/10.0017888

operators or heuristics are needed to initialize the tracks and combine the TDOA measurements corresponding to the individual Cuvier's beaked whales. The graph-based MTT method efficiently solves the data association problem to fuse the TDOA measurements among missed detections and false positives and tracks odontocetes in a computationally tractable manner. In addition, new and more extended tracks of the Cuvier's beaked whales could be extracted using the GCC-WIN algorithm, which normalizes the CPSD by the estimated instrument noise PSD to whiten the instrument noise.

A few factors need to be considered when adapting the processing chain for applications with another set of bioacoustic data of odontocetes. Because the tracks are distinguished based on spatial information, the data processing chain is species-agnostic. The preprocessing step needs careful customization, e.g., T_g, frequency filters, etc., such that the echolocation clicks from species of interest are processed. A manual or automated classification step would be desirable for scientific purposes. In addition, the GCC-WIN technique is only recommended if the instrument noise is dominant and can be estimated; otherwise, other GCC adaptations, such as SCOT and PHAT, are potentially better suited for the estimation of the TDOAs.

The data processing chain could be further extended to yield more accurate and longer tracks of odontocetes. As observed in the second scenario (Fig. 11), closely spaced tracks, i.e., a track coalescence, can pose a challenge. To mitigate, the future model in the MTT framework could incorporate statistics of ICIs to distinguish among multiple echolocating odontocetes. It is observed that the clustering described in Sec. IV A to ensure the assumption of a single measurement per target has hindered the track results under coalescence.

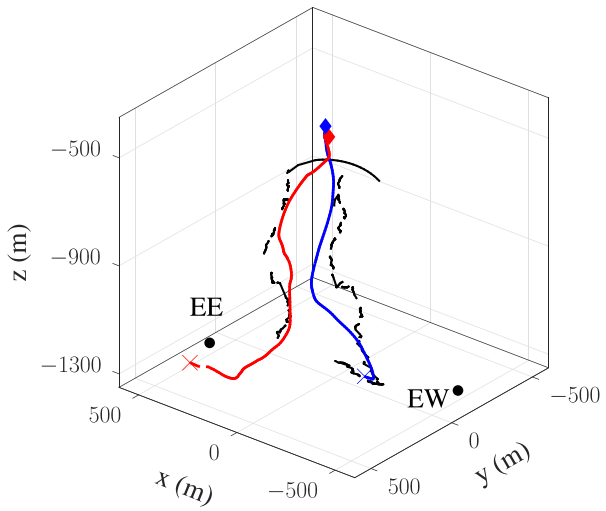

FIG. 11. (Color online) A comparison between the tracks generated from the hand-annotated data (dashed lines) and estimated tracks provided by the proposed MTT approach (solid lines) in 3-D. Two closely spaced Cuvier's beaked whales are simultaneously diving into deeper waters. The diamond and the cross indicate each track's start and end, respectively. The proposed MTT approach can successfully distinguish the two individuals. The data from July 1, 2018, are considered.

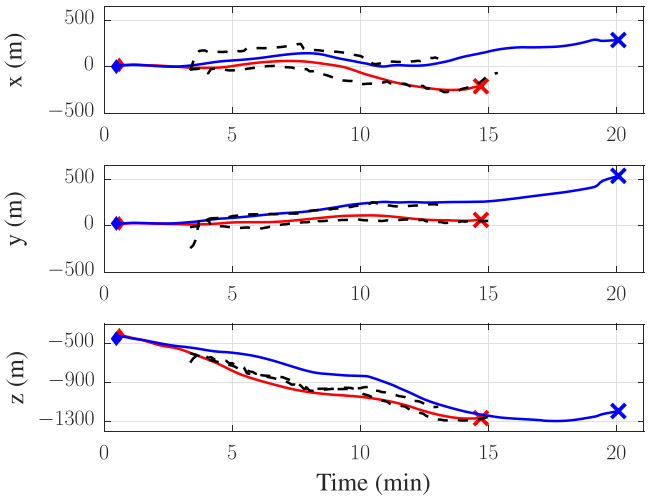

FIG. 12. (Color online) A comparison between the track generated from the hand-annotated data (dashed line) and estimated tracks provided by the proposed MTT approach (solid lines). Each axis of the three-dimensional domain is shown individually. The data from July 1, 2018, are considered.

Another challenge is that the model based on a constant probability of detection is not entirely reflective of reality. Even though the probability of detection is affected by the SNR of the bioacoustic signal, their irregularity, i.e., occurrences of a burst of echolocation clicks followed by long gaps of silence, needs to be considered. For example, there is approximately a 2 min long gap of TDOA detections starting at 40 min in HARP EW in Fig. 7 that is likely due to the Cuvier's beaked whale looking away from the array. With prior information on the echolocation click directionality and motion of the odontocetes at each given time, the detector's performance could be modeled more accurately. Alternatively, a time-varying detection probability could be estimated together with the target states.[63]

Furthermore, although the used, nearly constant velocity motion model of the odontocetes yielded good tracking results in this work, a tracking algorithm that selects the most suitable motion model for the odontocete motion during runtime is desirable. By incorporating interacting multiple models,[64] the tracker would not only capture time-varying behaviors or motions of the odontocetes more accurately but also be more suitable for tracking different species of odontocetes.

Finally, some species of odontocetes are highly sociable and, thus, sometimes move together in close spatial proximity.[65,66] In such a case, it would be challenging to track an individual odontocete, but subgroups could be tracked.[15] The question of the feasibility of the MTT frameworks to follow pods of odontocetes instead of individuals poses an interesting research problem.

VIII. CONCLUSION

Tracking acoustically active marine animals using passive acoustics can provide better understanding of their behaviors underwater, which are difficult to observe otherwise. However, human operators are often required to

annotate the bioacoustic data to associate the acoustic recordings with the marine animals that generated the signals. With an increasing amount of available underwater bioacoustic recordings and ultimate goal of real-time monitoring, tracking processes that involve human operators, who are imperfect, are not sustainable. Hence, it is crucial to research automated methods that are computationally tractable and accurate for passive acoustic tracking of marine animals. This will not only generate tracks that are more objective but also facilitate scientists to study the marine animals more efficiently.

In this paper, a data processing chain for the fully automated detection and tracking of odontocetes in 3-D from echolocation clicks is developed. The detection rate of the echolocation clicks is improved by using a GCC algorithm designed to whiten the instrument noise. Multiple odontocetes are detected and tracked simultaneously by applying two stages of a graph-based tracking method that efficiently solves the data association problem. Graph-based detection and tracking are first performed for each hydrophone pair individually in the TDOA domain and, subsequently, in the three-dimensional domain. The ability to track multiple odontocetes simultaneously without manual data selection by a human operator is demonstrated based on real acoustic data provided by two volumetric hydrophone arrays. In particular, tracking results in scenarios with two echolocating Cuvier's beaked whales (*Ziphius cavirostris*) are presented. In addition, the simulation results suggest that the presented processing chain can be used to track a larger number of whales in a scalable manner and is, thus, particularly appealing for future PAM systems. These results show how the proposed data processing chain can simplify scientists' steps to study the deep-diving echolocating odontocetes.

ACKNOWLEDGMENTS

This research was supported in part by the National Science Foundation (NSF) under CAREER Award No. N2146261, the Office of Naval Research (ONR) under YIP Award No. N00014-15-1-2587, and the Cooperative Ecosystems Study Unit under Cooperative Agreement No. N62473-18-2-0016 for the U.S. Navy, U.S. Pacific Fleet. We thank the students and staff of the Scripps Marine Bioacoustics Collaborative for fieldwork, data curation, and data preparation. Finally, we thank Mr. Wenyu Zhang for providing an implementation of the MTT based on particle flow.

NOMENCLATURE

TDOA	Time-difference-of-arrival
MTT	Multi-target tracking
PAM	Passive acoustic monitoring
SNR	Signal-to-noise ratio
GCC	Generalized cross-correlation
GCC-WIN	Generalized cross-correlation for whitening instrument noise
HARP	High-frequency acoustic recording package
DOA	Direction-of-arrival
MHT	Multiple hypothesis tracker
PHD	Probabilistic hypothesis density
GM-PHD	Gaussian mixture probabilistic hypothesis density
SPA	Sum-product algorithm
PDF	Probability density function
PSD	Power spectral density
CPSD	Cross-power spectral density
SCOT	Smoothed coherence transform
PHAT	Phase transform
MMSE	Minimum mean square error
MOU	Measurement-origin uncertainty
PT	Potential target
STD	Standard deviation
NST	Nonsequential tracking
SBT	Single Bernoulli tracker
LTSA	Long-term spectral average
FFT	Fast Fourier transform
ADCP	Acoustic Doppler current profiler
ICI	Inter-click-interval

[1]T. A. Marques, L. Thomas, S. W. Martin, D. K. Mellinger, J. A. Ward, D. J. Moretti, D. Harris, and P. L. Tyack, "Estimating animal population density using passive acoustics," Biol. Rev. **88**(2), 287–309 (2013).
[2]J. A. Hildebrand, S. Baumann-Pickering, K. E. Frasier, J. S. Trickey, K. P. Merkens, S. M. Wiggins, M. A. McDonald, L. P. Garrison, D. Harris, T. A. Marques, and L. Thomas, "Passive acoustic monitoring of beaked whale densities in the Gulf of Mexico," Sci. Rep. **5**, 16343 (2015).
[3]M. Gassmann, S. M. Wiggins, and J. A. Hildebrand, "Three-dimensional tracking of Cuvier's beaked whales' echolocation sounds using nested hydrophone arrays," J. Acoust. Soc. Am. **138**(4), 2483–2494 (2015).
[4]A. M. Thode, S. B. Blackwell, A. S. Conrad, K. H. Kim, T. Marques, L. Thomas, C. S. Oedekoven, D. Harris, and K. Bröker, "Roaring and repetition: How bowhead whales adjust their call density and source level (Lombard effect) in the presence of natural and seismic airgun survey noise," J. Acoust. Soc. Am. **147**(3), 2061–2080 (2020).
[5]A. Krumpel, A. Rice, K. E. Frasier, F. Reese, J. S. Trickey, A. E. Simonis, J. P. Ryan, S. M. Wiggins, A. Denzinger, H.-U. Schnitzler, and S. Baumann-Pickering, "Long-term patterns of noise from underwater explosions and their relation to fisheries in Southern California," Front. Mar. Sci. **8**, 796849 (2021).
[6]E. A. Falcone, G. S. Schorr, S. L. Watwood, S. L. DeRuiter, A. N. Zerbini, R. D. Andrews, R. P. Morrissey, and D. J. Moretti, "Diving behaviour of Cuvier's beaked whales exposed to two types of military sonar," R. Soc. Open Sci. **4**(8), 170629 (2017).
[7]R. T. Buxton, M. F. McKenna, D. Mennitt, K. Fristrup, K. Crooks, L. Angeloni, and G. Wittemyer, "Noise pollution is pervasive in U.S. protected areas," Science **356**(6337), 531–533 (2017).
[8]C. Erbe, S. A. Marley, R. P. Schoeman, J. N. Smith, L. E. Trigg, and C. B. Embling, "The effects of ship noise on marine mammals—A review," Front. Mar. Sci. **6**, 606 (2019).
[9]E. Pirotta, C. G. Booth, D. P. Costa, E. Fleishman, S. D. Kraus, D. Lusseau, D. Moretti, L. F. New, R. S. Schick, L. K. Schwarz, S. E. Simmons, L. Thomas, P. L. Tyack, M. J. Weise, R. S. Wells, and J. Harwood, "Understanding the population consequences of disturbance," Ecol. Evol. **8**(19), 9934–9946 (2018).
[10]E. L. Hazen, B. Abrahms, S. Brodie, G. Carroll, M. G. Jacox, M. S. Savoca, K. L. Scales, W. J. Sydeman, and S. J. Bograd, "Marine top predators as climate and ecosystem sentinels," Front. Ecol. Environ. **17**(10), 565–574 (2019).
[11]W. M. X. Zimmer, *Passive Acoustic Monitoring of Cetaceans*, 1st ed. (Cambridge University Press, Cambridge, UK, 2011).
[12]W. W. Au and M. O. Lammers, *Cetacean Acoustics* (Springer, New York, 2007), pp. 843–875.

[13]C. W. Clark, *Acoustic Behavior of Mysticete Whales* (Springer US, Boston, MA, 1990), pp. 571–583.
[14]S. Baumann-Pickering, M. A. McDonald, A. E. Simonis, A. Solsona Berga, K. P. B. Merkens, E. M. Oleson, M. A. Roch, S. M. Wiggins, S. Rankin, T. M. Yack, and J. A. Hildebrand, "Species-specific beaked whale echolocation signals," J. Acoust. Soc. Am. **134**(3), 2293–2301 (2013).
[15]P. Gruden, E.-M. Nosal, and E. Oleson, "Tracking time differences of arrivals of multiple sound sources in the presence of clutter and missed detections," J. Acoust. Soc. Am. **150**(5), 3399–3416 (2021).
[16]L. Bouffaut, K. Taweesintananon, H. Kriesell, R. Rørstadbotnen, J. Potter, M. Landro, S. E. Johansen, J. Brenne, A. Haukanes, O. Schjelderup, and F. Storvik, "Eavesdropping at the speed of light: Distributed acoustic sensing of baleen whales in the Arctic," Front. Mar. Sci. **9**, 901348 (2022).
[17]J. Barlow, E. T. Griffiths, H. Klinck, and D. V. Harris, "Diving behavior of Cuvier's beaked whales inferred from three-dimensional acoustic localization and tracking using a nested array of drifting hydrophone recorders," J. Acoust. Soc. Am. **144**(4), 2030–2041 (2018).
[18]S. Fregosi, D. V. Harris, H. Matsumoto, D. K. Mellinger, J. Barlow, S. Baumann-Pickering, and H. Klinck, "Detections of whale vocalizations by simultaneously deployed bottom-moored and deep-water mobile autonomous hydrophones," Front. Mar. Sci. **7**, 721 (2020).
[19]E. Henderson, S. Martin, R. Manzano-Roth, and B. Matsuyama, "Occurrence and habitat use of foraging Blainville's beaked whales (*Mesoplodon densirostris*) on a U.S. navy range in Hawaii," Aquat. Mamm. **42**, 549–562 (2016).
[20]R. Gibb, E. Browning, P. Glover-Kapfer, and K. E. Jones, "Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring," Methods Ecol. Evol. **10**(2), 169–185 (2019).
[21]L. Tenorio-Hallé, A. M. Thode, M. O. Lammers, A. S. Conrad, and K. H. Kim, "Multi-target 2D tracking method for singing humpback whales using vector sensors," J. Acoust. Soc. Am. **151**(1), 126–137 (2022).
[22]Y. M. Barkley, E.-M. Nosal, and E. M. Oleson, "Model-based localization of deep-diving cetaceans using towed line array acoustic data," J. Acoust. Soc. Am. **150**(2), 1120–1132 (2021).
[23]A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 2nd ed. (McGraw-Hill, Inc., New York, 1984).
[24]C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," IEEE Trans. Acoust., Speech, Signal Process. **24**(4), 320–327 (1976).
[25]S. M. Wiggins, B. J. Thayre, J. S. Trickey, S. Baumann-Pickering, and J. A. Hildebrand, "MPL technical memorandum 631: Beaked whale passive acoustic tracking offshore of Cape Hatteras 2017," N62470-15-D-8006, MPL, La Jolla, CA (2018).
[26]P. Gruden and P. R. White, "Automated extraction of dolphin whistles—A sequential Monte Carlo probability hypothesis density approach," J. Acoust. Soc. Am. **148**(5), 3014–3026 (2020).
[27]Y. Bar-Shalom, P. K. Willett, and X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms* (Yaakov Bar-Shalom, Storrs, CT, 2011).
[28]D. B. Reid, "An algorithm for tracking multiple targets," IEEE Trans. Autom. Control **24**(6), 843–854 (1979).
[29]R. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," IEEE Trans. Aerosp. Electron. Syst. **39**(4), 1152–1178 (2003).
[30]R. Mahler, *Statistical Multisource-Multitarget Information Fusion* (Artech House, Norwood, MA, 2007).
[31]A. A. Saucan, M. Coates, and M. Rabbat, "Multi-sensor multi-Bernoulli filter," IEEE Trans. Signal Process. **65**(20), 5495–5509 (2017).
[32]F. Meyer, T. Kropfreiter, J. L. Williams, R. Lau, F. Hlawatsch, P. Braca, and M. Z. Win, "Message passing algorithms for scalable multitarget tracking," Proc. IEEE **106**(2), 221–259 (2018).
[33]F. Meyer, P. Braca, P. Willett, and F. Hlawatsch, "A scalable algorithm for tracking an unknown number of targets using multiple sensors," IEEE Trans. Signal Process **65**(13), 3478–3493 (2017).
[34]F. Meyer, A. Tesei, and M. Z. Win, "Localization of multiple sources using time-difference arrival measurements," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* (2017), pp. 3151–3155.
[35]A. Tesei, F. Meyer, and R. Been, "Tracking of multiple surface vessels based on passive acoustic underwater arrays," J. Acoust. Soc. Am. **147**(2), EL87–EL92 (2020).
[36]P. M. Baggenstoss, "A multi-hypothesis tracker for clicking whales," J. Acoust. Soc. Am. **137**(5), 2552–2562 (2015).
[37]P. M. Baggenstoss, "Separation of sperm whale click-trains for multipath rejection," J. Acoust. Soc. Am. **129**(6), 3598–3609 (2011).
[38]T. Kropfreiter, F. Meyer, S. Coraluppi, C. Carthel, R. Mendrzik, and P. Willett, "Track coalescence and repulsion: MHT, JPDA, and BP," in *Proc. Int. Conf. Inf. Fusion* (2021), pp. 1–8.
[39]B. N. Vo and W. K. Ma, "The Gaussian mixture probability hypothesis density filter," IEEE Trans. Signal Process. **54**(11), 4091–4104 (2006).
[40]C. O. Tiemann, M. B. Porter, and L. N. Frazer, "Localization of marine mammals near Hawaii using an acoustic propagation model," J. Acoust. Soc. Am. **115**(6), 2834–2843 (2004).
[41]E.-M. Nosal, "Methods for tracking multiple marine mammals with wide-baseline passive acoustic arrays," J. Acoust. Soc. Am. **134**(3), 2383–2392 (2013).
[42]J. Macaulay, J. Gordon, D. Gillespie, C. Malinka, and S. Northridge, "Passive acoustic methods for fine-scale tracking of harbour porpoises in tidal rapids," J. Acoust. Soc. Am. **141**(2), 1120–1132 (2017).
[43]J. A. Nelder and R. Mead, "A simplex method for function minimization," Comput. J. **7**, 308–313 (1965).
[44]N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," J. Chem. Phys. **21**, 1087–1092 (1953).
[45]H. W. Kuhn, "The Hungarian method for the assignment problem," Nav. Res. Logist. Q. **2**(1-2), 83–97 (1955).
[46]E. Snyder, S. M. Wiggins, S. Baumann-Pickering, and J. A. Hildebrand, "Cuvier's beaked whale tracks in Southern California," J. Acoust. Soc. Am. **146**(4), 2939 (2019).
[47]S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-6**, 721–741 (1984).
[48]F. Meyer and J. L. Williams, "Scalable detection and tracking of geometric extended objects," IEEE Trans. Signal Process. **69**, 6283–6298 (2021).
[49]F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," IEEE Trans. Inf. Theory **47**(2), 498–519 (2001).
[50]F. Daum and J. Huang, "Nonlinear filters with log-homotopy," in *Proc. Soc. Photo-Opt. Instrum. Eng.* (2007), pp. 423–437.
[51]W. Zhang and F. Meyer, "Graph-based multiobject tracking with embedded particle flow," in *Proc. IEEE Radar Conf.* (2021), pp. 1–6.
[52]C. G. Carter, A. H. Nuttall, and P. G. Cable, "The smoothed coherence transform," Proc. IEEE **61**(10), 1497–1498 (1973).
[53]Y. Bar-Shalom, T. Kirubarajan, and X.-R. Li, *Estimation with Applications to Tracking and Navigation* (Wiley, New York, 2002).
[54]M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," IEEE Trans. Signal Process. **50**(2), 174–188 (2002).
[55]S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice-Hall, Upper Saddle River, NJ, 1993).
[56]F. Daum and J. Huang, "Curse of dimensionality and particle filters," in *Proceedings of IEEE Aerospace-03* (2003), Vol. 4, pp. 1979–1993.
[57]Y. Li and M. Coates, "Particle filtering with invertible particle flow," IEEE Trans. Signal Process. **65**(15), 4102–4116 (2017).
[58]B. Ristic, B.-T. Vo, B.-N. Vo, and A. Farina, "A tutorial on Bernoulli filters: Theory, implementation and applications," IEEE Trans. Signal Process. **61**(13), 3406–3430 (2013).
[59]S. M. Wiggins and J. A. Hildebrand, "High-frequency acoustic recording package (HARP) for broad-band, long-term marine mammal monitoring," in *Proc. IEEE Int. Symp. Underw. Technol.* (2007), pp. 551–557.
[60]L. Rabiner, J. McClellan, and T. Parks, "FIR digital filter design techniques using weighted Chebyshev approximation," Proc. IEEE **63**(4), 595–610 (1975).
[61]A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, 1st ed. (Prentice-Hall, Englewood Cliffs, NJ, 1989).
[62]F. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," Proc. IEEE **66**(1), 51–83 (1978).
[63]G. Soldi, F. Meyer, P. Braca, and F. Hlawatsch, "Self-tuning algorithms for multisensor-multitarget tracking using belief propagation," IEEE Trans. Signal Process. **67**(15), 3922–3937 (2019).
[64]H. A. P. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with Markovian switching coefficients," IEEE Trans. Autom. Control **33**(8), 780–783 (1988).
[65]R. C. Connor, J. Mann, P. L. Tyack, and H. Whitehead, "Social evolution in toothed whales," Trends Ecol. Evol. **13**(6), 228–232 (1998).
[66]K. McHugh, "Odontocete social strategies and tactics along and inshore," in *Ethology and Behavioral Ecology of Odontocetes* (Springer, Cham, Switzerland, 2019), pp. 165–182.