Vision-based monitoring and measurement of bottlenose dolphins' daily habitat use and kinematics

Joaquin Gabaldon[1*], Ding Zhang[2], Lisa Lauderdale[3], Lance Miller[3], Matthew Johnson-Roberson[1,4], Kira Barton[1,2], K. Alex Shorter[2],

1 Robotics Institute, University of Michigan, Ann Arbor, MI, USA
2 Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI, USA
3 Chicago Zoological Society, Brookfield Zoo, Brookfield, IL, USA
4 Naval Architecture and Marine Engineering, University of Michigan, Ann Arbor, MI, USA

* gabaldon@umich.edu

Abstract

This research presents a framework to enable computer-automated observation and monitoring of bottlenose dolphins (*Tursiops truncatus*) in a professionally managed environment. Results from this work provide insight into the dolphins' movement patterns, kinematic diversity, and how changes in the environment affect their dynamics. Fixed overhead cameras were used to collect ~ 100 hours of observations, recorded over multiple days including time both during and outside of formal training sessions. Animal locations were estimated using convolutional neural network (CNN) object detectors and Kalman filter post-processing. The resulting animal tracks were used to quantify habitat use and animal dynamics. Additionally, Kolmogorov-Smirnov analyses of the swimming kinematics were used for high-level behavioral mode classification. The detectors achieved a minimum Average Precision of 0.76. Performing detections and post-processing yielded 1.24×10^7 estimated dolphin locations. Animal kinematic diversity was found to be lowest in the morning and peaked immediately before noon. Regions of the habitat displaying the highest activity levels correlated to locations associated with animal care specialists, conspecifics, or enrichment. The work presented here demonstrates that CNN object detection is not only viable for large-scale marine mammal tracking, it also enables automated analyses of dynamics that provide new insight into animal movement and behavior.

Introduction

Direct observation of animals in both free ranging and managed settings has been key to developing an understanding of the behavior and dynamics of these biological systems. How the animals behave in the presence of other animals, interact and engage with their environment, or are affected by changes to their environment are all questions of interest. Ideally, these observations are made without modifying animal behavior, and in a manner that facilitates a quantitative comparison between conditions in the environment. In managed settings there is a strong emphasis on behavioral monitoring to inform welfare practices [1–3]. Bottlenose dolphins, the most common cetacean in zoos and aquariums, are generally regarded as a species that thrives in a managed environment, though data-driven studies of behavior and welfare have been limited [4, 5].

The ability to quantify animal motion and location, both in the environment and with respect to other animals, is therefore critical in understanding their behavior. Here we present an automated computer vision framework inspired by methods found in the field of robotics for persistently and robustly tracking animal position and kinematics.

Biomechanics and behavioral studies depend on animal-based measurements that are considered reliable and repeatable for the species of interest [2, 6–8], but direct measurements of animals in the marine environment can be challenging. As a result, researchers tend to use direct observation and expert knowledge to classify and parameterize animal behavior in both wild and managed settings. In the wild, measurements of animal motion are often made using animal-borne tracking systems. The sensors used to collect data from animals tend to be packaged together into minimally-invasive (removable) tagging systems [9]. These tags can be used to directly measure parameters such as animal speed, acceleration, position at the surface or orientation in their environment without introducing significant modifications to the animals' swimming dynamics [10]. When combined with direct observations of behavior, tag data can be used to quantify the animals' behaviors during a period of interest, such as foraging [11]. Sensor data and behavioral observations have also been used to train algorithms to automatically detect behavioral states [12, 13]. These trained algorithms can then be used to detect and parameterize behavioral states from large amounts of sensor data that lack direct observations of animal behavior.

In contrast, tag-based measurements of marine mammals in managed settings are less common, and location measurements in indoor habitats are not possible with GPS. Instead of tags, animals in these environments tend to be monitored using external sensors, such as cameras and hydrophones, placed in the environment [14, 15]. These sensor networks can be used to observe a majority of the animals' environment with a relatively small number of sensors. While it is possible to continuously record the animals' environmental use and social interactions, these videos must be heavily processed to convert them into useful information. This processing is often performed by a trained expert, who watches and scores behavioral or tracking information from the data [2, 16–18]. This hand-tracking is time consuming and can be inefficient when hundreds of hours of data have been collected from multiple sensors. Recent efforts have been made to automate this process for cameras, primarily through heuristically-crafted computer-vision techniques [19, 20]. However, these techniques were either limited in execution due to prohibitive costs (e.g. funds for the hardware/installation of an extended multi-camera array), or required manual tuning to account for changing environmental conditions (e.g. lighting shifts throughout the day).

To address these gaps, this work investigates day-scale swimming kinematics using a neural network based computer-automated framework to quantify the positional states of multiple animals simultaneously in a managed environment. Neural networks have demonstrated flexibility and robustness in extracting information on biological systems from image and video data [21–23], and were chosen for use in this research for these properties. In this study, video recordings of the animals from a two-camera system were analyzed using convolutional neural network (CNN) object-detection techniques and were post-processed via Kalman filtering to extract animal kinematics. The resulting kinematic states were used to quantify bottlenose dolphin habitat usage, kinematic diversity, and movement profiles during daily life. The framework and results presented here demonstrate the capabilities of robotics/computer vision-inspired techniques in extracting dynamic information from biological systems that can be used to gain new insights into behaviors and biomechanics.

Materials and methods

In this work, camera data were used to monitor the behavior of a group of marine mammals both qualitatively and quantitatively in a managed setting. Camera-based animal position data were used to quantify habitat usage, as well as where and how the group of animals moved throughout the day. The position data were decomposed into kinematics metrics, which were used to discriminate between two general movement states — static and dynamic — using the velocities of the tracked animals. A general ethogram of the animals' behaviors monitored in this research is presented in Table 1. The kinematics metrics were further used to refine our understanding of the behavioral states the animals experienced both in and out of training sessions through a combination of Kolmogorov-Smirnov statistical analyses and joint differential entropy computations. The study protocol was approved by the University of Michigan Institutional Animal Care and Use Committee and the Brookfield Zoo.

Table 1. Behavior condition ethogram of dolphins under professional care.

Category	Behavior	Definition
ITS (In Training Session)	Animal Care Session	Time period in which animal care specialists work with the dolphins to learn new behaviors or practice known behaviors without public audience.
ITS	Formal Presentation	Time period in which animal care specialists work with the dolphins in front of an audience to present educational information to the public.
OTS (Out of Training Session)	Static	Animal movement state with little to no active fluking at a rate of speed less than $0.5\,\mathrm{m\,s^{-1}}$.
OTS	Dynamic	Animal movement state with active fluking at a rate of speed greater than $0.5\,\mathrm{m\,s^{-1}}$.

Experimental environment

Seven bottlenose dolphins with an average age of 17 ± 12 yrs and length of 247 ± 17 cm were observed using a dual-camera system in the Seven Seas building of the Brookfield Zoo, Brookfield IL. The complete environment consists of a main habitat with public viewing, two smaller habitats behind the main area, and a medical habitat (not shown) between the two smaller habitats (Fig. 1). The main habitat is 33.5 m across, 12.2 m wide, and 6.7 m deep. The back habitats have circular diameters of 10.7 m and are 4.3 m deep, and the medical area is 7.6 m in diameter and 2.4 m deep. The habitats are connected through a series of gates. During formal training sessions in the main habitat, animal care specialists primarily engage with the animals on the island between the gates to the other areas. There are underwater observation windows for the viewing public on the far side of the main habitat from the island (not shown), and smaller windows looking into the offices of the animal care specialists on the island and next to the right gate (Fig. 2, top). Recordings of the main habitat took place across multiple days (between Feb. 6 and March 27, 2018), for varying portions of each day, for a total of 99.5 hours over 20 recordings. Data collection began at the earliest at 07:41 and ended at the latest at 16:21. During the recorded hours, the dolphins

participated in four formal training sessions according to a regular, well-defined schedule set by the animal care specialists (ACSs).

Fig 1. Diagram of the experimental setup. TOP: Illustration of the main habitat, with camera placements (blue enclosures) and fields of view (gray cones). BOTTOM: x-y view of example tracklets (red and green on gray lines) of two dolphins (highlighted light orange), which are also visible in the top of this figure. BOTTOM-ZOOM (RIGHT): Vector illustrations of the two example tracks. Example notation for tracklet j (red): position ($\vec{p}^{(j,t')}$), velocity ($\vec{v}^{(j,t')}$), yaw ($\theta^{(j,t')}$), and yaw rate ($\dot{\theta}^{(j,t')}$). BOTTOM-ZOOM (LEFT) Illustration of tracklet generation, with detections (stars) and tracklet proximity regions (dashed). Example notation for tracklet j (red): position ($\vec{p}^{(j,t)}$), velocity ($\vec{v}^{(j,t)}$), Kalman-predicted future position ($\hat{\vec{p}}^{(j,t+1)}$), true future position ($\vec{p}^{(j,t+1)}$), and future animal detection ($\vec{u}^{(j,t+1,i')}$).

Fig 2. Combined figure demonstrating camera overlap, bounding box meshing, and animal position uncertainty. TOP: Transformed individual camera views, with objects in the habitat marked. Yellow – Dolphin bounding boxes, Green – Drains, Red – Gates between regions, Orange – Underwater windows (3 total). Correlated bounding boxes are indicated by number, and the habitat-bisecting lines (l_s) for each camera frame in solid red. Distances from Box 2 to the closest frame boundary (d_b) and the boundary to the bisecting line (d_l) are highlighted in yellow. MIDDLE: Combined camera views including dolphin bounding boxes (yellow), with the location uncertainty distribution (A) overlaid for Box 2. BOTTOM: 2D uncertainty distribution (A) with major (a-a, black) and minor (b-b, red) axes labeled and separately plotted.

A formal training session consisted of time in which the ACSs work with the dolphins to learn new behaviors or practice known behaviors. At the beginning of each formal training session, the dolphins were asked to maintain positions directly in front of the ACS (formally known as "stationing"). The animal care specialists then presented discriminative stimuli or gestures that indicated which behaviors they requested each dolphin produce. If a dolphin produced the desired behavior, they received a reward (i.e., reinforcement). If they chose not to produce the behavior, a specialist may request the same behavior again or move on to a different behavior. When the animals were in a formal training session (abbreviated ITS), they experienced two formats of training during the data collection period: non-public animal care sessions and formal public presentations. Time outside of formal training sessions (abbreviated OTS) was defined as when the animals were not interacting with ACSs. During the OTS time periods, the ACSs would provide enrichment objects for the animals to interact with and select which parts of the habitat the animals could access using gates on either side of the main island. The time intervals for the OTS and ITS blocks are displayed in Table 2.

Experimental equipment

Two AlliedVision Prosilica GC1380C camera sensors with Thorlabs MVL5M23 lenses were separately mounted in Dotworkz D2 camera enclosures, which were attached to 80/20 T-slotted aluminum framing. On the frame, the cameras were spaced approximately 2m apart. The frame was mounted to a support beam directly above the main habitat, with the cameras angled to give full coverage of the area when combined. Figure 1, top, illustrates the habitat, camera placement, and field of view coverage. For data collection, the cameras were connected through the Gigabit Ethernet protocol to a central computer with an Intel i7-7700K CPU. Recordings were executed using the MATLAB Image Acquisition Toolbox, in the RGB24 color format at a frame rate of

Table 2. Block time intervals

	Time Interval	
Block	OTS	ITS
1	08:00 – 09:30	09:30 – 10:00
2	10:00 – 11:30	11:30 – 12:00
3	12:00 – 13:00	13:00 – 13:30
4	13:30 – 14:30	14:30 – 15:00
5	15:00 – 16:00	N/A

The ITS blocks (1 and 3) are animal care sessions, and the OTS blocks (2 and 4) are formal presentations.

20Hz. Each camera was connected to a separate Ethernet port on an internal Intel PRO/1000 Pt PCIe card. A separate computer system was used for detection inference, and was outfitted with an Intel i7-8700K processor clocked to 4.8 GHz and a Nvidia Titan V graphics processing unit in Tesla Compute Cluster mode.

Dolphin detection

Approximately 99.5 hours of data from two cameras were collected for this work, resulting in ~ 14 million individual frames of data. To extract spatial information about habitat use and swimming kinematics, we first needed to identify animals in the frames. These detections were filtered and associated with short trajectories (tracklets) from individual animals. Kinematic data (position, velocity and heading) from the tracklets were then used to parameterize and form probability distributions that were used to identify tendencies in animal motion during in training (ITS) and out of training session (OTS) swimming.

Neural network methods

The first step in the analysis process was dolphin detection from the captured video frames using Faster R-CNN, a machine-learning object detection method [24]. The method consisted of two primary modules: a Region Proposal Network (RPN), and a Fast R-CNN detector network. The RPN identified regions in an image that may enclose objects of interest, and presented these to the Fast R-CNN detector to verify which regions did in fact contain objects the detector sought to identify. These two modules when combined form one large network capable of returning a bounding box tightly enclosing an object's location within an image. For a more complete explanation of the method please refer to [24].

All modules used in the implementation were present in the MATLAB Deep Learning Toolbox excepting the Parametric Rectified Linear Unit (PReLU) activation function, which was defined with a custom neural network layer per directions in the MATLAB online documentation [25, 26]. The convolutional neural network (CNN) structure used in the Faster R-CNN framework is as follows. For the input layer, the size was chosen to be similar to the smallest bounding boxes in the set of manually scored dolphin profiles, in the format of (l,l,3), where l is $2\times$ the side length of the smallest bounding box major axis. The input layer had a third dimension of 3 as input images were in the RGB colorspace. The feature extraction layers had the following structure: four sets of 2D 3×3 convolution layers, each followed by batch normalization, PReLU activation, and 2×2 max pooling (stride 2) layers, in that order. The four convolution layers had, in order: 64, 96, 128, and 128 filters. Each convolution was performed with one layer of zero padding along the edges of the inputs to avoid discounting the

corners/edges. The classification layers used the extracted features from the previous layers to identify an image region as either a dolphin or the background. They consisted of: 1) A fully connected layer, length 512, to extract features from the final convolution layer, followed by a PReLU activation; 2) A fully connected layer, length 2, to determine non-scaled classification weights; 3) A softmax function layer to convert these weights into the final probabilities of the image region's classification. The highest probability from the softmax layer corresponded to the most likely classification for the region, and the magnitude of this probability indicated the confidence of the classification.

Training the network

Ground truth data were scored by a trained observer who manually defined bounding boxes that identified the locations of the dolphins in the training/testing frames (Fig. 2, A). These ground truth data were selected over a range of lighting conditions and dolphin locations to ensure robustness of the detection network. For each camera, 100 frames were extracted from each of 11 separate recordings, with evenly spaced time intervals between frames. The recordings were collected in May 2017, and February, March, and August 2018. Over 940 frames from each of the left and right cameras were found to contain usable dolphin locations, i.e. human-detectable dolphin profiles. Each usable dolphin location in the selected frames was manually given a bounding box tightly enclosing the visible profile. The detector for the left camera was trained on 1564 profiles and tested on 662, and the detector for the right camera was trained on 1482 profiles and tested on 662. The dolphin detectors were trained using the MATLAB implementation of Faster R-CNN, employing the previously-defined CNN structure as the classification method.

Detection processing

Detections were performed over all 99.5 hours of recorded data from both cameras, at 10Hz intervals (total of 7.16×10^6 frames), using a 95% minimum confidence threshold to ensure accuracy. The fields of view of the two cameras overlap for a portion of the habitat, resulting in some dolphins being detected simultaneously by both cameras. This yielded multiple sets of conflicting detection bounding boxes spanning the two fields of view, which necessitated associating the most likely left/right box pairs. Before conflict identification was performed, the detection boxes were first transformed into a common plane of reference termed the world frame. Using known world point coordinates, homographies from each camera to the world frame were generated using the normalized Direct Linear Transform method [27]. These homographies were used to convert the vertices of the bounding boxes to the world frame using a perspective transformation. Intersecting boxes were identified by evaluating polygonal intersections, and Intersection over Union (IoU) metrics were computed for intersecting boxes to measure how well they matched. Associations were identified between pairs of left/right intersecting boxes with the highest mutual IoU values.

Associated boxes' world frame centroid locations were meshed using a weighted mean. First, the boundaries of each camera's field of view were projected into the world frame, allowing us to obtain the line in the world frame y-direction defining the center of the overlap region, denoted $l_s = x_{mid}$ (Fig. 2, middle). x_{mid} is the x-coordinate in the world frame midway between the physical placement of the cameras. For each detection (u), the distance (d_b) in the x-direction from u to the nearest projected camera boundary line (b_n) was then determined. Next, the distance (d_l) in the x-direction from line l_s through u to b_n was found. Finally, the weight for the camera corresponding to b_n was calculated as $w_n = d_b/2d_l$, with the weight for the other (far)

camera as $w_f = 1 - w_n$. This ensured that if detection u was on l_s, then $w_n = w_f = 0.5$, and as u moved closer to b_n, we would have $w_n \to 0$ and $w_f \to 1$.

In specific circumstances, the shapes of the drains at the bottom of the habitat were warped by the light passing through rough surface water, and resulted in false dolphin detections. Separate (smaller) image classifiers for each camera were trained to identify these false positive drain detections, and were run on any detections that occurred in the regions of the video frames containing the drains. These detectors were strictly CNN image classifiers and were each trained on over 350 images and tested on over 150 images. For the drain detector, the input layer size had the format of $(l_d, l_d, 3)$, where l_d is the mean side length of the detection bounding boxes being passed through the secondary classifiers. The feature detection layers had the same general structure as the Faster R-CNN classifier network, except in this case the convolution layers had, in order: 32, 48, 64, and 64 filters each. In the classification layers, the first fully connected layer had a length of 256.

Tracklet formation

Each experimental session involved the detection of multiple animals throughout their habitat. However, animal detections were done independently for each frame of the video. To extract kinematic information from the animals in the video, the detection associations needed to be preserved across frames. In this work, short continuous tracks (i.e. *tracklets*) were generated for a detected animal by identifying the most likely detection of that animal in the subsequent frame. To generate multiple individual tracklets in series of video frames, an iterative procedure of *prediction* and *association* was conducted under a Kalman filter framework with a constant velocity model.

The position of the i-th detected animal in one video frame at time t is denoted as $\mathbf{u}^{(t,i)} = [u_x^{(t,i)}, u_y^{(t,i)}]$. Each detection, $\mathbf{u}^{(t,i)}$ was either associated with a currently existing tracklet or used to initialize a new tracklet. To determine which action was taken, for each tracklet, denoted as $\mathbf{T}^{(k)}$ for the k-th tracklet, this process first predicted the state of the tracked animal in the next frame ($\widehat{\mathbf{T}}^{(k,t+1)}$) based on the current state information of the animal $\mathbf{T}^{(k,t)}$.

$$\mathbf{T}^{(k,t)} = [\mathbf{p}^{(k,t)}, \mathbf{v}^{(k,t)}] \tag{1}$$
$$= [p_x^{(k,t)}, p_y^{(k,t)}, v_x^{(k,t)}, v_y^{(k,t)}] \tag{2}$$
$$\widehat{\mathbf{T}}^{(k,t+1)} = [\hat{\mathbf{p}}^{(k,t+1)}, \hat{\mathbf{v}}^{(k,t+1)}] \tag{3}$$
$$= [\hat{p}_x^{(k,t+1)}, \hat{p}_y^{(k,t+1)}, \hat{v}_x^{(k,t+1)}, \hat{v}_y^{(k,t+1)}] \tag{4}$$

where $\mathbf{p}^{(k,t)} = [p_x^{(k,t)}, p_y^{(k,t)}]$ denotes the filtered position of the animal tracked by the k-th tracklet at time t and $\mathbf{v}^{(k,t)} = [v_x^{(k,t)}, v_y^{(k,t)}]$ is the corresponding velocity. Under a constant velocity model, the predicted next frame position $\hat{\mathbf{p}}^{(k,t+1)} = [\hat{p}_x^{(k,t+1)}, \hat{p}_y^{(k,t+1)}]$ was obtained by integrating the current velocity over one frame period and summing this to the current frame position. The predicted velocity remained constant.

$$\hat{p}_x^{(k,t+1)} = p_x^{(k,t)} + v_x^{(k,t)}\Delta t \tag{5}$$
$$\hat{p}_y^{(k,t+1)} = p_y^{(k,t)} + v_y^{(k,t)}\Delta t \tag{6}$$
$$\hat{v}_x^{(k,t+1)} = v_x^{(k,t)} \tag{7}$$
$$\hat{v}_y^{(k,t+1)} = v_y^{(k,t)} \tag{8}$$

Using the predicted position, the k-th tracklet checked whether there existed a closest detection in the next frame that was within the *proximity region* of the predicted position. If true, that detection, denoted as $\mathbf{u}^{(k,t+1,i)}$ for the i-th detection in frame $t+1$ associated with the k-th tracklet, was used as the reference signal of the Kalman filter to update the state (position and speed) of tracklet $\vec{T}^{(k)}$. If false, the unassociated tracklet continued propagating forward, assuming the animal maintained a constant velocity. If a tracklet continued to be unassociated for 5 consecutive frames (empirically determined), it was considered inactive and was truncated at the last confirmed association. All information related to the k-th tracklet was saved after its deactivation:

$$\mathbf{T}^{(k)} = [\mathbf{T}^{(k,t_{start})}, \ldots, \mathbf{T}^{(k,t-1)}, \mathbf{T}^{(k,t)}, \mathbf{T}^{(k,t+1)}, \ldots, \mathbf{T}^{(k,t_{end})}]^T \tag{9}$$

As illustrated in Fig. 1, the tracklet formation operation linked each animal's individual detections ($\mathbf{u}$) over consecutive frames and returned not only the positions ($\mathbf{p}$) of the animals, but also the forward speed (v), yaw (heading, θ), and turning rate ($\dot{\theta}$), which could then be used to parameterize the positional states of the animals.

Position uncertainty

There was a general position uncertainty for each animal detection due to noise in the Faster R-CNN detections. This was caused by a combination of limited camera resolution, as well as distortion of an animal's image from waves and ripples on the surface of the water. Additionally, since animal depth could not be measured, there were errors in the world-frame x-y location estimates (caused by camera perspective and light refraction effects) that could not be corrected. In this work, the detection uncertainty was represented as a 2D probability density function (PDF), whose size and shape depended on the location of the detection with respect to the cameras (Fig. 2, bottom). The short (minor) axis, D_1, was a Gaussian uncertainty distribution defined according to a heuristically estimated error in the camera detections (~ 0.2 m), and represented the general position uncertainty in the Faster R-CNN detections (Fig. 2, bottom, b-b). The long (major) axis of the distribution, D_2, represented the position uncertainty caused by the perspective and refraction effects (uncertainty from unknown depth). A 1D PDF was defined according to previously measured animal depth data (total of 9.8 hours during separate OTS time blocks), obtained via non-invasive tagging, which represented the general distribution of depths occupied by the animals. This was convolved with D_1 to produce the general shape of D_2 (Fig. 2, bottom, a-a). The x-axis length scale for D_2 for a particular detection was obtained from the maximum position error in the detection's x-y location. This was the magnitude of the x-y position difference (original versus corrected x-y position) if the detection happened to be at maximum depth (~ 7 m). This magnitude varied dependent on the world-frame original location of the detection. Details on the depth-based location correction can be found in [28].

Mapping animal kinematics to habitat

Heatmaps of dolphin position and speed were used to map animal positional state to the habitat. The dolphins were defined to be static or minimally mobile (drifting) when they were traveling at speeds below $0.5\,\mathrm{m\,s^{-1}}$, and dynamic otherwise. To generate the positional heat maps, a blank 2D pixel map of the main habitat, M, was first created. Then, for each pixel representation p of a detection u, the maximum possible magnitude of location error due to depth was determined, defined as e_m (pixels, scale 1 pix = 5 cm), along with the orientation of the error propagation, ψ_m (radians). The perimeter of the habitat served as a hard constraint on the location of the animals, thus e_m was

truncated if the location of the point with the maximum possible shift, $[p_x + e_m \cos(\psi_m), p_y + e_m \sin(\psi_m)]$, fell outside this boundary. The minor axis of the 2D uncertainty distribution, D_1, was a 1D PDF in the form of a Gaussian kernel with $\sigma_{gauss} = 0.2s$ (0.2 meters scaled to pixels by scaling factor $s = 20$). Next, the depth PDF was interpolated to be e_m pixels long, and was convolved with D_1 (to account for measurement uncertainty in the camera detections). This yielded the major axis 1D PDF, D_2. The 2D (unrotated) occupancy PDF, $E = D_1^\top D_2$, was then computed, where D_1, D_2 were horizontal vectors of the same length. The 2D rotated occupancy PDF, F, was calculated by rotating E by an angle of ψ_m through an interpolating array rotation. The MATLAB implementation of `imrotate` was used for this calculation. F was then normalized to ensure the distribution summed to 1. Finally, F was locally summed into M, centered at location $[x_u, y_u] = [p_x + 0.5e_m \cos(\psi_m), p_y + 0.5e_m \sin(\psi_m)]$, to inject the occupancy probability distribution for u into map M. This process was then repeated for all detections. For the sake of visibility, all heatmaps were sub-sampled down to the scale of 1 pix = 1 meter.

A similar process was used to form the speed heatmaps. In a speed heatmap, the values of F are additionally scaled by the scalar speed of the animal, v, that corresponds to detection u, and then locally summed into a separate map, N (sum $F \cdot v$ into N centered at $[x_u, y_u]$). Element-wise division of N by M was performed to generate S, a map of the average speed per location.

Lastly, the direction of motion of the animals throughout the monitored region was described using a quiver plot representation. To formulate the quiver plot, two separate heatmaps were generated, Q_x and Q_y, one each for the x and y components of the animals' velocities. Q_x was created using a similar method to the speed heatmap, but in this case F was scaled by the x-component of the animal's velocity (sum $F \cdot v\cos(\theta)$ into Q_x centered at $[x_u, y_u]$), where θ was the heading of the animal corresponding to detection u. Similarly for Q_y, F was scaled by the y-component of the animal's velocity (sum $F \cdot v\sin(\theta)$ into Q_y centered at $[x_u, y_u]$). The vector components Q_x and Q_y combined represented the general orientation of the animals at each point in the habitat.

Probability distributions of metrics and entropy computation

For each time block of OTS and ITS, the PDFs of speed ($\mathrm{m\,s^{-1}}$) and yaw (rad) were numerically determined. These were obtained by randomly extracting 10^5 data samples of both metrics from each time block of OTS and ITS, and producing PDFs for each metric and time block from these data subsets.

Additionally, the joint differential entropies of speed and yaw were computed for each time block of OTS and ITS. In this case, the joint entropy of animal speed and yaw represents the coupled variation in these metrics for the animals. This indicates that speed-yaw joint entropy can be considered a proxy for measuring the diversity of their kinematic behavior. To compute the joint entropy h for one time block, the randomly sampled speed (continuous random variable $\mathbf{S}$) and yaw (continuous random variable $\mathbf{\Psi}$) data subsets (S and Ψ, respectively) of that time block were used to generate a speed/yaw joint PDF: $f(s, \psi)$, where $s \in S$, $\psi \in \Psi$. f was then used to compute h with the standard method:

$$h(\mathbf{S}, \mathbf{\Psi}) = -\int_{S,\Psi} f(s, \psi) \ln[f(s, \psi)] ds d\psi \tag{10}$$

Kolmogorov-Smirnov statistics

To evaluate the statistical differences in animal dynamics between time blocks, the two-sample Kolmogorov-Smirnov (K-S) distances (Δ_{ks}) and their significance levels (α)

were computed for each of the following metrics: speed ($\mathrm{m\,s^{-1}}$), yaw (rad), yaw rate ($\mathrm{rad\,s^{-1}}$), and the standard deviations of each [29]. These were done by comparing randomly-sampled subsets of each time block, with each subset consisting of 10^4 data samples per metric. Only time blocks of similar type were compared (i.e. no ITS blocks were compared to OTS blocks, and vice-versa). The computations were performed using the MATLAB statistics toolbox function `kstest2`.

Results

Detector and filter performance

During evaluation, the Faster R-CNN detectors for the left and right cameras achieved Average Precision scores of 0.76 and 0.78, respectively. The CNN drain classifiers for the left and right cameras achieved respective accuracy scores of 92% and 94%. Processing all 99.5 hours of recordings yielded 5.92×10^6 detections for the left camera and 6.35×10^6 detections for the right. The initial set of detections took ~ 8.4 days to compute when performed on the Titan V computer system. Of these, 3.83×10^4 (0.65%) detections from the left camera and 3.02×10^4 (0.48%) detections from the right camera were found to be drains misclassified as dolphins. After removing the misclassified detections, meshing the left and right detection sets yielded a total of 1.01×10^7 individual animal detections within the monitored habitat. The tracklet generation method used in this work associated animal track segments containing gaps of up to 4 time steps. As a result, the prediction component of its Kalman filter implementation was used to fill in short gaps in the tracking data. Generating tracklets from the meshed detections yielded a total of 1.24×10^7 estimated dolphin locations, from 3.44×10^5 total tracklets.

Spatial distribution — position

During OTS, the tracked animals were found to be in a dynamic swimming state $\sim 77\%$ of the time and a static state for $\sim 23\%$ of the time. The static OTS behavior tended to be associated with particular features of their habitat: the gates that lead to the other areas of the habitat or at the underwater windows that offered views of the animal care specialist staff areas (Fig. 3). When swimming dynamically during OTS, the dolphins tended to spend more time near the edges of their habitat, with the most time focused on the island side with the gates and the windows (Fig. 4, left column). This was especially true during Block 5, with additional weight placed along the edge of the central island.

Fig 3. Static position distributions for OTS and ITS. A note on the format of the training sessions: Dolphins spent more time stationed at the main island during public presentations than non-public animal care sessions. During formal public presentations, ACSs spend a higher portion of the training session on the main island because it is within view of all of the public attending the presentation. Non-public animal care sessions are more fluid in their structure than public sessions. ACSs often use the entire perimeter of the habitat throughout the session.

Throughout ITS, the dolphins were asked to engage in dynamic swimming tasks $\sim 62\%$ of the time, and were at station (in front of the ACSs) for the remaining $\sim 38\%$ of the time. During ITS, the dolphins had a heavy static presence in front of the central island, where the animals were stationed during formal training programs. Less emphasis was placed on the edges, contrasted to their locations during OTS (Fig. 5, left

Fig 4. Spatial distributions for dynamic OTS, with position distributions along the first column and speed distributions/quiver plots along the second column. Prior to the first full training session of the day at 9:30 a.m., the dolphins were engaged in low intensity (resting) swimming clockwise around the perimeter of the habitat, with the highest average OTS speeds recorded after the 9:30 sessions. From there, speeds trail off for the subsequent two time periods. The 1:30-2:30 p.m. time block is characterized by slower swimming in a predominantly counterclockwise pattern. There is an increase in speed and varied heading pattern during the 3:00-4:00 time block.

column). During ITS, the ACSs presented discriminative stimuli or gestures corresponding to specific animal behavior, which defined the spatial distributions of the dolphins' movements during these time blocks. Additionally, there were spatial distribution similarities between training sessions of similar type, e.g. Blocks 1, 3 were animal care and husbandry sessions, and 2, 4 were formal public presentations. Note the structure of the spatial distributions across the top of their habitat, where during the care sessions (Blk. 1, 3) the dolphins' positions were focused on specific points in the area, while during the presentations (Blk. 2, 4) their positions were distributed across the edge of the central island. This captured the formation used during presentations with animals distributed more uniformly across the island.

Fig 5. Spatial distributions for dynamic ITS, with position distributions along the first column and speed distributions/quiver plots along the second column. Speeds across the entire habitat are higher during public presentations than non-public animal care sessions because high-energy behaviors (e.g., speed swims, porpoising, and breaches) are typically requested from the group several times throughout the presentation. Though non-public presentations include high-energy behaviors, non-public animal care sessions also focus on training new behaviors and engaging in husbandry behaviors. Public presentations provide the opportunity for exercise through a variety of higher energy behaviors, and non-public sessions afford the ability to engage in comprehensive animal care and time to work on new behaviors.

Spatial distribution — speed/quiver

In Block 1 of OTS, the dolphins had relatively low speeds (mean $1.30\,\mathrm{m\,s^{-1}}$) across their habitat, and based on the vector field of the quiver plot for the block, were engaged in large, smooth loops along the edges of the habitat (Fig. 4, right column). This was contrasted with Block 2, which saw a higher general speed (mean $1.57\,\mathrm{m\,s^{-1}}$) as well as diversified movement patterns, with the right half exhibiting counter-clockwise chirality while the left half maintained the clockwise motion pattern. Blocks 3-5 exhibited higher mean speeds (Blk. 3: $1.45\,\mathrm{m\,s^{-1}}$, Blk. 4: $1.41\,\mathrm{m\,s^{-1}}$, Blk. 5: $1.43\,\mathrm{m\,s^{-1}}$) than Block 1, and lower than 2, with the dolphins' movement patterns shifting changing between each OTS block.

During ITS, the care blocks' (Blk. 1, 3) speed distributions and vector fields qualitatively demonstrated similar structures, while those of the presentations (Blk. 2, 4) were more mixed, with more similarities along the left and right far sides, but fewer in the center (Fig. 5, right column). The mean speeds did not share particular similarities between blocks of similar type (Blk. 1: $1.39\,\mathrm{m\,s^{-1}}$, Blk. 2: $1.45\,\mathrm{m\,s^{-1}}$, Blk. 3: $1.44\,\mathrm{m\,s^{-1}}$, Blk. 4: $1.39\,\mathrm{m\,s^{-1}}$).

Statistical comparison of metrics

Figure 6, top, displays the overlaid PDFs of the speed and yaw metrics during OTS, and Figure 6, middle, displays the PDFs during ITS. The K-S distances for all six metrics were reported in Table 3, with all values rounded to 3 digits of precision. For OTS, we saw from the K-S results that Blocks 1 and 2 varied the most with respect to the others in terms of speed, which was observed in Figure 6, top, while the yaw values were not generally significantly different, again observed in Fig. 6 (given the high number of samples used to generate the K-S statistics, we were able to compare the significance levels to a stronger threshold of $\alpha_{crit} = 0.001$). Across the board, Block 2 generally differed significantly from the rest of the OTS blocks for the most metrics, with Block 1 following close behind. In contrast, Blocks 3-5 differed the least significantly from each other, indicating similarities in the dolphins' dynamics patterns for Blocks 3-5.

Table 3. Kolmogorov-Smirnov Session Comparison

			Speed		**Yaw**		**Yaw Rate**	
	Blk.		Δ_{ks}	α	Δ_{ks}	α	Δ_{ks}	α
OTS	1	2	0.187	< 0.001	0.028	< 0.001	0.047	< 0.001
	1	3	0.095	< 0.001	0.021	0.025	0.034	< 0.001
	1	4	0.080	< 0.001	0.019	0.049	0.057	< 0.001
	1	5	0.079	< 0.001	0.021	0.027	0.035	< 0.001
	2	3	0.096	< 0.001	0.028	< 0.001	0.017	0.099
	2	4	0.111	< 0.001	0.025	0.003	0.028	< 0.001
	2	5	0.110	< 0.001	0.023	0.012	0.016	0.148
	3	4	0.026	0.002	0.019	0.046	0.025	0.004
	3	5	0.026	0.003	0.022	0.012	0.010	0.685
	4	5	0.018	0.093	0.013	0.403	0.030	< 0.001
ITS	1	2	0.059	< 0.001	0.028	< 0.001	0.022	0.017
	1	3	0.021	0.019	0.020	0.039	0.008	0.871
	1	4	0.059	< 0.001	0.028	0.001	0.021	0.020
	2	3	0.061	< 0.001	0.023	0.009	0.028	< 0.001
	2	4	0.043	< 0.001	0.010	0.638	0.008	0.940
	3	4	0.068	< 0.001	0.029	< 0.001	0.028	< 0.001

			Speed σ		**Yaw σ**		**Yaw Rate σ**	
	Blk.		Δ_{ks}	α	Δ_{ks}	α	Δ_{ks}	α
OTS	1	2	0.047	< 0.001	0.035	< 0.001	0.076	< 0.001
	1	3	0.012	0.434	0.026	0.002	0.053	< 0.001
	1	4	0.025	0.004	0.029	< 0.001	0.062	< 0.001
	1	5	0.014	0.249	0.015	0.222	0.040	< 0.001
	2	3	0.047	< 0.001	0.031	< 0.001	0.033	< 0.001
	2	4	0.065	< 0.001	0.039	< 0.001	0.043	< 0.001
	2	5	0.051	< 0.001	0.048	< 0.001	0.043	< 0.001
	3	4	0.025	0.005	0.016	0.153	0.014	0.264
	3	5	0.008	0.889	0.026	0.002	0.026	0.002
	4	5	0.025	0.003	0.032	< 0.001	0.035	< 0.001
ITS	1	2	0.033	< 0.001	0.108	< 0.001	0.092	< 0.001
	1	3	0.027	0.001	0.012	0.423	0.016	0.139
	1	4	0.040	< 0.001	0.096	< 0.001	0.086	< 0.001
	2	3	0.046	< 0.001	0.103	< 0.001	0.100	< 0.001
	2	4	0.014	0.303	0.014	0.264	0.026	0.003
	3	4	0.056	< 0.001	0.093	< 0.001	0.095	< 0.001

For ITS, we note that the significant differences in metrics generally followed the

structure type of each ITS block: comparisons between Blocks 1 vs. 3, and 2 vs. 4, were found to be significantly different the least often. This was to be expected, given Blocks 1 and 3 were animal care sessions, and 2 and 4 were presentations. Of particular note are the yaw std. dev. and yaw rate std. dev. metrics, with entire order of magnitude differences in K-S distances when comparing similar vs. different types of ITS blocks.

Speed and yaw joint entropy

The joint differential entropies of speed and yaw per time block are displayed in Figure 6, bottom, with values reported in Table 4. The time blocks in this figure were presented in chronological order, and with that in mind we observed that the first blocks of each OTS and ITS had the least joint entropy (variation in speed and yaw throughout the time block), followed immediately by a peak in the second block of each. Subsequent time blocks for both OTS and ITS then yielded lower entropies that were sustained. Overall, ITS blocks were observed to have higher speed-yaw joint entropy than OTS blocks in similar time windows.

Fig 6. Speed and yaw probability distributions and joint differential entropies, respective to time block. TOP: Probability density functions of animal speed ($\mathrm{m\,s^{-1}}$) for OTS (left) and ITS (right). MIDDLE: Probability density functions of yaw (rad) for OTS (left) and ITS (right). BOTTOM: Joint differential entropy of speed and yaw for each block of OTS (left) and ITS (right), with limited-range y-axes to more clearly show value differences.

Table 4. Speed and Yaw Joint Differential Entropy

	OTS					ITS			
Block	1	2	3	4	5	1	2	3	4
Entropy	2.358	2.599	2.543	2.508	2.541	2.521	2.675	2.584	2.605

Discussion

Automatic dolphin detection

This research presents a framework that enables the persistent monitoring of managed dolphins through external sensing, performed on a scale that would otherwise require a prohibitively high amount of human effort. Both the Faster R-CNN dolphin detection and CNN drain detection methods displayed reliable performance in testing, and enabled large-scale data processing at rates not achievable by humans. Given that the total duration of video processed was $\sim$ 199 hours (2 cameras $\times$ 99.5 hours each), an inference time of $\sim$ 202 hours (1.013$\times$) represents at minimum an order-of-magnitude increase in processing speed when compared to human data annotation. This estimate was obtained from the authors' prior experience in manual animal tracking, which could take over 10 hours of human effort per hour of video (frame rate of 10 Hz) annotated for a *single* animal. As such, the performance of this detection framework presents new opportunities in long-term animal monitoring, and enables the automated processing of longer duration and more frequent recording sessions. In this research, use of the monitoring framework enabled the large-scale animal position and kinematic state data necessary to yield insights into animal behavior and spatial use within their environment.

Animal kinematics and habitat use

Kinematic diversity

Joint dynamic entropy was used to quantify differences in animal kinematic diversity throughout the day to explore how temporal changes in the dolphins' habitat would result in modified kinematic diversity levels (Fig. 6, bottom). The use of entropy as a proxy for kinematic diversity has been applied in the past to characterize prey motion unpredictability for predator evasion, however in this work it serves to provide a measure of animal engagement [30]. We observed the lowest kinematic diversity in the mornings as the animal care specialists were arriving at work and setting up for the day. The highest kinematic diversity when not interacting with animal care specialists then occurred immediately after the first ITS time block. In general, the first time blocks of both OTS and ITS showed the lowest kinematic diversity of their type, the second of each showed the highest, and the following blocks stabilized between the two extremes. The speed/quiver plots (Figs. 4-5, right) provide a qualitative understanding of the entropy results. For example, in Block 1 of OTS (Fig. 4, top-right) the dolphins engaged in slow swimming throughout their habitat in smooth consistent cycles along the environment edge, yielding the lowest joint entropy. Joint entropy then increased during both the morning ITS and OTS blocks and remained elevated for the rest of the day, representing higher animal engagement through the middle of their waking hours.

This is consistent with previous research on animal activity and sleep patterns, which reports a diurnal activity cycle for managed animals [17]. However, it is interesting to note that changes in animal kinematic diversity throughout the day during OTS are not gradual: the OTS time block displaying the minimum value is immediately followed by the block displaying the maximum, and are only separated by the first training session (30 minute duration). This sudden shift may not be fully explained by only the dolphins' diurnal activity cycle, and may be related to the fact that their first daily interactions with the ACSs occur between these two OTS time blocks. A finer time-scale analysis of their kinematic diversity trends is necessary to determine which is the cause for this change in animal engagement.

Habitat use

The kinematic data also enabled the investigation into how features in the habitat influenced animal behavior and spatial use, particularly during OTS. The animals tended to have a general focus on the area between the gates along the edge of the central island (Fig. 4, left). Additionally, throughout the OTS position plots (including static, Fig. 3, left) four animal-preferred locations were observed. The two hot spots to the left and right of the central island are gates (Fig. 1, middle, Fig. 2, top), where the dolphins could communicate with conspecifics when closed or pass through to other areas of their habitat when open. Conversely, the two hot spots nearer the middle of the island edge corresponded to underwater windows that led to an ACS work area (two central windows in Fig. 2, top/middle). Through these windows the dolphins may observe the ACSs, view conspecifics in one of the back habitats (through an additional window, not shown in Fig. 2), or observe enrichment occasionally placed on the other side of the glass (mirrors, videos, etc.). Regions of the habitat in proximity to these two windows experienced some of the highest occupancy in all OTS position plots, both static and dynamic. This indicates that particular attractors for the dolphins' attention were observable through those windows, whether they were the ACSs, conspecifics, or enrichment.

These attractors also influenced the dolphins' kinematics and activity levels. Of all the regions in the environment, only the positions in front of the central windows consistently recorded peak or near-peak location-specific animal swimming speeds for all

OTS time blocks (Fig. 4, right). When combined with the results from the position distributions (Fig. 4, left), this implies that these dolphins not only focused their attention on these attractors, their presence correlated to higher activity levels in the dolphins when swimming in their vicinity.

Behavior classification from dynamics metrics

During ITS blocks, ACSs asked for specific behaviors from the dolphins and these behaviors were often repeated. Elements of public educational presentations (ITS 2/4) were varied to include a mixture of both high and low energy segments, and this blend resulted in similar dynamic patterns for the public sessions. In contrast, the non-public animal husbandry and training sessions (ITS 1/3) were less dynamic overall, and yielded similar dynamic patterns for these sessions. Qualitative similarities in the pairs of animal training sessions were observable in both the position and speed/quiver plots in Fig. 5, and the probability density functions presented in Fig. 6.

The K-S statistics were used to quantify the similarities and differences between time blocks within both OTS and ITS. As the ACSs requested similar behaviors during ITS blocks of the same type, we expected similarities in the dynamics metrics for Blocks 1 vs. 3 and Blocks 2 vs. 4, and differences between the metrics for blocks of different types. The pattern displayed by the K-S statistics in Table 3 (particularly in the std. devs.) showed by far the most significant differences between time blocks of different types, and the fewest for blocks of the same type. Without prior knowledge of the block types, it would be possible to use this pattern to identify that Blocks 1 and 3 were likely the same type, as were 2 and 4. This demonstrated that the presented method of obtaining and analyzing dolphins' dynamics metrics was sufficient to differentiate between general behavior types.

This was useful for analyzing the OTS results, as the position and speed/quiver plots in Fig. 4 only showed patterns in the animals' location preferences within their habitat. In contrast, the K-S statistics gave a clearer view of the differences between OTS time blocks. Block 2 separated itself significantly from all other time blocks in nearly every metric, while Block 1 was in a similar position (though not as pronounced). Blocks 3-5 showed few significant differences for metrics comparisons between each other. This indicated that the dolphins had more distinct dynamics for Blocks 1 and 2, and maintained similar dynamics patterns throughout Blocks 3-5. When combined with the joint differential entropy values, these results indicated there may be three general OTS behavior types for the dolphins in this dataset (in terms of kinematic diversity [KD]): "Low KD" at the beginning of the day (Block 1), "High KD" immediately after the first training session (Block 2), and "Medium KD" for the remainder of the day (Blocks 3-5). A fine-scale temporal analysis of animal kinematic diversity should reveal whether these behavior transitions are dependent on the ACSs or other factors.

Limitations and future work

Using a limited number of cameras meant full stereo coverage of the habitat was not possible, preventing a direct estimate of animal depth. Additionally, camera placements resulted in region-specific glare on the surface of the water that impeded the Faster R-CNN detector. To address these problems, cameras could be added in locations that allow for fully overlapping coverage, at angles that avoid glare in the same regions. Further, installing cameras capable of low-light recording could enable night monitoring sessions. An inherent problem with camera-based tracking is the fact that similarities between dolphin profiles make it challenging to identify individuals. This problem has been addressed in [28], where kinematic data from dolphin-mounted biologging tags were used to filter camera-based animal location data. This filtering process made it

more feasible to identify which location data points corresponded to specific tagged individuals, coupling the kinematic and location data streams for these animals. Fusing the coupled tag and camera data through methods similar to [28] or [31] would then provide high-accuracy localization information to contextualize the detailed kinematics data produced by the tags.

Conclusions

Through this research we have demonstrated a monitoring framework that offers new options for long-term managed dolphin observation, while significantly enhancing the efficiency of both data collection and analysis. This work demonstrated the feasibility of a camera-based computer-automated marine animal tracking system, and explored its capabilities by analyzing the behavior and habitat use of a group of managed dolphins over a large time scale. From the results, we were able to quantify day-scale temporal trends in the dolphins' spatial distributions, dynamics patterns, and kinematic diversity modes. These in turn revealed that habitat features associated with particular attractors served as focal points for this group of dolphins: these features were correlated with higher animal physical proximity, kinematic diversity (specifically ACS presence), and activity levels.

Acknowledgments

The authors would like to thank the Brookfield Zoo for its aid in facilitating this research. Rita Stacey and the Seven Seas Animal Care Specialists were instrumental in helping to acquire such a large volume of data, and the help of the Zoo's administration made this research a possibility. Finally, the authors would like to thank Sarah Breen Bartecki and William Zeigler of the Chicago Zoological Society for their continued support.

Competing interests

The authors have no competing interests to declare.

Funding

This research was funded by The Granger Foundation and The Chicago Zoological Society.

References

1. Kagan R, Carter S, Allard S. A Universal Animal Welfare Framework for Zoos. Journal of Applied Animal Welfare Science. 2015;18. doi:10.1080/10888705.2015.1075830.

2. Miller LJ, Mellen J, Greer T, Kuczaj SA. The effects of education programmes on Atlantic bottlenose dolphin (Tursiops truncatus) behaviour. Animal Welfare. 2011;20(2):159–172.

3. Whitham JC, Wielebnowski N. New directions for zoo animal welfare science. Applied Animal Behaviour Science. 2013;147(3-4):247–260. doi:10.1016/j.applanim.2013.02.004.

4. Mason GJ. Species differences in responses to captivity: Stress, welfare and the comparative method. Trends in Ecology and Evolution. 2010;25(12):713–721. doi:10.1016/j.tree.2010.08.011.

5. Alex Shorter K, Shao Y, Ojeda L, Barton K, Rocho-Levine J, van der Hoop J, et al. A day in the life of a dolphin: Using bio-logging tags for improved animal health and well-being. Marine Mammal Science. 2017;33(3):785–802. doi:10.1111/mms.12408.

6. Clegg ILK, Borger-Turner JL, Eskelinen HC. C-Well: The development of a welfare assessment index for captive bottlenose dolphins (Tursiops truncatus). Animal Welfare. 2015;24(3):267–282. doi:10.7120/09627286.24.3.267.

7. Ugaz C, Valdez RA, Romano MC, Galindo F. Behavior and salivary cortisol of captive dolphins (Tursiops truncatus) kept in open and closed facilities. Journal of Veterinary Behavior: Clinical Applications and Research. 2013;8(4):285–290. doi:10.1016/j.jveb.2012.10.006.

8. Waples KA, Gales NJ. Evaluating and minimising social stress in the care of captive bottlenose dolphins (Tursiops aduncus). Zoo Biology. 2002;21(1):5–26. doi:10.1002/zoo.10004.

9. Johnson MP, Tyack PL. A digital acoustic recording tag for measuring the response of wild marine mammals to sound. IEEE Journal of Oceanic Engineering. 2003;28(1):3–12. doi:10.1109/JOE.2002.808212.

10. Zhang D, van der Hoop JM, Petrov V, Rocho-Levine J, Moore MJ, Shorter KA. Simulated and experimental estimates of hydrodynamic drag from bio-logging tags. Marine Mammal Science. 2020;36(1):136–157. doi:10.1111/mms.12627.

11. Aguilar Soto N, Johnson MP, Madsen PT, Díaz F, Domínguez I, Brito A, et al. Cheetahs of the deep sea: Deep foraging sprints in short-finned pilot whales off Tenerife (Canary Islands). Journal of Animal Ecology. 2008;77(5):936–947. doi:10.1111/j.1365-2656.2008.01393.x.

12. Sibal R, Zhang D, Rocho-Levine J, Shorter KA, Barton K. Bidirectional LSTM Recurrent Neural Network Plus Hidden Markov Model for Wearable Sensor-Based Dynamic State Estimation. ASME Letters in Dynamic Systems and Control. 2021;1(2). doi:10.1115/1.4046685.

13. Zhang D, Alex Shorter K, Rocho-Levine J, Van Der Hoop J, Moore M, Barton K. Behavior inference from bio-logging sensors: A systematic approach for feature generation, selection and state classification. In: ASME 2018 Dynamic Systems and Control Conference, DSCC 2018. vol. 2. American Society of Mechanical Engineers (ASME); 2018.Available from: `http://asmedigitalcollection.asme.org/DSCC/proceedings-pdf/DSCC2018/51906/V002T21A005/2376728/v002t21a005-dscc2018-9213.pdf`.

14. Ames AE, Macgregor RP, Wielandt SJ, Cameron DM, Kuczaj SA, Hill HM. Pre-and post-partum whistle production of a bottlenose dolphin (Tursiops truncatus) social group. International Journal of Comparative Psychology. 2019;32:1–17. doi:10.46867/ijcp.2019.32.02.02.

15. Harvey BS, Dudzinski KM, Kuczaj SA. Associations and the role of affiliative, agonistic, and socio-sexual behaviors among common bottlenose dolphins (Tursiops truncatus). Behavioural Processes. 2017;135:145–156. doi:10.1016/j.beproc.2016.12.013.

16. Clegg ILK, Rödel HG, Cellier M, Vink D, Michaud I, Mercera B, et al. Schedule of human-controlled periods structures bottlenose dolphin (tursiops truncatus) behavior in their free-time. Journal of Comparative Psychology. 2017;131(3):214–224. doi:10.1037/com0000059.

17. Sekiguchi Y, Kohshima S. Resting behaviors of captive bottlenose dolphins (Tursiops truncatus). Physiology and Behavior. 2003;79(4-5):643–653. doi:10.1016/S0031-9384(03)00119-7.

18. Walker RT, Miller LJ, Kuczaj SA, Solangi M. Seasonal, diel, and age differences in activity budgets of a group of bottlenose dolphins (Tursiops truncatus) under professional care. International Journal of Comparative Psychology. 2017;30. doi:10.46867/ijcp.2017.30.00.05.

19. Karnowski J, Hutchins E, Johnson C. Dolphin detection and tracking. Proceedings - 2015 IEEE Winter Conference on Applications of Computer Vision Workshops, WACVW 2015. 2015; p. 51–56. doi:10.1109/WACVW.2015.10.

20. Rachinas-Lopes P, Ribeiro R, Dos Santos ME, Costa RM. D-Track—A semi-automatic 3D video-tracking technique to analyse movements and routines of aquatic animals with application to captive dolphins. PLoS ONE. 2018;13(8):e0201614. doi:10.1371/journal.pone.0201614.

21. Brunetti A, Buongiorno D, Trotta GF, Bevilacqua V. Computer vision and deep learning techniques for pedestrian detection and tracking: A survey. Neurocomputing. 2018;300:17–33. doi:10.1016/j.neucom.2018.01.092.

22. Jiménez-García B, Aznarte J, Abellán N, Baquedano E, Domínguez-Rodrigo M. Deep learning improves taphonomic resolution: high accuracy in differentiating tooth marks made by lions and jaguars. Journal of The Royal Society Interface. 2020;17(168):20200446. doi:10.1098/rsif.2020.0446.

23. Arafati A, Morisawa D, Avendi MR, Amini MR, Assadi RA, Jafarkhani H, et al. Generalizable fully automated multi-label segmentation of four-chamber view echocardiograms based on deep convolutional adversarial networks. Journal of The Royal Society Interface. 2020;17(169):20200267. doi:10.1098/rsif.2020.0267.

24. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017;39(6):1137–1149. doi:10.1109/TPAMI.2016.2577031.

25. He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision. vol. 2015 Inter; 2015. p. 1026–1034.

26. Mathworks. Define a Custom Deep Learning Layer with Learnable Parameters; 2019. Available from: `https://www.mathworks.com`.

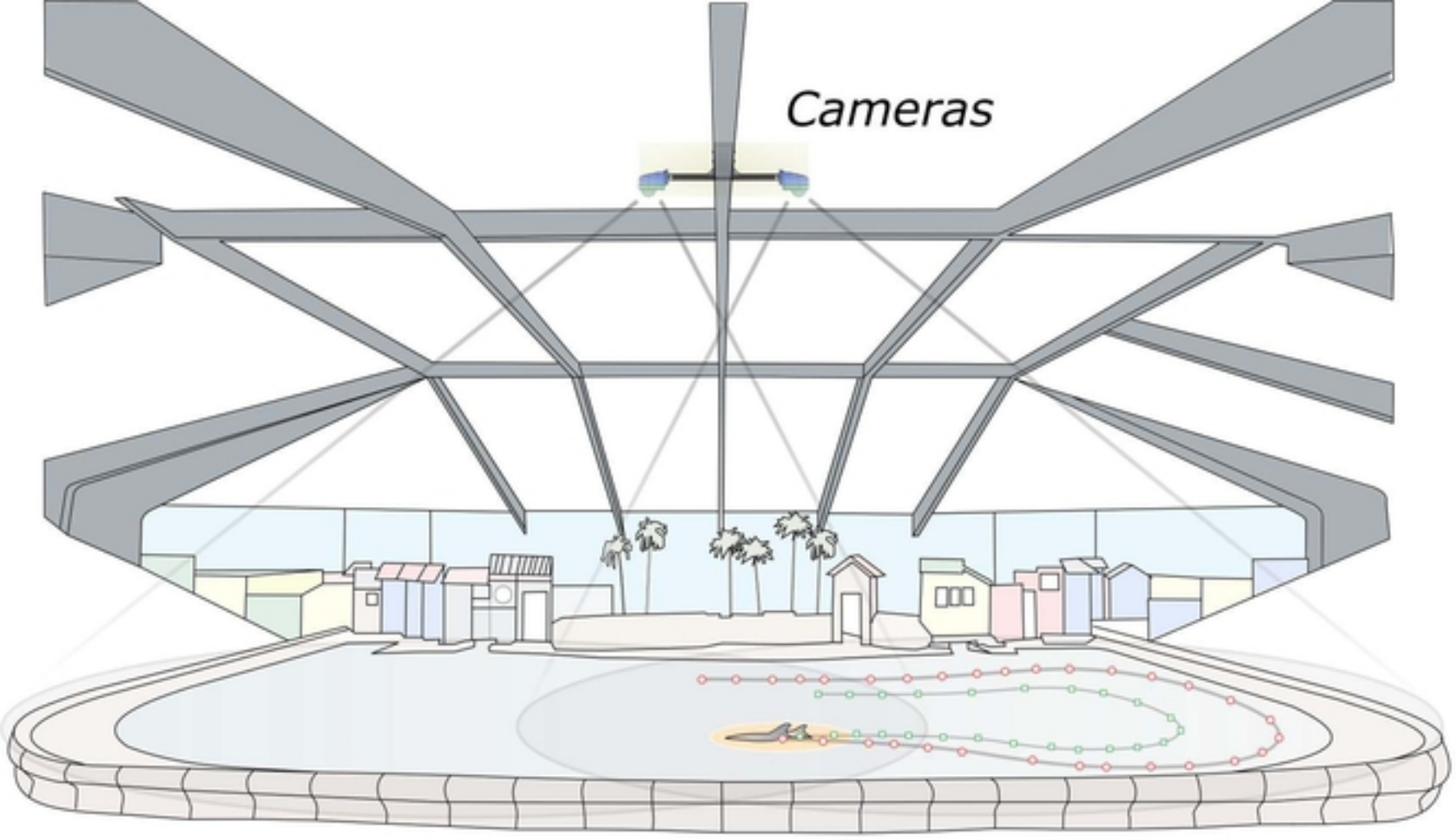
27. Hartley RI. In Defense of the Eight-Point Algorithm. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE. 1997;19(6).

28. Zhang D, Gabaldon J, Lauderdale L, Johnson-Roberson M, Miller LJ, Barton K, et al. Localization and tracking of uncontrollable underwater agents: Particle filter based fusion of on-body IMUs and stationary cameras. In: Proceedings - IEEE International Conference on Robotics and Automation. vol. 2019-May. Institute of Electrical and Electronics Engineers Inc.; 2019. p. 6575–6581.

29. Porter FC. Testing Consistency of Two Histograms; 2008. Available from: http://arxiv.org/abs/0804.0380.

30. Moore TY, Cooper KL, Biewener AA, Vasudevan R. Unpredictability of escape trajectory explains predator evasion ability and microhabitat preference of desert rodents. Nature Communications. 2017;8(1):1–9. doi:10.1038/s41467-017-00373-2.

31. Gabaldon J, Zhang D, Barton K, Johnson-Roberson M, Shorter KA. A framework for enhanced localization of marine mammals using auto-detected video and wearable sensor data fusion. In: IEEE International Conference on Intelligent Robots and Systems. vol. 2017-Septe. Institute of Electrical and Electronics Engineers Inc.; 2017. p. 2505–2510.



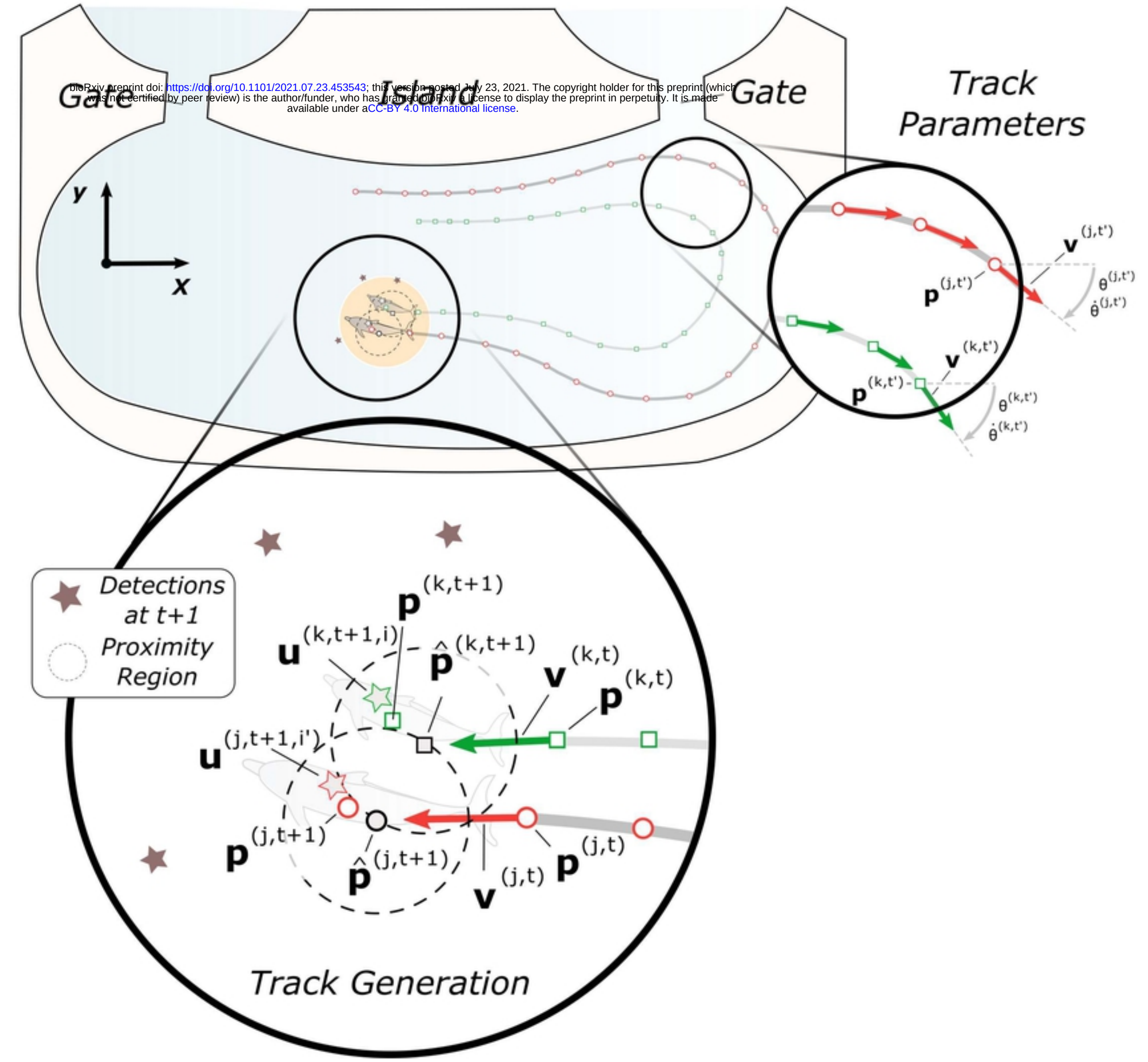
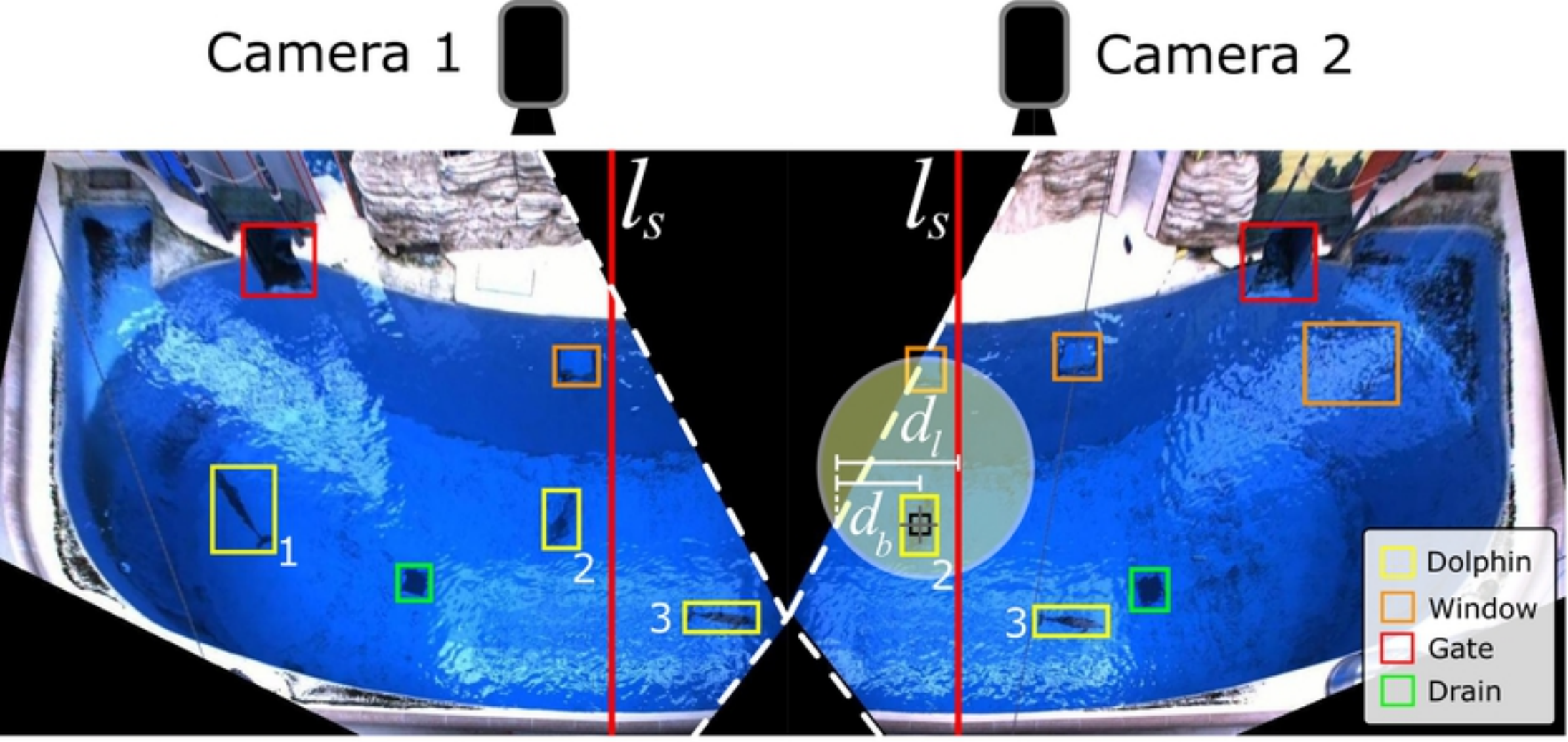
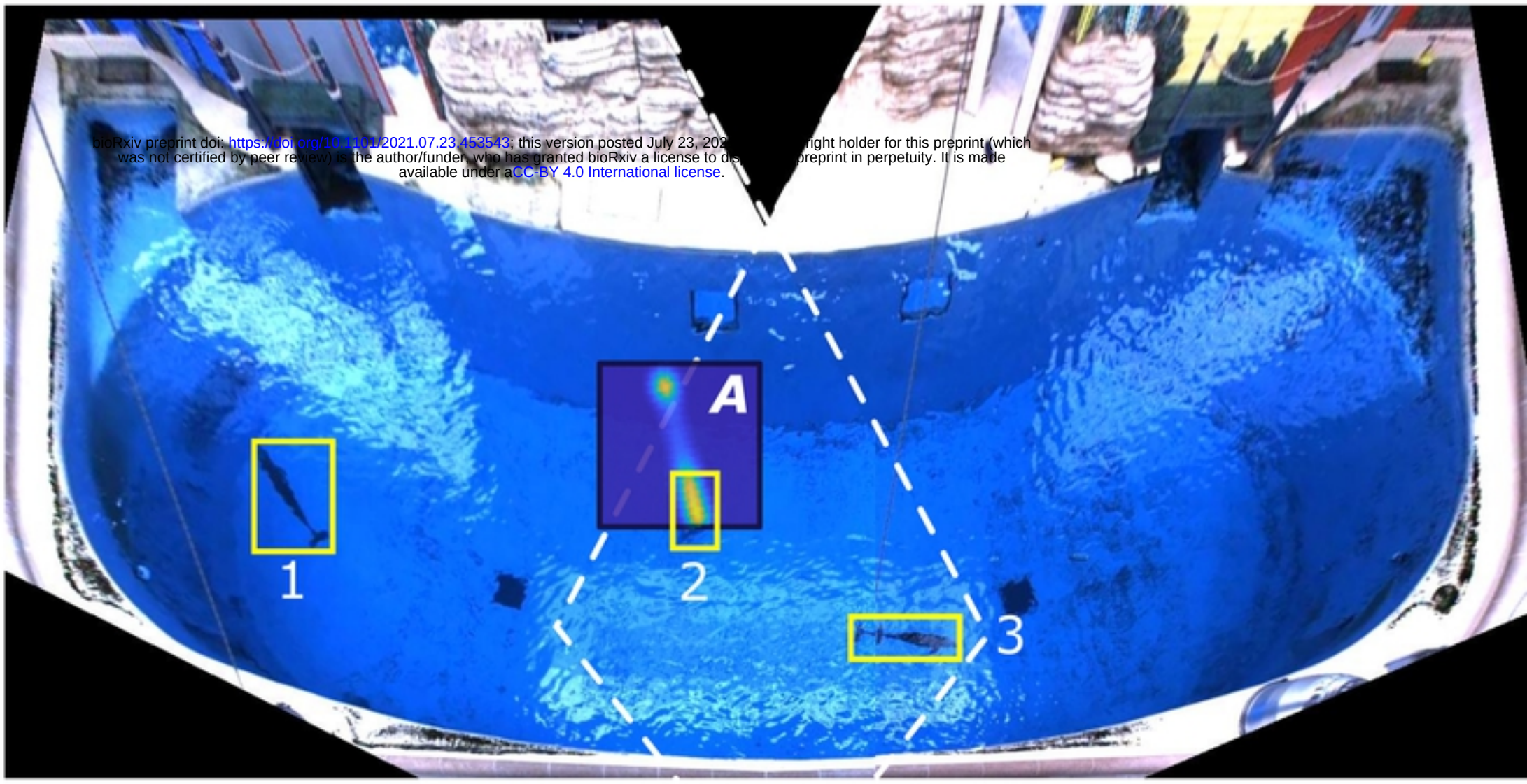
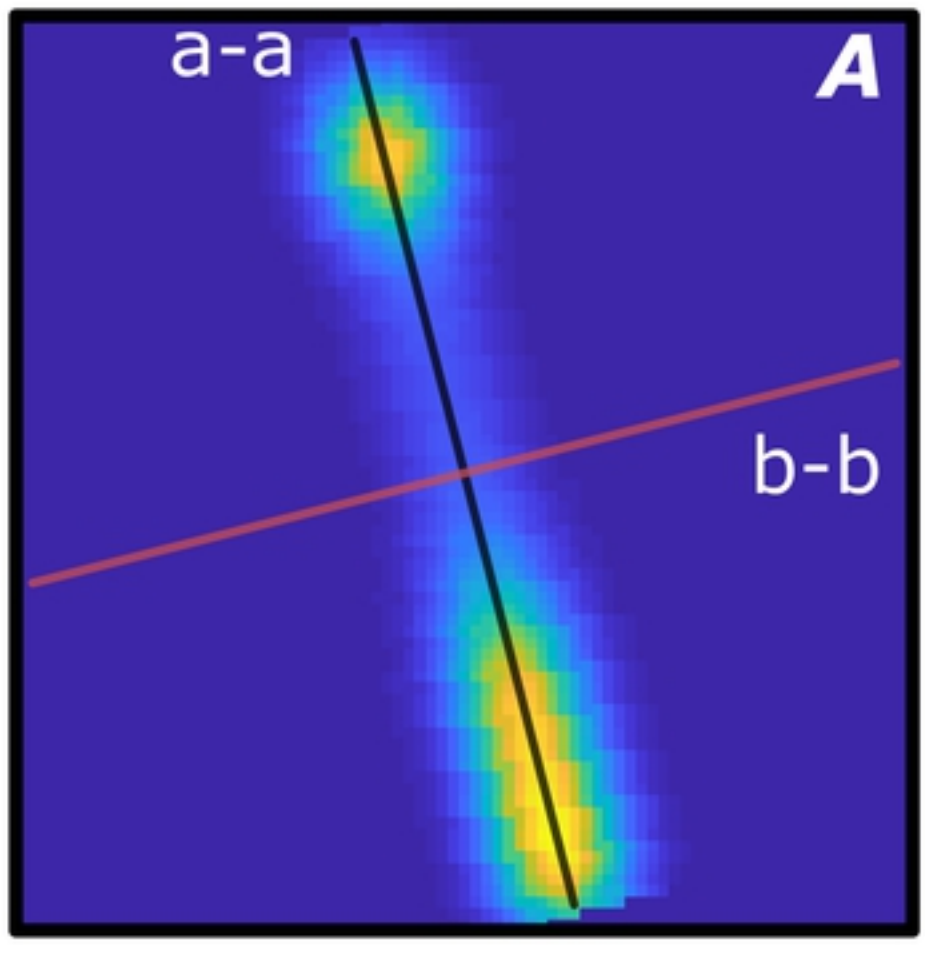


Figure 1


Camera 1
Camera 2
l_s
l_s
d_l
d_b
1
2
3
2
3
Dolphin
Window
Gate
Drain


A
1
2
3


a-a
A
b-b

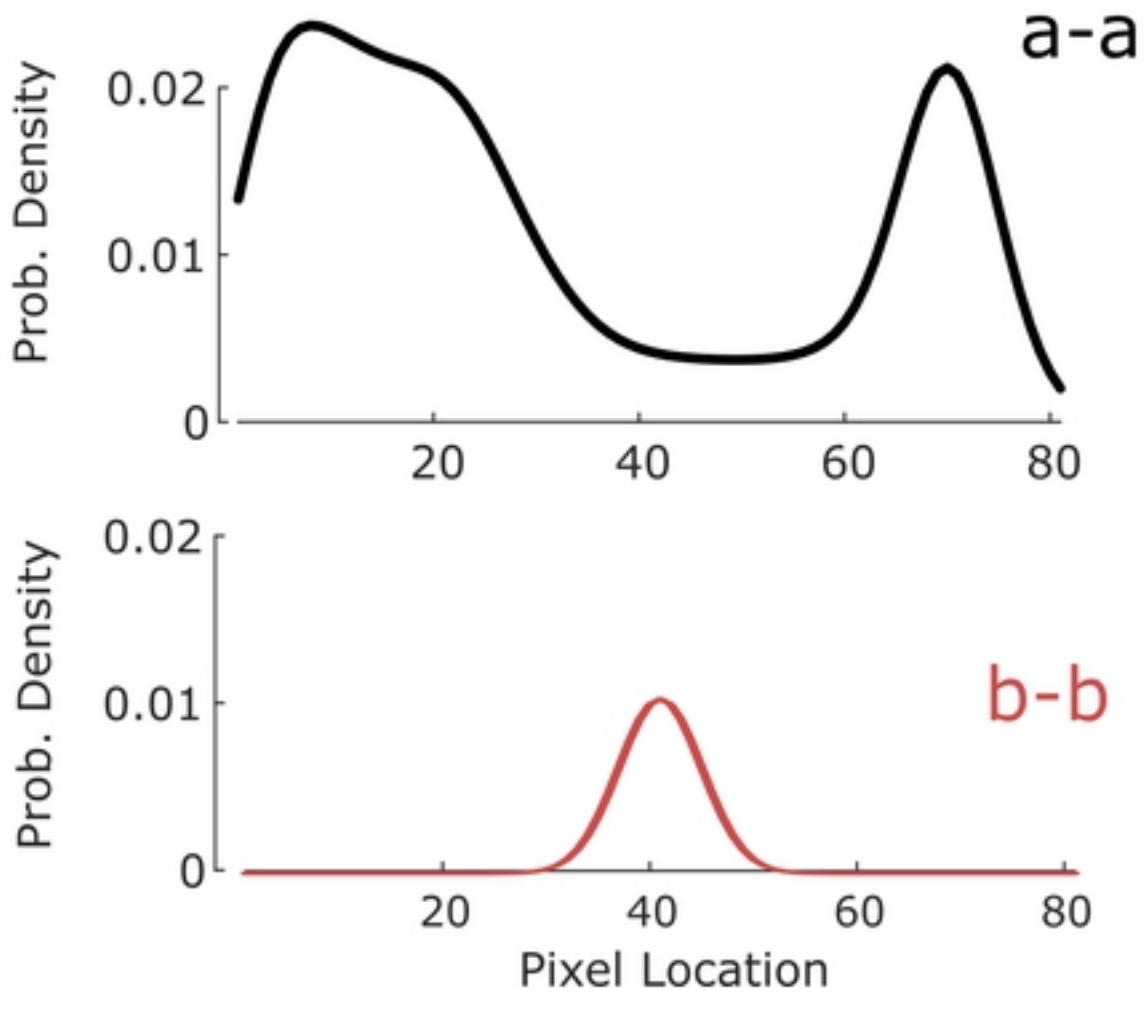

a-a
Prob. Density
0.02
0.01
0
20
40
60
80
b-b
Prob. Density
0.02
0.01
0
20
40
60
80
Pixel Location

Figure 2

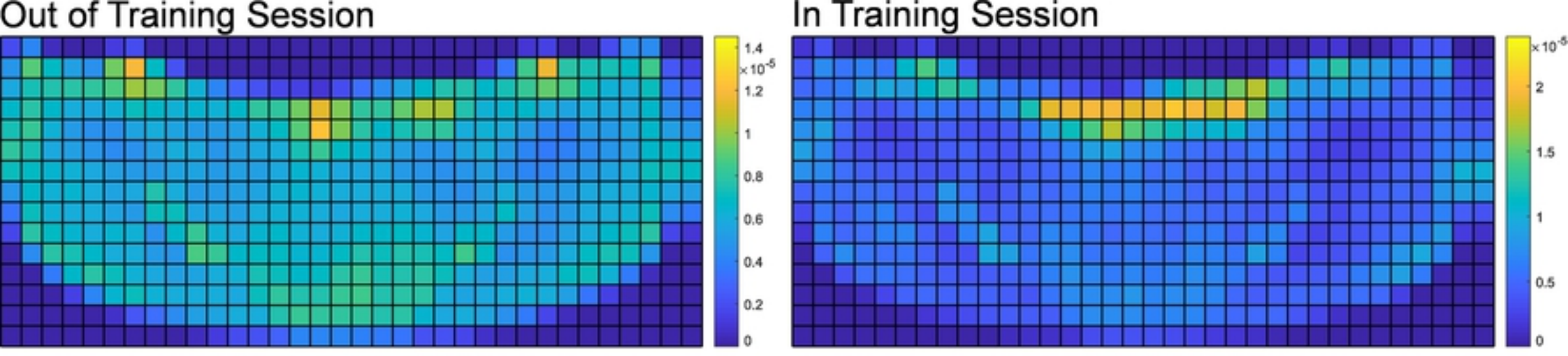

Out of Training Session
In Training Session
1.4
×10-5
1.2
1
0.8
0.6
0.4
0.2
0
×10-5
2
1.5
1
0.5
0

Figure 3

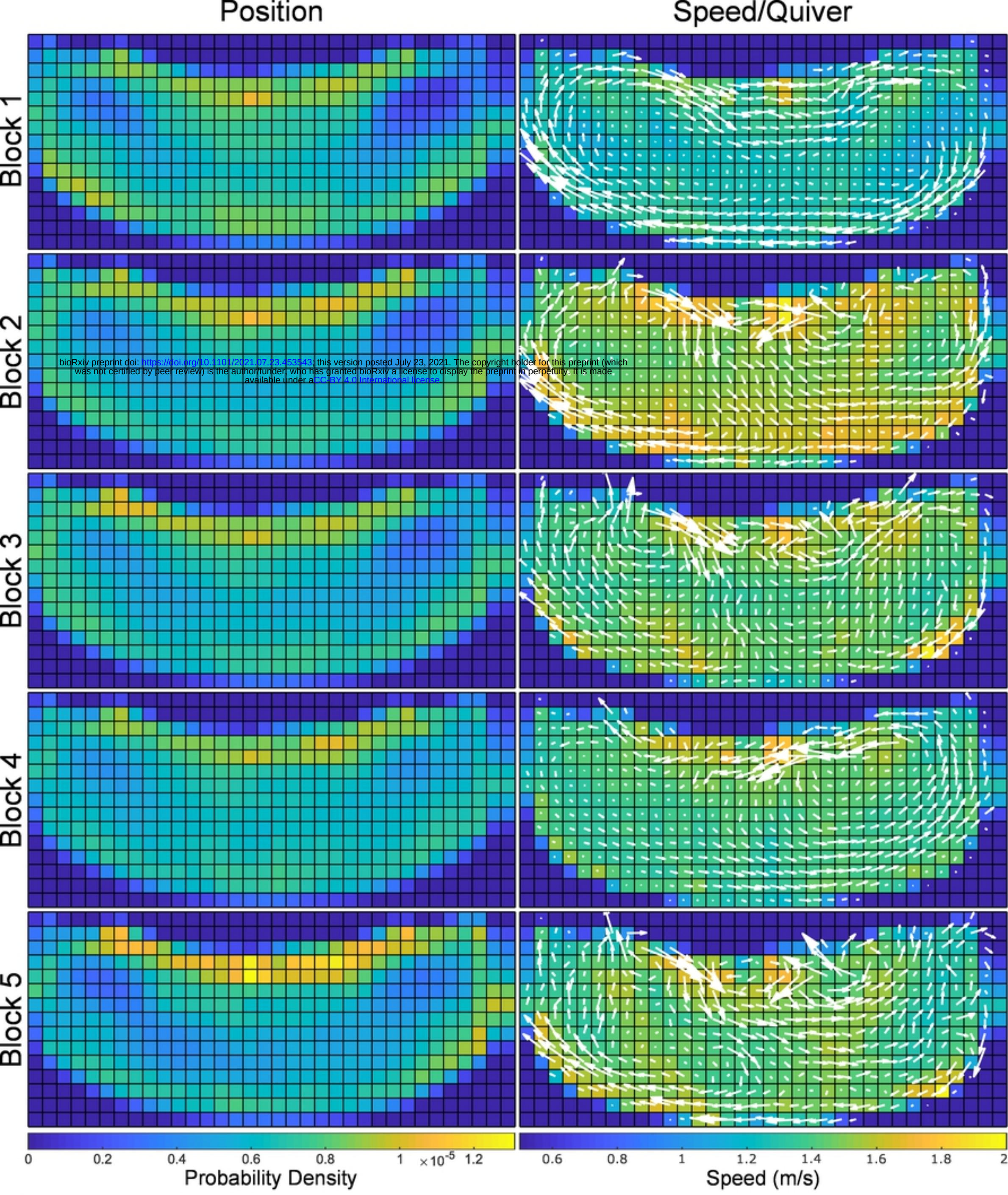


Figure 4

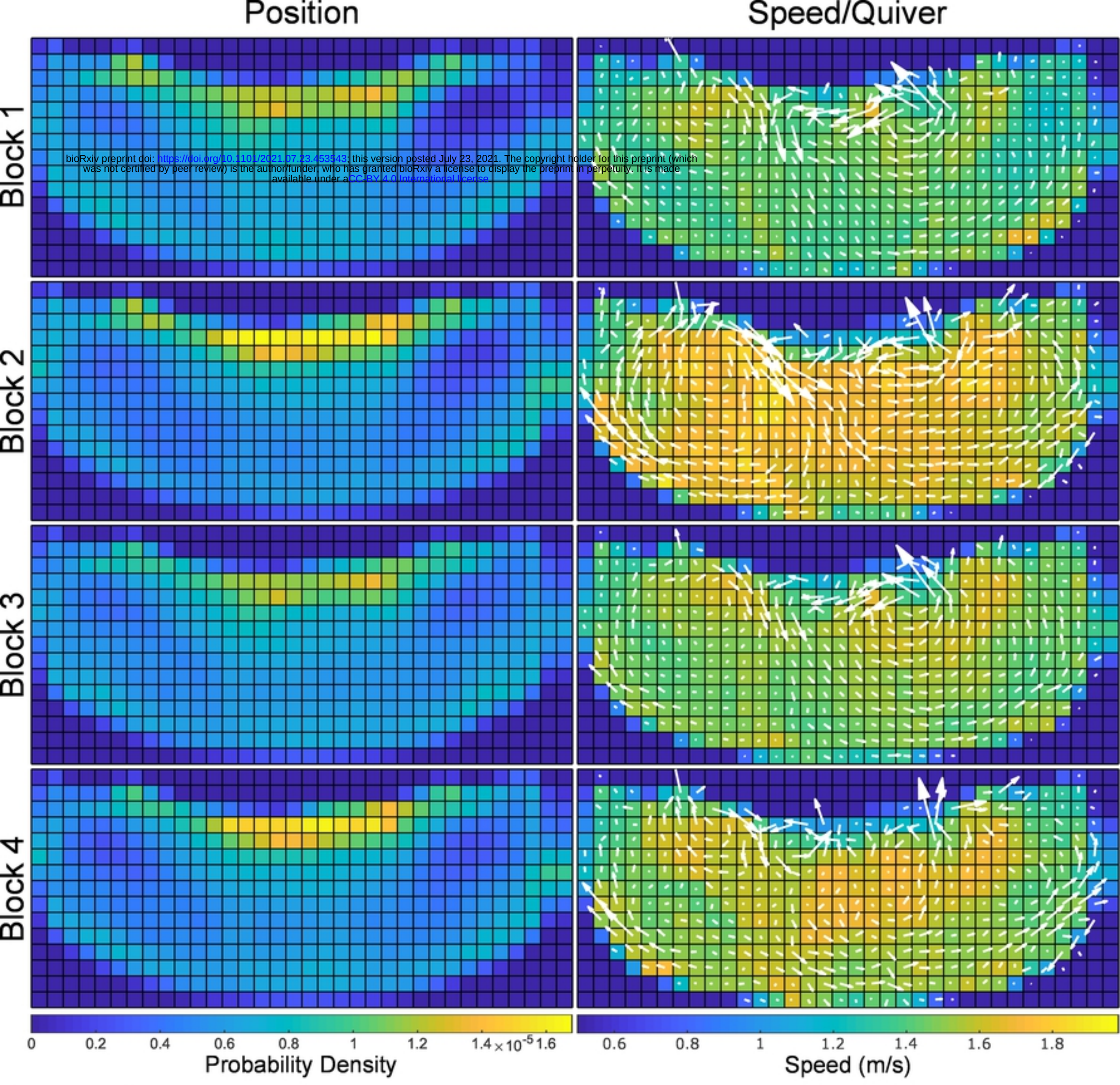


Figure 5

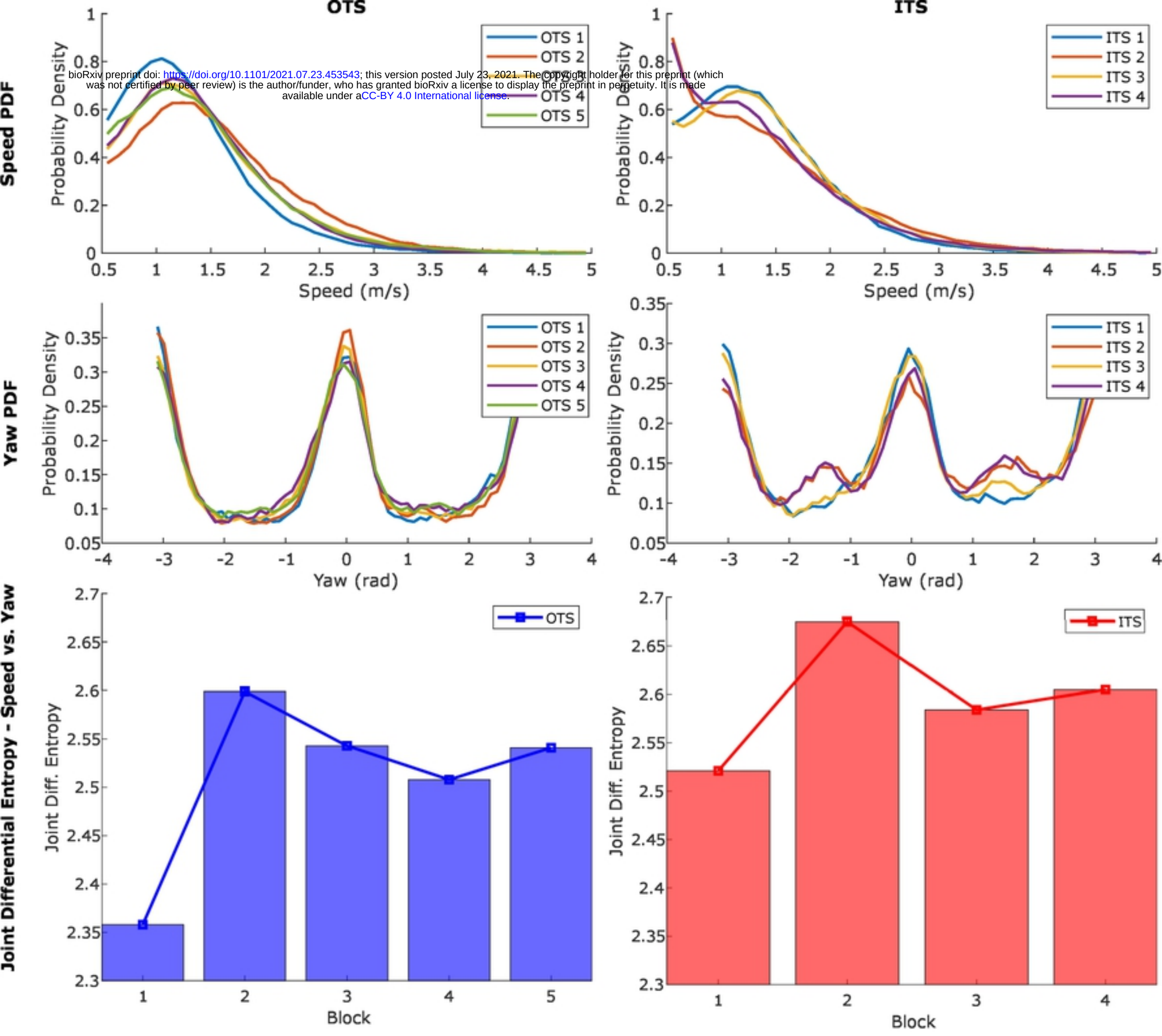


Figure 6