Journal of Experimental Marine Biology and Ecology 576 (2024) 152012

Contents lists available at ScienceDirect

Journal of Experimental Marine Biology and Ecology

journal homepage: www.elsevier.com/locate/jembe

Investigating leatherback surface behavior using a novel tag design and machine learning

Rick Rogers [a,*], Kate H. Choate [a], Leah M. Crowe [a], Joshua M. Hatch [b], Michael C. James [c], Eric Matzen [b], Samir H. Patel [d], Christopher R. Sasso [e], Liese A. Siemann [d], Heather L. Haas [b]

[a] Integrated Statistics Under Contract to the Northeast Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Woods Hole, MA 02543, USA
[b] NOAA National Marine Fisheries Service, NEFSC, Woods Hole, 02543, MA, USA
[c] Population Ecology Division, Fisheries and Oceans Canada, Bedford Institute of Oceanography, Dartmouth, NS B2Y 4A2, Canada
[d] Coonamessett Farm Foundation, 277 Hatchville Road, East Falmouth, 02536, MA, USA
[e] NOAA National Marine Fisheries Service, SEFSC, Miami, 33149, FL, USA

ARTICLE INFO

Keywords:
Sea turtle
Availability
Behavior
Video
Machine learning

ABSTRACT

Understanding the surfacing behavior of marine wildlife is an important component for improving abundance estimates derived from visual surveys. We monitored the behavior of 18 leatherback sea turtles (*Dermochelys coriacea*) in coastal habitats off Massachusetts, USA, using a high-resolution camera and satellite tag package (HiCAS - High Resolution Camera and Satellite) that we assembled from commercially available components which work independently. We used nine data streams derived from the multiple sensors and a video camera to explore four different depth thresholds defining surface zones. We compared classification of video images by a human to classification of those images by a machine learning algorithm. We calculated four metrics to describe surface behavior for each of the nine data streams. The mean percent time at the surface was the only behavior metric that changed systematically as data streams were used to assess different visible depth thresholds, increasing as the depth threshold increased. Other behavior metrics (mean surface duration, mean dive duration and number of surfacing events per hour) were less similar across data streams, making them unreliable for estimating surface availability. This study highlights the need for sustained data collection to better inform the availability bias estimates used to calculate abundance from visual observations.

1. Introduction

Sea turtles have two key physiologically important habitats where they are most often observed: on nesting beaches and at shallow depths at sea. Population monitoring on nesting beaches has become a standard component in conservation assessments (e.g, Wallace et al., 2013, http://www.iucnredlist.org, NMFS and USFWS, 1992, NMFS and USFWS, 2008). The main advantage of nesting-beach monitoring is the relative accessibility of the habitat and low cost, but it has drawbacks because it often only accounts for the nests and, at some sites, the number of reproductive females, which represent a small and variable proportion of the population (Ceriani et al., 2019, Seminoff and Shanker, 2008, Warden et al., 2017). Other important monitoring occurs through aerial and shipboard visual surveys in the surface waters of migratory corridors and foraging areas where sea turtles are observed at the ocean surface (Archibald and James, 2016; Benson et al., 2020; Shoop and Kenney, 1992), and through satellite tagging of animals for extended periods at broad spatial scales. Monitoring turtles at sea increases the probability of encountering more sex-age and age classes (though not small turtles; Schroeder and Thompson, 1987), but is more resource intensive. In addition to standard aerial and shipboard line transect surveys, research at sea may require alternative strategies unique to each habitat, potentially including the use of remotely operated vehicles, baited underwater video, animal-borne cameras, or divers conducting direct observations (Smolowitz et al., 2015; Letessier et al., 2014; Heaslip et al., 2012; Schofield et al., 2006).

As turtles approach the surface of the ocean from deeper water, they become more available to visual observers surveying from boats or planes. When animals are too deep to be visually observed, they are considered to be unavailable to the survey, an issue termed ‘availability

* Corresponding author.
E-mail address: rick.a.rogers@noaa.gov (R. Rogers).

https://doi.org/10.1016/j.jembe.2024.152012
Received 19 May 2023; Received in revised form 10 April 2024; Accepted 15 April 2024
Available online 7 May 2024

bias'. This can cause abundance estimates from line transect surveys to be biased low unless animal behavior data are used to inform the availability bias (Marsh and Sinclair, 1989; Laake et al., 1997; Borchers et al., 2013; Hatch et al., 2022). An estimate of availability bias (A) can be combined with an estimate of the number of animals at the surface (N_s) to produce a "corrected" abundance estimate (N_c) of the number of animals throughout the water column ($N_c = N_s / A$, Heide-Jørgensen and Laidre, 2015). When sea turtle surface behavior is incorporated into abundance estimates from line transect surveys, the abundance can be an order of magnitude higher than the estimate without the availability bias correction (*Northeast Fisheries Science Center*, 2011). Hence, knowing information about turtle behavior in the visible depth threshold allows researchers to correct line transect animal counts into more accurate abundance estimates (Dunstan et al., 2020, Odzer et al., 2022, Barco et al., 2018, Fuentes et al., 2015, Thomson et al., 2012, 2013, DiMatteo et al., 2024). Depth sensors and wet-dry sensors can be used to translate sea turtle surfacing behavior into estimates of availability bias, but research is lacking on how accurate and precise these sensors are in an applied setting such as when carried on an animal-borne tag.

Research on sea turtle availability during visual surveys has been hindered by a two-pronged problem associated with demarcating the visible depth threshold. First, the precise depth of this threshold is difficult to define because visual observers in an airplane cannot accurately estimate how deep they can see. Experimental field studies have begun to address this issue by assessing the availability of sunken objects to visual observers (Benson et al., 2007, Barco et al., 2018, Robbins et al., 2014, Fuentes et al., 2015, Odzer et al., 2022), but the depth of the visible surface layer is dynamically affected by ocean turbidity, sea state, solar intensity and angle, as well as the altitude and speed of the observing platform and the size and coloration of the turtle (Laake et al., 1997). Second, the precise depth of tagged turtles is difficult to assess because pressure sensors on animal-borne tags (especially those used for deep-diving species) often have low resolution near the surface, and detection in surface waters can be highly influenced by subtle variations in depth (Heide-Jørgensen and Lage, 2022). The use of multiple sensors (e.g., wet/dry sensors and pressure sensors) on high resolution animal-borne tags has the potential to better define animal depth in shallow depth thresholds, particularly when animals are at the air-sea interface.

The Northwest Atlantic leatherback sea turtle (*Dermochelys coriacea*) population has recently been assessed as having a high risk of extinction (NMFS and USFWS, 2020). Most of the abundance and productivity information in the assessment was obtained from nesting data, with the declining nesting trends in recent years being an important consideration (NMFS and USFWS, 2020). Abundance estimates based on aerial survey data add critical information for monitoring sea turtle populations (Warden et al., 2017), particularly when understanding potential impacts within short time frames are of interest. Understanding the impacts of anthropogenic activities, such as the development of wind energy, exploration of oil extraction lease areas, and the development of offshore aquaculture requires a better understanding of leatherback behavior, distribution, and abundance particularly at broad spatial scales.

In this study, we used results from custom-built high-resolution tags deployed on leatherback sea turtles to create multiple data streams to examine leatherback surface behavior. The instrumentation on the tag included a video camera, a wet/dry sensor, and multiple depth sensors. We compared the results of each data stream to the presence of turtles at the surface confirmed by human classification of animal-borne video; we used machine learning to investigate the automation of video classification; and we explored the implications of varying the visible depth threshold on quantifying surface behavior.

2. Methods

2.1. High-resolution camera and satellite (HiCAS) tag deployments

2.1.1. Specifications

We assembled a high-resolution camera and satellite tag package (HiCAS - High Resolution Camera and Satellite) by coupling a hydrodynamic and buoyant base block of syntactic foam (model MZ-22, Engineered Syntactic Systems) with several commercially-available components which work independently. Our initial prototype (20.3 × 6.2 × 3.5 cm) was designed to hold a camera, VHF radio transmitter, and global positioning system (GPS) logger. A second version (24.5 × 4.5 × 3.2 cm, Fig. 1) was designed to hold the same camera and a Mk10 satellite data logger (Wildlife Computers, Washington, USA). The syntactic foam was painted orange to increase visibility and was equipped with two suction cups: a 3.25-in. diameter cup at the anterior end to facilitate attachment to study animals and a 3-in. rear stabilizing cup (integrating holes to prevent suction) at the posterior end. This stabilizing cup acted as a bumper to prevent the foam block from rubbing or bouncing against the carapace. As a substitute for the integrated timed-release mechanisms for suction cup instrument deployments on leatherbacks (Migneault et al., 2023; Wallace et al., 2015, 2018), we used galvanic timed releases (International Fishing Devices,Inc., Florida, USA); either AA0.5 or AA1 releases to achieve 3 to 4-h deployment durations. The corrodible galvanic release was secured to the anterior suction cup using zip ties such that the cup would predictably separate from the foam block. (See Fig. 2.)

The electronic components of the tag allowed for the collection of multiple, simultaneous data streams of surface behavior. Electronic components included a camera with integrated temperature-depth sensors, radio transmitter, GPS unit, and the satellite data logger. All tags were equipped with a Paralenz dive camera recording at a resolution of 1080p and 30 frames per second. The Paralenz cameras also had a temperature sensor and time-depth recorder (TDR) that logged every second at resolutions of ±0.1 °C and ± 0.1 m. The satellite data logger was a Wildlife Computers SPLASH10, which had a depth resolution of 0.5 m and an accuracy of ±1%. Temperature and depth recorded by the Paralenz were embedded within the footage and saved in a log file. Battery life at 1080p resolution allowed approximately three hours of recording time. The camera was secured at the anterior end of the foam base using an aluminum action camera mount. To track the tags for recovery, they were equipped with VHF radio transmitters (AI-2, Holohil, Ontario, Canada) with 50 cm flexible antennas. Radio transmitters operated within the 150 MHz range and were secured to the foam base

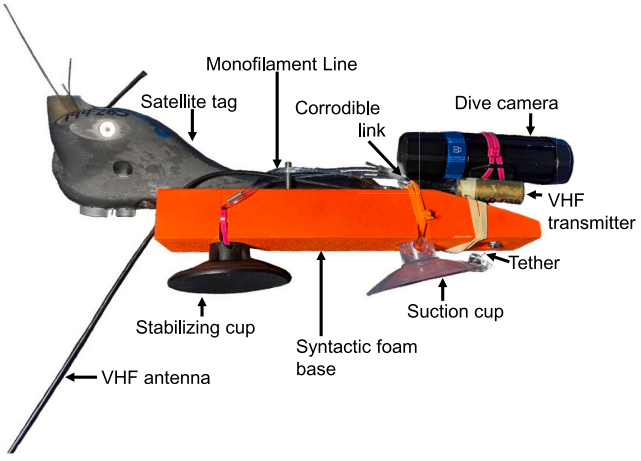

Fig. 1. High-resolution suction cup tag (HiCAS, version 2) as it would ride on the turtle (head to the right), with all instruments oriented on top of the foam block.

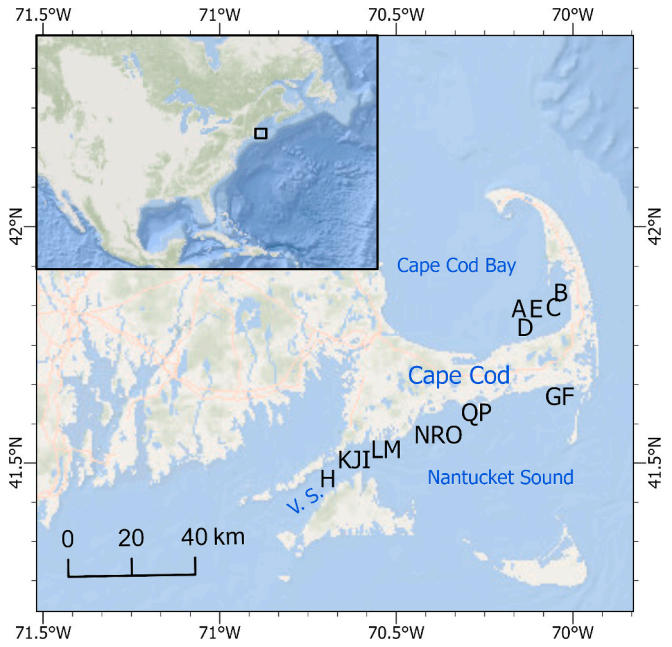

Fig. 2. Location of the study area and tag deployments indicated by the letters identifying each turtle. "V·S" refers to Vineyard Sound. Map credits Esri,GEBCO, DeLorme, and NaturalVue.

using a rubber band and cable ties. Six deployments used stand-alone GPS units that sampled at 10 Hz, resulting in several locations recorded during every surfacing. These GPS units were built in-house at Coonamessett Farm Foundation using Adafruit components (www.adafruit.com). Specifically, we used the Adafruit Feather, Adalogger FeatherWing - RTC + SD Add-on and the Adafruit Ultimate GPS FeatherWing, all powered by a 3.7 V, 500 mAh battery. GPS units were housed in a waterproof action camera case measuring 8.0 × 8.0 × 4.5 cm. Twelve deployments used the Wildlife Computer Mk10s with Fastloc® GPS (Wildlife Computers, 2023) instead of the in-house built GPS units. Depth and temperature were sampled by the satellite data loggers every second and the duty cycle was set to attempt data transmissions and receive GPS locations at every surfacing. The satellite data loggers had a wet/dry sensor to enable transmissions when the tag was out of water. A 50-cm monofilament line coupled with a galvanic release secured the satellite data logger to the foam base.

When the corrodible galvanic link broke, the tether raised the lip of the suction cup, releasing the tag from the turtle while retaining the suction cup. When free floating, the weight of the tag's electrical components caused it to flip upside down, with the suction cups and VHF antenna at the surface of the water and facing skyward. The breaking of the corrodible galvanic link also caused the monofilament securing the satellite data logger to the foam base to unravel. This allowed the satellite data logger to remain attached to the foam base but float right-side up, unencumbered by other tag components. With both the satellite data logger and VHF antennas breaking the surface of the ocean, the HiCAS tag could be readily located and recovered using a goniometer and a VHF receiver.

2.1.2. Deployments

We deployed tags on leatherbacks in coastal Massachusetts, USA, where the species is regularly observed during summer and autumn (Dodge et al., 2014; James et al., 2006; Lazell, 1980; Shoop and Kenney, 1992). Cape Cod Bay, on the northern side of Cape Cod, Massachusetts, is a coastal embayment with a maximum depth of 63 m. Vineyard Sound and Nantucket Sound are connected basins on the southern side of Cape Cod and are relatively shallow areas (max depth of 26 m) with strong, semidiurnal tidal currents.

An aerial spotter in a small aircraft (Piper Super Cub) was used to help locate leatherbacks, and a 7-m research vessel with a tagging platform modified from Heaslip et al., 2012 was used for tag deployments and retrievals. The tagging platform was fitted to the forward starboard side of the boat and was placed near water level allowing a researcher to lay flat across it and reach a surfacing leatherback. Surfacing turtles were approached and when they were within reach, the boat engines were shifted to neutral and a researcher on the tagging platform placed the tag near the anterior edge of the carapace (Fig. 3) providing a forward-facing view that typically included the turtle's head and a portion of its neck so that it was possible to document breathing and surfacing events in the recorded video. When a tag released from a turtle, we recovered it and downloaded the high-resolution archived data, which we used to create the data streams.

2.2. Data streams

We used data from multiple sensors on the HiCAS tag and four depth thresholds to create nine data streams to explore leatherback surface behavior (Table 1). Two of those data streams were surfacing counts from the human and machine classified video, and one was wet/dry sensor data. The other six data streams were derived from the two depth (pressure) sensors. An approximate time offset between the satellite data logger and the camera data streams was determined by noting the time recorded when the turtle made its first dive. All data streams were trimmed to include only the times when the video showed that the tag was attached to the turtle.

We established four depth thresholds that could be monitored by the HiCAS tag or seen by visual observers in different platforms and environmental conditions. The topmost depth threshold (0.0 m) encompasses only the air directly above the sea-air interface; turtles here are considered to be at the surface. The shallow depth threshold (< 0.5 m) was selected because it encompasses turtles appearing just below the air-sea interface. The middle depth threshold (< 2.0 m) was defined to approximate depths most often used for assessing sea turtle availability to aerial observers (Barco et al., 2018; Fuentes et al., 2015; Hatch et al., 2022; James et al., 2006; Casey et al., 2014). The deepest threshold (< 4.0 m) was included to represent a larger visible layer which could be accurate for some survey platforms, larger animals, and clearer waters (Westgate et al., 2014; Odzer et al., 2022).

2.2.1. Classified video

To provide a best estimate of the time turtles spent at the surface of the water, a human classifier analyzed each frame of the footage using the Observer XT version 12 software package (Noldus Information Technology, Virginia, USA). We developed an ethogram that included codes for when the tag was deployed ("tag on the turtle"), when it detached ("tag off the turtle"), and when it was in the air ("camera at the surface"). We calculated the deployment duration (tag off time minus tag on time). The 'camera at the surface' parameter reflected a binary classification scheme we used to indicate whether the camera (and turtle by proxy) was at or below the surface. See Fig. 4 for a description and examples of the classification rules. Although it was potentially subjective, we considered human-classified video to be our best estimate of surface behavior because the surface could be demarcated well, the video could be reviewed, and it showed some of the least variability in mean percent time at the surface.

To explore the possibility of using automated pattern recognition to achieve efficiencies beyond human coding, we used machine learning to create a second data stream from the videos. The machine learning model used a convolutional neural network (CNN) with eleven layers. A complete description of the model and the software used to create and run it are given in the Appendix. In general, footage was subsampled such that the first frame from each second of video was classified by the machine learning model to create the data stream. To train the machine learning algorithm we applied a supervised learning approach, using

Fig. 3. Hand placement of a HiCAS tag on a leatherback sea turtle. Photo Credit: Joshua Hatch NOAA. Tagging under Endangered Species Act permit (22218).

Table 1
Definitions of the data streams we used to describe leatherback surface behavior. The size of the depth threshold increases from the top to the bottom of the table. The name of each data stream is a concatenation of the Depth Threshold, Sensor Type (HV = human classified video, MV = machine classified video, WD = wet/dry sensor, PR = pressure sensor), and Inst (Instrument; C = camera and S = satellite data logger). For example 0.0-HV-C is the name of the data stream described in the first row of the table.

Depth Threshold (m)	SensorType	Instr	Criteria for defining surfacing events
0.0	HV	C	Human classified video with sky visible in frame
0.0	MV	C	Machine classified video using supervised ML model
0.0	WD	S	Wet/Dry sensor data >80
0.5	PR	C	Zero offset corrected depth data <0.5 m
0.5	PR	S	Zero offset corrected depth data <0.5 m
2.0	PR	C	Zero offset corrected depth data <2.0 m
2.0	PR	S	Zero offset corrected depth data <2.0 m
4.0	PR	C	Zero offset corrected depth data <4.0 m
4.0	PR	S	Zero offset corrected depth data <4.0 m

small video sequences of "at surface" or "below surface" verified by a human classifier. Additional sequences were added until the training validation loss was consistently below 0.1 with 5489 video frames.

The relationship between the machine-classified video and the human-classified video was examined by summing the number of video frames classified as "at surface" by the human coder for each turtle and plotting them against the sum of machine-classified surface frames.

2.2.2. Wet/dry sensor (0.0-WD-S)

The wet/dry sensor on the satellite data loggers was designed to determine if the sensor is in or out of the water (Wildlife Computers, 2021). Once per second, the satellite data logger recorded a wet/dry value that ranged between 0 and 255. A tag fully submerged in high salinity seawater has maximum conductivity and wet/dry sensor readings near 20; while a dry tag unable to measure any conductivity has wet/dry sensor readings near 255 (Wildlife Computers, 2019). As a tag emerges out of the water, a film of water can create residual conductivity even when the tag is in the air. Conversely if bubbles are created as the turtle dives, they can cause the sensor to read drier. To separate these sensor readings into wet and dry categories, we used the same threshold value that Wildlife Computers uses for the initial setting in their dynamic threshold calculations; all wet/dry sensor values <80 were classified as "below the surface" and values greater than or equal to 80 as "at the surface".

2.2.3. Time depth records (0.5-PR-C, 0.5-PR-S, 2.0-PR-C, 2.0-PR-S, 4.0-PR-C, 4.0-PR-S)

At one second intervals during deployments, independent sensors in the camera and the satellite data logger recorded time and depth. The clock on the camera depth sensor had to be set manually, while the satellite data logger clock was set from the satellite. The resolution of the camera depth sensor on the camera was 0.1 m, and the satellite data logger was 0.5 m. Because pressure transducers can drift over time and cause recorded depths to vary from actual depths (Luque and Fried, 2011, Heide-Jørgensen and Lage, 2022), both depth data streams were calibrated using zero offset correction methods. The satellite data logger calculated a "Corrected Depth" automatically by applying a zero offset correction (based on recorded depths when the tag is dry) to its depth sensor readings. The metadata for the camera did not describe any calibrations to address pressure sensor drift, and we knew (from the associated video) that the camera depth recorder was not reporting zero depth at the surface, so we applied a zero-offset correction to the camera depth data. We set the zero-offset correction for each turtle to the value of the second percentile of the dive depth data for that turtle. We selected the second percentile rather than the smallest value to discount outliers. We checked the validity of the zero-offset correction method (for the camera) by applying this same approach to the uncorrected satellite data logger data. When the zero-offset was applied to uncorrected satellite data logger depth values the result resembled the corrected data as calculated by the satellite data logger itself. We used

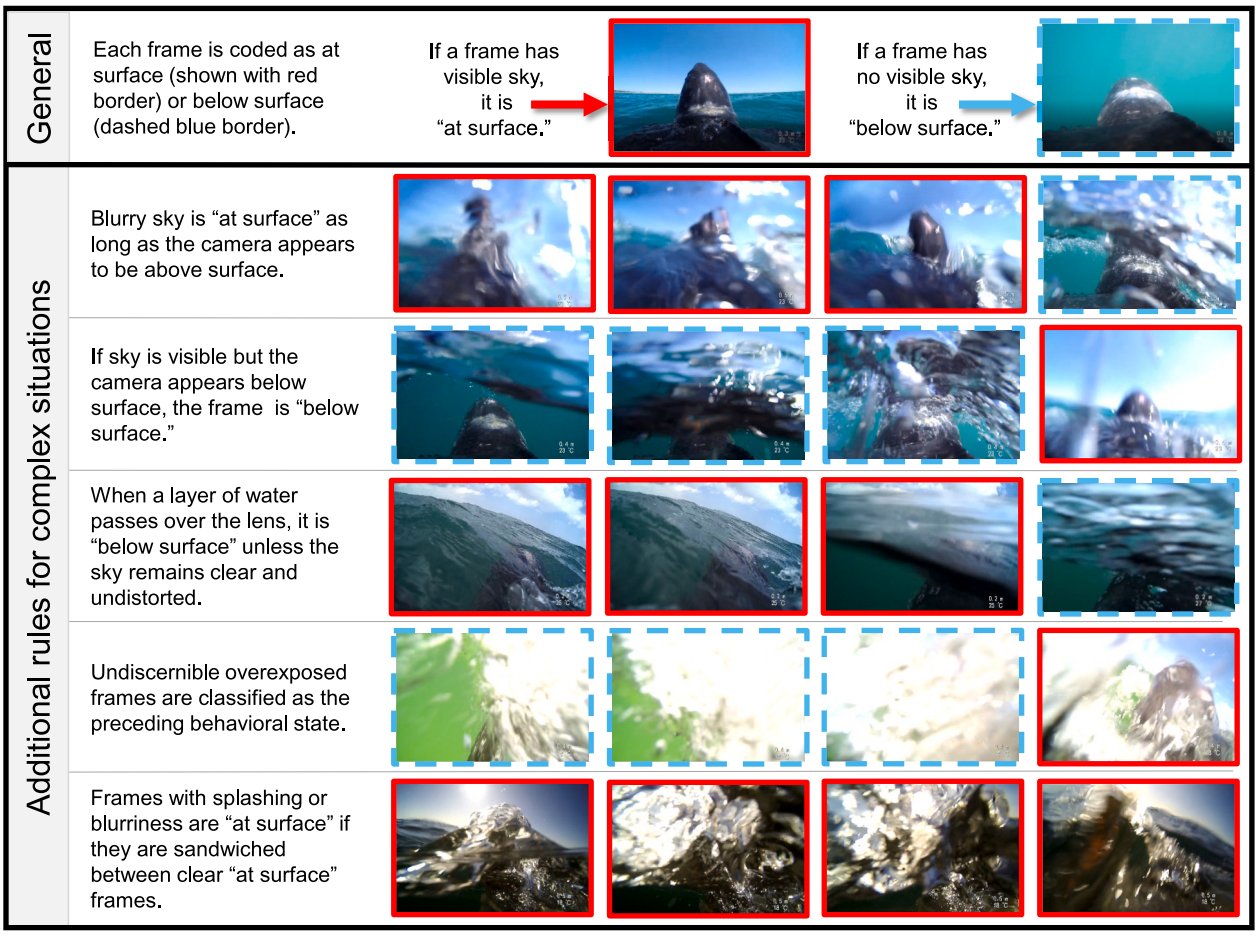

Fig. 4. Video Classification Rules. In these examples, the solid red outlines around each frame indicate when the frame was classified as 'at surface', and the dashed blue outlines indicate when the frame was classified as 'below surface'. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

corrected depth data in all of our calculations.

2.3. Behavior metrics

We calculated four metrics to describe surface behavior for each of the nine data streams. We focused on the weighted mean percent time at surface (as a percent of total tag duration for each turtle) because it can be calculated from a variety of sensors (video, wet/dry sensor, and depth sensors), and aligns with other research on sea turtle availability (Seminoff et al., 2014; Fuentes et al., 2015; Barco et al., 2018; Hatch et al., 2022). In order to provide more ecological information and supply data for more complex availability analysis (Laake et al., 1997), we also calculated the weighted mean surface duration and weighted mean dive duration, as well as the standard error of those metrics. Mean values were calculated for each turtle, and a weighted mean calculated from those means, weighted by deployment durations. In addition, we calculated the number of surfacings per hour for each turtle to better understand relationships between event durations and frequency. The four definitions of "surface" were explored for each of the behavioral metrics as described in Table 1, and "diving" was considered "any time the turtle was not at the surface", using the appropriate definition of "surface".

To better illustrate how individual turtles used surface waters, we calculated the percent of time each turtle spent in each of the shallowest 5 m, as well as the percent time deeper than 5 m. We used the depth data from the camera because it had the highest resolution and was available for all but one of the turtles, and we selected the 5 m threshold to

Table 2
Description of HiCAS Tag Deployments. "ID" refers to the identifier for each turtle. "S" indicates whether the satellite data logger was present (Y = yes, N = no). "Tag on" refers to the time of day (local time) when the HiCAS tag was placed on the turtle. "Deployment Duration" refers to the length of the video (minutes) while the tag was on the turtle. "Offset" refers to the depth adjustment (meters) applied to correct the camera depth readings.

ID	S	Date	Tag on Time	DeploymentDuration	Offset
A	N	27-Aug-2019	11:29	198.8	0.4
B	N	10-Sep-2019	12:24	37.1	0.3
C	N	15-Sep-2019	11:37	184.4	0.3
D	Y	21-Sep-2019	10:54	95.7	NA
E	Y	30-Sep-2019	11:00	32.8	0.3
F	N	5-Oct-2019	14:33	157.3	0.5
G	Y	5-Oct-2019	15:08	57.5	0.3
H	Y	25-Oct-2019	10:45	130.5	0.2
I	N	25-Oct-2019	10:54	188.4	0.5
J	N	25-Oct-2019	11:54	156.0	0.3
K	Y	24-Sep-2020	10:02	170.7	0.3
L	Y	4-Oct-2020	10:43	118.6	0.4
M	Y	4-Oct-2020	11:45	119.7	0.5
N	Y	19-Oct-2020	9:51	179.5	0.5
O	Y	19-Oct-2020	10:18	171.2	0.4
P	Y	19-Oct-2020	11:09	161.9	0.1
Q	Y	22-Oct-2020	15:36	109.6	0.6
R	Y	23-Oct-2020	13:16	184.2	0.4

provide a buffer beyond the depths that are typically visible to aerial observers.

3. Results

3.1. HiCAS tag deployments

Between August and October of 2019 and 2020, we deployed the HiCAS tag on 18 leatherback turtles (Table 2). Deployments occurred during daylight hours, with a mean deployment duration of 136.1 min and a range of 32.8–198.9 min.

3.2. Data streams

Most deployments ($n = 12$) had both the camera and the satellite data logger, and therefore yielded all nine expected data streams (Table 1). Five deployments had only the camera (video and TDR functionality), and one deployment (Turtle D) had the camera with only video functionality. After trimming and aligning to the time of the turtle's first dive, there was general agreement of surfacing events across sensors, however there was often several seconds of offset between sensors. An example of the combined results for one turtle (Turtle Q) is shown in Fig. 5.

3.2.1. Classified video

The video frames for all 18 turtles were classified as "at surface" or "below surface" by the human classifier for the duration that the tags were on the turtles. The overall time invested in this analysis by the human classifier totaled approximately 175 h, with each hour of video footage taking a mean time of four hours to classify. The machine learning model and utility code were coded by one developer working part time over six months (altogether ~600 lines of Python code), including the labeling of 4591 training video frames. The model was then trained once, overnight, unattended, as detailed in the Appendix. It was used for all turtles in this study and should not need to be retrained for future video classification. It classified the same video as the human classifier in ~14 h, again unattended. In most cases the machine-classified results were close to the human-classified results (Fig. 6). One turtle (H) had substantially more machine-classified surface images than human-classified surface images. The source video for turtle H showed consistent images of the sun shining through the water in the frames that were classified differently by human and machine. We suspect the angle of the sun caused the machine learning algorithm to classify the frame as "at the surface", even though the video and the corresponding depth records showed the turtle was clearly below the surface. (Appendix Fig. A.3. Example Frame Misclassified as At Surface).

3.2.2. Wet/dry sensor

The distribution of the wet/dry data was bimodal, with data clustered near wet (20) and dry values (250). See Supplemental Fig. S2 for more details.

3.2.3. Time depth records

The time and depth records indicated when the sea turtle was at the surface or below, based on the pressure (depth) readings from the camera and satellite tag sensors. Setting the zero-offset correction (for data streams from the camera) to the value of the second percentile of the dive depth data for each deployment resulted in depth records that more accurately represented known turtle behavior (see Fig. 5 showing both corrected and uncorrected depth data). The zero offset correction appropriately resulted in tags approaching the surface (where depth = 0), but not rising above it; whereas the uncorrected depth data had tags inappropriately rising above the surface of the water (where depth > 0). These corrected depth records are the basis of the values developed in section 3.3 Behavior Metrics.

3.3. Behavior metrics

The mean percent time at surface was the only behavior metric that changed systematically as data streams incorporated deeper thresholds (Table 3). The metric increased as the size of the depth threshold increased, which is what we would expect if all sensors were working correctly. The video-based data streams (0.0-HV-C, 0.0-MV-C) estimated the lowest mean percent time at surface at just over 7% (Table 3). When

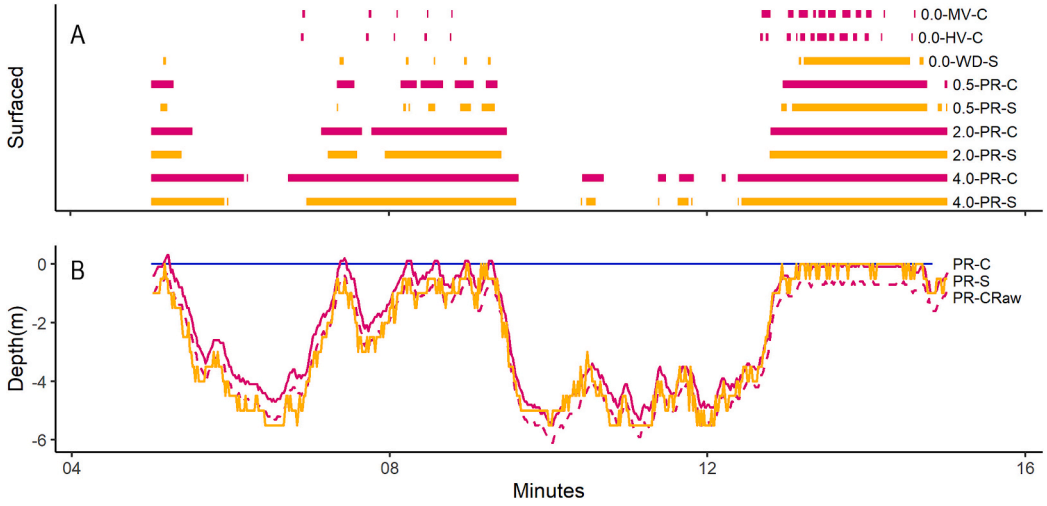

Fig. 5. Turtle Q Dive Trace. To show detail, only a subset of the deployment is shown (from 4 to 16 min). Panel A shows the 9 data streams that have been classified into surface and non-surface behavior. The markings indicate when the turtle was at the surface or the indicated near-surface depth, and the absence of markings indicate when the turtle was at or below the threshold depth. Descriptions of the variables are shown in Table 1. We did not further synchronize the time between the video and pressure sensors data streams as the offset does not affect our main focus (mean percent time as the surface). Panel B shows the corresponding depth readings. The solid orange line shows depth data from the satellite data logger (PR-S) and the solid magenta line shows corrected depth data from the camera (PR-C). The dashed magenta line shows uncorrected depth data from the camera (PR-CRaw). The horizontal light blue line marks the surface of the water (zero depth). The need for camera depth data correction can be seen, as without it, the turtle never appears to rise to the surface. Values from the camera tag are shown in magenta, values from the satellite data logger are in orange. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

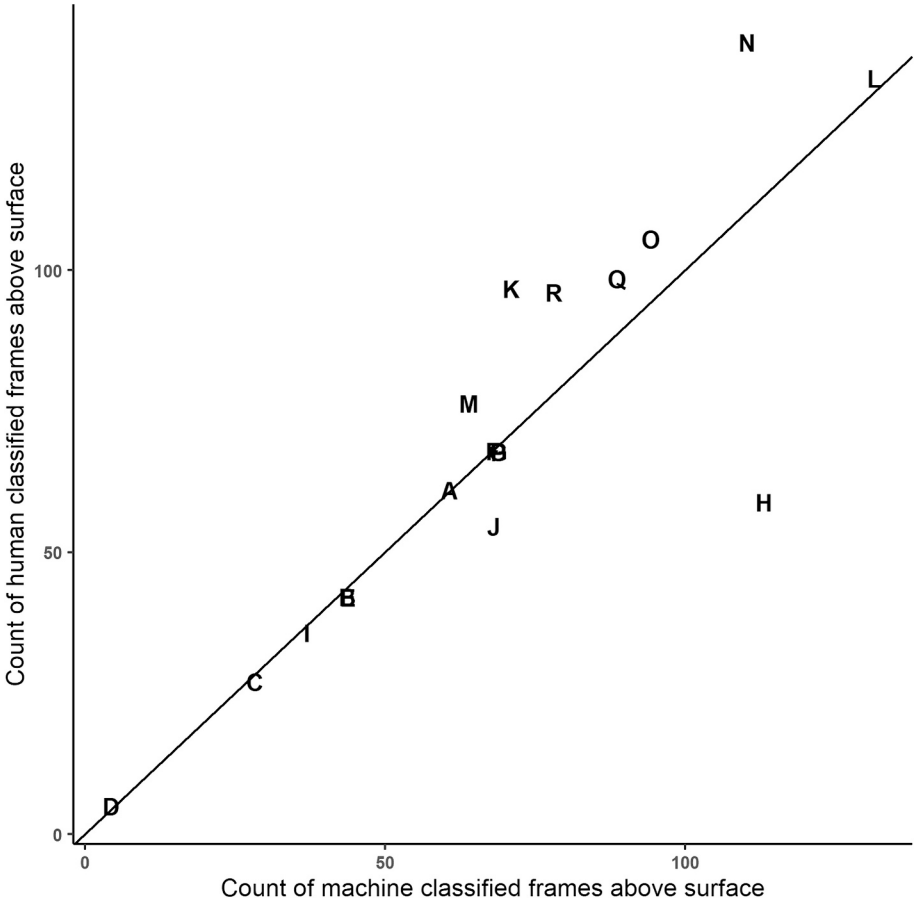

Fig. 6. Comparison of the number of frames classified as at the surface by manual analysis and machine classification. Letters indicate the tag deployment. The solid line indicates when the human and machine classifications agreed as to the number of frames above the surface.

Table 3
Results across all turtles for each data stream: mean surface duration in seconds (mSDur), mean dive duration in seconds (mDDur), number of surfacings per hour (nSurfs/h), and mean percent time at the surface (mPTaS). Note that the size of the depth threshold increased from the top to the bottom of the table.

Data	Behavior Metrics			
Streams	mSDur	mDDur	nSurfs/h	mPTaS
0.0-HV-C	3.6	44.2	72.3	7.5
0.0-MV-C	3.7	47.0	69.1	7.4
0.0-WD-S	17.9	93.3	19.8	15.1
0.5-PR-C	13.4	53.0	48.0	19.6
0.5-PR-S	12.9	52.8	36.9	19.3
2.0-PR-C	71.3	99.9	21.1	41.1
2.0-PR-S	23.2	38.1	34.5	37.7
4.0-PR-C	97.2	58.9	22.0	53.3
4.0-PR-S	38.2	25.3	34.1	49.7

using the wet/dry sensor, this metric doubled to about 15% (Table 3). The depth-based data streams from the camera and satellite data logger produced similar estimates as each other. Using data streams from the 0.5 m surface layer (0.5-PR-C and 0.5-PR-S) increased mean percent time at surface to almost 20%. Using a 2 m surface layer (2.0-PR-C and 2.0-PR-S) nearly doubled the mean percent time at surface to roughly 40%, and using a 4 m surface layer (4.0-PR-C and 4.0-PR-S) increased it to roughly 50% (Table 3). The variability in this metric increased in data streams which incorporated deeper surface layers (Fig. 7, Supplement Table S.4.).

Within a sensor, the mean surface duration roughly increased as the depth threshold increased, however there were minor inconsistencies in the trend and substantial inconsistencies between sensors (Table 3). The mean surface durations from the video-based data streams (0.0-HV-C, 0.0-MV-C) were shortest (approx. 4 s), and the surface duration from the wet/dry sensor was larger (approx. 18 s) as expected, but the mean surface duration from the 0.5 m surface layer (0.5-PR-C and 0.5-PR-S) were intermediary (approx. 13 s) in between the two shallower depth thresholds. The mean surface duration for the >2 m depth threshold was higher than shallower thresholds, but the values from the different sensors were not similar (2.0-PR-C approx. 71 s; 2.0-PR-S approx.23 s). The surface durations for the deepest thresholds were the largest calculated from each sensor, but again the values from the different sensors were dissimilar (4.0-PR-C approx. 97 s; 4.0-PR-S approx. 38 s). As the thresholds were defined more deeply, there was more variability in mean surface duration (Fig. 7). Many of these trends were also evident at the individual turtle level (Fig. S.2) with more variability in individual values, as expected.

The mean dive duration did not show a consistent trend with respect to increasing depth threshold (Table 3). The satellite data logger TDR showed shorter mean dive durations at deeper depth thresholds, but notably the mean dive duration of the camera TDR at a depth threshold of 0.5 m (0.5-PR-C approx. 53 s) is shorter than the mean dive duration of the camera TDR at 2.0 m (2.0-PR-C approx 100 s). These complex patterns may result from a combination of factors. The satellite data logger TDR has lower resolution than the camera TDR which may explain why the estimates derived from these two sensors differ. The inconsistent pattern within the camera TDR data occurs because shallow depth thresholds consider a longer portion of a dive trace as part of the dive (which increases the overall dive durations); yet, shallow depth thresholds also produce more surfacings per hour, resulting in more

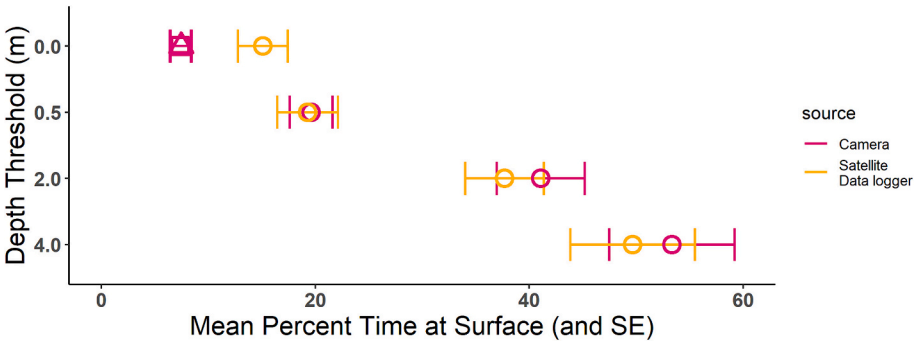

Fig. 7. Mean percent time at surface across all turtles, for each depth threshold. Values from camera (magenta) and satellite data logger (orange) sensors are shown. Human classification results are shown as a triangle, machine learning results are shown as a square, and results derived from depth sensor readings are shown as circles. The error bars indicate the high and low standard errors for each value. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

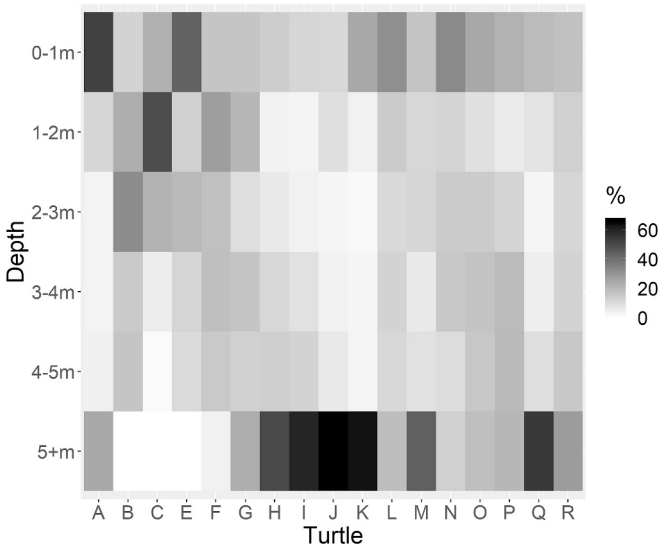

Fig. 8. Percent Time in depth range. The percentage of total time that each turtle spent in the indicated depth ranges (m), using the camera depth data. Camera depth data was not recovered for Turtle D.

short dives (which decreases the mean dive duration). This inconsistent pattern can also be seen in the TDR data from the camera tag shown in Fig. 8 where the intermediate (2 m) depth threshold produced the longest mean dive duration (0.5-PR-C = 82 s, 2.0-PR-C = 129 s, and 4.0-PR-C = 41.7 s, for the subset of data shown in Fig. 8).

The number of surfacings per hour across deployments was highest in the video-based data streams (roughly 70 surfacings per hour) and lowest in the wet/dry data stream (0.0-WD-S around 20). The rest of the data streams were inconsistent in the relationship between the number of surfacings and the considered depth of the surface layer (Table 3). We did not present the total number of dives per hour because they were nearly identical to the number of surfacings per hour (due to our methodology).

Variability in the values of the behavior metrics tended to increase as depth threshold increased. For example, the standard error of the data streams quantifying the mean percent time at the air-water interface (0.0-HV-C, 0.0-MV-C, and 0.0-WD-S) were approx 1 to 2 s, whereas the standard error for the mean percent time < 2 m and < 4 m ranged from 4 to 6 s (Fig. 7 and Supplement Table S.4).

At the individual turtle level, there was variability in the pattern of depth use within the upper 5 m (Fig. 8). Some turtles (H, I, J, K, Q) spent considerable time below the top 5 m, and in these cases, the surface behavior tended to be more concentrated in the upper 0.0–1.0 m. Some turtles (L, O, P, R) tended to have a more uniform distribution across all the depth thresholds. There were three turtles (B, C, E) that spent nearly all of their time in the top 5 m, sometimes spending substantial time in the 1.0–2.0 m or 2.0–3.0 m thresholds.

4. Discussion

The modular design of the HiCAS tags allowed for multiple data streams to be collected and combined to better understand leatherback surfacing behavior and how it is measured by different sensors. Ideally surface behavior would be monitored by high resolution sensors that report behavior every second over a long time period, but from a practical standpoint, most surface behavior estimates are based either on high resolution data from short-term tags, or based on summarized data from long-term satellite tags (though see Heide-Jørgensen and Lage, 2022 for a nice comparison of the two approaches). By integrating the satellite data logger with other sensors, we were able to record and analyze different data streams, including those typically used to assess sea turtle surface behavior and calculate availability estimates (Hatch et al., 2022; Barco et al., 2018; Fuentes et al., 2015; Roberts et al., 2022). Estimated surface behavior varied based on tag sensor and the definition of the depth threshold.

When comparing the human- and machine-classified video, we found high similarity in the results. This provides confidence that the machine learning algorithm can be used in future video sets with smaller amounts of human classifications as ground truthing. The high correlation between human- and machine-classified video suggests that machine learning can be an effective replacement for human classification, at least for simple tasks, such as whether the video camera is underwater or not. We avoided coding the video footage for when the turtle was at other positions in the water column due to the differences in lighting and camera position between each video. Although it would have been useful to have a human-classified version of near surface position equivalent to the 2.0 m or 4.0 m thresholds, it was not possible to consistently tell where in the water column the turtle was when below the surface. Conveniently, the camera used in the HiCAS tag does embed the depth reading in the video frame, however, as seen in Fig. S1, this can be off by 0.5 m or more due to common pressure sensor limitations like drift, time delays, and poor resolution at near-surface depths. Image classification is an active research area in the field of machine learning, and there are multiple algorithms that could be used (Chen et al., 2021). There is an opportunity to explore the use of different (possibly simpler) algorithms in the future, as well as trying different combinations of algorithm parameters.

Compared to the other data streams, we have more confidence in the video-based data streams (0.0-HV-C and 0.0-MV-C) because these have robust methodology resulting from an easy to demarcate depth threshold, frequent sampling rates, and high information content in each

visual sample. The definition of the depth threshold was easy to demarcate by limiting the surfacings to only those instances when a turtle broke through the surface of the water and into the air. The video format provided 30 frames per second, and each high resolution (1920 × 1080) frame could be viewed as a still image. Although human-classified video is labor intensive, it is thought to have a low error rate (Fleuret et al., 2011). The video-based data streams in this analysis also had the benefit of having a relatively low level of variation between turtles. The variation in the mean surface duration and mean percent time at surface from the video assessments was lower than for any other data stream.

The wet/dry sensor also measured behavior with respect to the water-air interface, and had the advantage of being easily transmitted from standard satellite data loggers; but it has notable shortcomings. While the mean percent time at surface from the wet/dry sensor was the closest to the video data streams, it was more than twice the estimate (0.0-HV-C and 0.0-MV-C ~ 7%; 0.0-WD-S ~ 15%). Despite measuring behavior in the same theoretical surface layer as the video methods, the wet/dry sensor differed substantially with respect to the duration and number of surfacings. This may be related to the cut off value used to determine wet vs dry (as it is not a binary reading) or to a delay in the switch between wet and dry when the tag comes to the surface and resubmerges.

When aggregated across turtles, only the mean percent time at surface had systematic trends that aligned well with a priori expectations. This metric increased as the depth threshold increased, and there was similarity between the data streams coming from the camera and satellite data logger depth sensors. The rest of the behavior metrics (mean surface duration, mean dive duration, and surfacings per hour) had no consistent trend making their validity as meaningful metrics for estimating surface availability questionable. For all three metrics, the two depth sensors (data logger and camera) produced dissimilar results and there was no consistent increase in value corresponding to the increase in depth threshold. Neither the camera nor the satellite data logger TDR produced behavior metrics that were reasonably close to the metrics calculated from the video data streams. Although the satellite data logger had a lower depth resolution than the camera TDR, the surface durations and number of surfacing events from the satellite data logger were more similar to the estimates from the video data streams. Importantly, metrics related to the duration and number of surfacings appear to be highly sensitive to the ways that surfacings are lumped or split. This likely skews the perception of near-surface behavior from studies relying only on satellite transmitters for data, particularly when many short consecutive surfacings may be lumped into one longer continuous surfacing. These differences highlight the importance of carefully selecting appropriate behavior metrics and matching the tag sensors and the definition of surface layer to the applied need for behavioral information.

The variability in the patterns of depth use within the upper 5 m (Fig. 8) illustrates the behavioral differences between turtles, and it suggests that extra caution is warranted when interpreting surface behavior data, particularly when sample sizes are small or when there is unaddressed spatial or temporal structure. When the leatherbacks in this study spent substantial time in deep water (> 5 m), their occupation of the 1–4 m depth threshold was minimal, perhaps indicating that these leatherbacks were only transiting through the 2–4 m portion of the water column between breathing at the surface and foraging at depth. In these cases, the calculation of the mean percent time at surface is not particularly sensitive to how the surface layer is defined. However, turtles B, C, and E were located in a shallow area of Cape Cod Bay where water depths are around 5 m, and, therefore, had less available habitat to be at our deepest depth threshold. Consequently, our data show they spent little time below 5 m and exhibited substantial variability in the pattern of depth use across the upper 3 m. For these turtles, the calculation of the mean percent time at surface is highly sensitive to how the depth threshold is defined. As a simplified example, if the surface abundance in a given area is 100 turtles, and the availability bias correction was based on turtles behaving similar to Turtle B (which spent little time below 5 m), the corrected abundance estimate using a 3 m depth threshold (surface availability = 0.6827) would be ~146, but it would increase to ~785 when using a 1 m depth threshold (surface availability = 0.1274). Large sample sizes of randomly selected animals can appropriately amalgamate these various types of turtle behavior when abundance estimates are not spatially or temporally explicit. However, when abundance estimates are spatially or temporally explicit, accurate abundance estimates will only be expected if animal behavior is sampled throughout and in proportion to the spatial and temporal structuring. This highlights the need to consider behavioral and environmental variation when defining the depth threshold and setting sample size determinations for calculating availability bias corrections.

In the context of availability bias estimates for line transect surveys, the trend we documented of increasing variability in animal behavior with increasing depths of the surface layer is further confounded by the uncertainty of how deep a visual observer can see at any given moment. When aerial survey programs record turtles that are seen at depth (in addition to turtles that are at the surface), the programs benefit by avoiding the complication of discerning whether or not the turtle is at the surface and by increasing the count of turtles available for analysis. Recording turtles seen at depth, however, also introduces additional complexity and sources of variation, many of which rarely get propagated into subsequent abundance estimates. The differences we documented in surface metrics at various depth thresholds provide line transect survey analysts with realistic estimates of the magnitude of error that would be introduced if the depth thresholds used in surface behavior analysis do not match the visible depth threshold during the observation period. Because mean percent time at surface increases (and the range of observed values increases) as the depth threshold increases, an ideal survey would match the depth thresholds used in the surface behavior analysis to the actual depths visible to survey observers. Unfortunately, this is extremely difficult to do in practice because the depth of the visible threshold is difficult to measure and highly variable in space and time (Barco et al., 2018).

In cases where estimates of surface behavior are likely to have influential depth-related errors (such as when spatially and temporally explicit abundance estimates are needed across diverse environmental or behavioral zones), it may be useful for survey programs to explore the consequences of focusing on only turtles that are at the surface. This approach would avoid many of the depth related errors associated with the calculations of availability bias, but would also decrease the survey's sample size because turtles in the visible subsurface zone would not be included. A trial investigation of this approach could additionally log whether the animals are at the surface when sighted, and these data together with turtle behavioral data could be used to build simulations to evaluate the trade offs between decreases in the sample size and increases in precision that may result from only counting turtles at the surface (Sequeira et al., 2019). This type of research into the feasibility, merit, and shortcomings of producing abundance estimates and availability corrections from only animals at the surface can be used to refine future line transect surveys and animal tagging efforts.

Although mean percent time at surface (Fig. 7) represents useful information on the behavior of leatherbacks within our study region, we caution against applying our estimates of surface behavior as availability corrections for line-transect surveys, particularly in different times and areas. Our studies were made in daylight hours, mostly around mid-day, for short periods (usually <3 h), in a limited geographic area (Cape Cod Bay, Vineyard Sound, and Nantucket Sound), during a limited timeframe (late August through September of two years). Our results represent a small study area which may not be typical of offshore areas with high survey effort. The mean percent time at surface that we observed in this small-scale study are not unusual in comparison to other Atlantic leatherback surface behavior studies (Rider et al., n.d.), but the

populations that comprise the global distribution of this endangered species are known to have varied diving and surfacing behaviors so care must be taken to include spatial and temporal effects (Migneault et al., 2023; Rider et al., 2022; Casey et al., 2014; James et al., 2006; Wallace et al., 2015; Dodge et al., 2014). Additional studies performed in areas that are part of migration corridors and known foraging areas can add critical information.

This study provides important insights into methods that underlie marine protected species assessments. We used multiple data streams to illustrate how our understanding of surface behavior varies depending on the definition of the depth threshold and the sensors used to measure depth. We also demonstrated that machine learning can offload laborious tasks, such as human coding of recorded video. Animal surfacing data, such as those we present, can be highly influential to abundance estimates from line transect surveys (Barco et al., 2018; Fuentes et al., 2015). The methods for both satellite tagging and line transect survey analysis are well established, but research into the intersection of these fields is far less developed, especially with respect to fully incorporating variability and uncertainty. This study is a small step towards addressing these needs. Future research is needed to better understand the variability in animal behavior and the optimal ways to incorporate this information into line transect survey analysis. The collection and interpretation of animal behavior data is complex, and it is further complicated by anthropogenic development and climate change which are expected to modify animal habitat and behavior (Hawkes et al., 2009; Beever et al., 2017). Hence, in addition to methodological research, sustained data collection is needed to avoid the pitfalls associated with applying behavior data outside of its range. We caution against treating animal surfacing behavior as a simple, uniform, or static process.

CRediT authorship contribution statement

Rick Rogers: Writing – review & editing, Writing – original draft, Visualization, Software, Conceptualization. **Kate H. Choate:** Writing – review & editing, Visualization, Methodology. **Leah M. Crowe:** Writing – review & editing, Methodology, Investigation. **Joshua M. Hatch:** Writing – review & editing, Visualization, Methodology. **Michael C. James:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Conceptualization. **Eric Matzen:** Methodology. **Samir H. Patel:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Conceptualization. **Christopher R. Sasso:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Conceptualization. **Liese A. Siemann:** Writing – review & editing, Methodology, Investigation. **Heather L. Haas:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We thank Bailey Buckley for preliminary human video coding, George Breen and Rick Brown for aerial support, Lisa Conger for small boat support, Farrell Davis for building the GPS units, Michael Casso for technical advice and assistance in machining the foam bases, and Doug Sigourney for technical review of the manuscript. We also thank the anonymous reviewers of the manuscript for their contributions in improving it. This study was funded in part by the U.S. Department of the Interior, Bureau of Ocean Energy Management through Interagency Agreements M10PG00075, M14PG00005, and M19PG00007 with the U. S. Department of the Commerce, National Oceanic and Atmospheric Administration (NOAA), National Marine Fisheries Service (NMFS), Northeast Fisheries Science Center (NEFSC), and through the NOAA Bycatch Reduction and Engineering Program grant #NA17NMF4720268 awarded to Coonamessett Farm Foundation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jembe.2024.152012.

References

Archibald, D.W., James, M.C., 2016. Evaluating inter-annual relative abundance of leatherback sea turtles in Atlantic Canada. Mar. Ecol. Prog. Ser. 547, 233–246.
Barco, S.G., Burt, M.L., DiGiovanni Jr., R.A., Swingle, W.M., Williard, A.S., 2018. Loggerhead turtle Caretta caretta density and abundance in Chesapeake Bay and the temperate ocean waters of the southern portion of the mid-Atlantic bight. Endanger. Species Res. 37, 269–287.
Beever, E.A., Hall, L.E., Varner, J., Loosen, A.E., Dunham, J.B., Gahl, M.K., Lawler, J.J., 2017. Behavioral flexibility as a mechanism for coping with climate change. Front. Ecol. Environ. 15 (6), 299–308.
Benson, S.R., Forney, K.A., Harvey, J.T., Carretta, J.V., Dutton, P.H., 2007. Abundance, distribution, and habitat of leatherback turtles (*Dermochelys coriacea*) off California, 1990–2003. Fish. Bull. 105 (3), 337–347.
Benson, S.R., Forney, K.A., Moore, J.E., LaCasella, E.L., Harvey, J.T., Carretta, J.V., 2020. A long-term decline in the abundance of endangered leatherback turtles, Dermochelys coriacea, at a foraging ground in the California Current Ecosystem. Global Ecol. Conserv. 24, e01371. ISSN 2351-9894.
Borchers, D.L., Zucchini, W., Heide-Jørgensen, M.P., Cañadas, A., Langrock, R., 2013. Using hidden Markov models to deal with availability bias on line transect surveys. Biom 69, 703–713. https://doi.org/10.1111/biom.12049.
Casey, J.P., James, M.C., Williard, A.S., 2014. Behavioral and metabolic contributions to thermoregulation in freely swimming leatherback turtles at high latitudes. J. Exp. Biol. 217 (13), 2331–2337.
Ceriani, S.A., Casale, P., Brost, M., Leone, E.H., Witherington, B.E., 2019. Conservation implications of sea turtle nesting trends: elusive recovery of a globally important loggerhead population. Ecosphere 10 (11), e02936. https://doi.org/10.1002/ecs2.2936.
Chen, L., Li, S., Bai, Q., Yang, J., Jiang, S., Miao, Y., 2021. Review of image classification algorithms based on convolutional neural networks. Remote Sens. 13 (22), 4712. https://doi.org/10.3390/rs13224712.
DiMatteo, A., Roberts, J.J., Jones, D., Garrison, L., Hart, K.M., Kenney, R.D., McLellan, W.A., Lomac-MacNair, K., Palka, D., Rickard, M.E., Roberts, K.E., 2024. Sea turtle density surface models along the United States Atlantic coast. Endanger. Species Res. 53, 227–245.
Dodge, K.L., Galuardi, B., Miller, T.J., Lutcavage, M.E., 2014. Leatherback turtle movements, dive behavior, and habitat characteristics in ecoregions of the Northwest Atlantic Ocean. PLoS One 9 (3), e91726. https://doi.org/10.1371/journal.pone.0091726.
Dunstan, A., Robertson, K., Fitzpatrick, R., Pickford, J., Meager, J., 2020. Use of unmanned aerial vehicles (UAVs) for mark-resight nesting population estimation of adult female green sea turtles at Raine Island. PLoS One 15 (6), e0228524. https://doi.org/10.1371/journal.pone.0228524.
Fleuret, F., Li, T., Dubout, C., Wampler, E.K., Yantis, S., Geman, D., 2011. Comparing machines and humans on a visual categorization test. Proc. Natl. Acad. Sci. USA 108 (43), 17621–17625. https://doi.org/10.1073/pnas.1109168108. Epub 2011 Oct 17. PMID: 22006295; PMCID: PMC3203755.
Fuentes, M.M.P.B., Bell, I., Hagihara, R., Hamann, M., Hazel, J., Huth, A., Seminoff, J.A., Sobtzick, S., Marsh, H., 2015. Improving in-water estimates of marine turtle abundance by adjusting aerial survey counts for perception and availability biases. J. Exp. Mar. Biol. Ecol. 471, 77–83.
Hatch, J.M., Haas, H.L., Sasso, C.R., Patel, S.H., Smolowitz, R.J., 2022. Estimating the complex patterns of survey availability for loggerhead turtles. J. Wildl. Manag. 86, e22208 https://doi.org/10.1002/jwmg.22208.
Hawkes, L.A., Broderick, A.C., Godfrey, M.H., Godley, B.J., 2009. Climate change and marine turtles. Endanger. Species Res. 7 (2), 137–154.
Heaslip, S.G., Iverson, S.J., Bowen, W.D., James, M.C., 2012. Jellyfish support high energy intake of leatherback sea turtles (*Dermochelys coriacea*): video evidence from animal-borne cameras. PLoS One 7 (3), e33259. https://doi.org/10.1371/journal.pone.0033259 (Epub 2012 Mar 16. Erratum in: PLoS One. 2012;7(6). doi:10.1371/annotation/4722cac5-8305-4b03-b805-a59ced1eea49).
Heide-Jørgensen, M.P., Lage, J., 2022. On the Availability bias in Narwhal Abundance Estimates. NAMMCO Scientific Publications, p. 12.
Heide-Jørgensen, M.P., Laidre, K.L., 2015. Surfacing time, availability bias and abundance of humpback whales in West Greenland. Journal of Cetacean Research and Management 15, 1–8.

James, M.C., Ottensmeyer, C.A., Eckert, S.A., Myers, R.A., 2006. Changes in diel diving patterns accompany shifts between northern foraging and southward migration in leatherback turtles. Can. J. Zool. 84 (5), 754–765.
Laake, J.L., Calambokidis, J., Osmek, S.D., Rugh, D.J., 1997. Probability of detecting harbor porpoise from aerial surveys: estimating g(0). J. Wildlife Manage. 61 (1), 63–75 (Published by: Allen Press).
Lazell, J.D., 1980. New England waters: critical habitat for marine turtles. Copeia 1980 (2), 290–295. https://doi.org/10.2307/1444006.
Letessier, T.B., Bouchet, P.J., Reisser, J., Meeuwig, J.J., 2014. Baited videography reveals remote foraging and migration behaviour of sea turtles. Mar. Biodivers. https://doi.org/10.1007/s12526-014-0287-3.
Luque, S.P., Fried, R., 2011. Recursive filtering for zero offset correction of diving depth time series with gnu r package dive move. PLoS One 6 (1), e15850.
Marsh, H., Sinclair, D.F., 1989. Correcting for visibility Bias in strip transect aerial surveys of aquatic Fauna. J. Wildl. Manag. 53 (4), 1017–1024. https://doi.org/10.2307/3809604.
Migneault, A., Bennison, A., Doyle, T.K., James, M.C., 2023. High-resolution diving data collected from foraging area reveal that leatherback turtles dive faster to forage longer. Ecosphere 14 (8), e4576. https://doi.org/10.1002/ecs2.4576.
National Marine Fisheries Service, and U.S. Fish and Wildlife Service, 1992. Recovery Plan for Leatherback Turtles in the U.S. Caribbean, Atlantic, and Gulf of Mexico. National Marine Fisheries Service Office of Protected Resources and U.S. Fish and Wildlife Service, Washington, D.C.
National Marine Fisheries Service, and U.S. Fish and Wildlife Service, 2008. Recovery plan for the Northwest Atlantic Population of the Loggerhead Sea Turtle (*Caretta caretta*), Second revision. National Marine Fisheries Service Office of protected resources and U.S. Fish and Wildlife Service, Washington, D.C.
National Marine Fisheries Service, and U.S. Fish and Wildlife Service, 2020. Endangered Species Act status review of the leatherback turtle (*Dermochelys coriacea*). National Marine Fisheries Service Office of Protected Resources and U.S. Fish and Wildlife Service, Washington, D.C.
Preliminary Summer 2010 Regional Abundance Estimate of Loggerhead Turtles (*Caretta caretta*) in Northwestern Atlantic Ocean Continental Shelf Waters. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 11-03; 33, 2011. Northeast Fisheries Science Center. Available from: National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, or online at. http://www.nefsc.noaa.gov/nefsc/publications/.
Odzer, M.N., Brooks, A.M., Heithaus, M.R., Whitman, E.R., 2022. Effects of environmental factors on the detection of subsurface green turtles in aerial drone surveys. Wildl. Res. 49 (1), 79–88.
Rider, M.J., Haas, H., Sasso, C.R., 2022. Surface Availability Metrics of Leatherback Turtles (*Dermochelys coriacea*) Tagged off North Carolina and Massachusetts, United States. NEFSC Center Reference Document.
Rider, M.J., Avens, L., Haas, H.L., Patel, S.H., Sasso, C.R. (n.d.) (in preparation) Regional Variation in Leatherback Dive Behavior with Respect to environmental Conditions in Northwestern Atlantic Foraging Areas.
Robbins, W.D., Peddemors, V.M., Kennelly, S.J., Ives, M.C., 2014. Experimental evaluation of shark detection rates by aerial observers. PLoS One 9 (2), e83456.
Roberts, K.E., Garrison, L.P., Ortega-Ortiz, J., Hu, C., Zhang, Y., Sasso, C.R., Hart, K.M., 2022. The influence of satellite-derived environmental and oceanographic parameters on marine turtle time at surface in the Gulf of Mexico. Remote Sens. 14 (18), 4534.
Schofield, G., Katselidis, K.A., Dimopoulos, P., Pantis, J.D., Hays, G.C., 2006. Behaviour analysis of the loggerhead sea turtle *Caretta caretta* from direct in-water observation. Endanger. Species Res. 2, 71–79.
Schroeder, B.A., Thompson, N.B., 1987. Distribution of the loggerhead turtle, Caretta caretta, and the leatherback turtle, Dermochelys coriacea, in the Cape Canaveral, Florida area: results of aerial surveys. Witzell, WN (convener and editor). In: Ecology of East Florida Sea Turtles: Proceedings of the Cape Canaveral, Florida Sea Turtle Workshop. Miami, Florida. NOAA Technical Report NMFS. Vol. 53. 1987.
Seminoff, J.A., Shanker, K., 2008. Marine turtles and IUCN red listing: a review of the process, the pitfalls, and novel assessment approaches. J. Exp. Mar. Biol. Ecol. 356 (1–2), 52–68. ISSN 0022–0981.
Seminoff, J.A., Eguchi, T., Carretta, J., Allen, C.D., Prosperi, D., Rangel, R., Gilpatrick Jr., J.W., Forney, K., Peckham, S.H., 2014. Loggerhead Sea turtle abundance at a foraging hotspot in the eastern Pacific Ocean: implications for at-sea conservation. Endanger. Species Res. 24 (3), 207–220.
Sequeira, A.M., Heupel, M.R., Lea, M.A., Eguíluz, V.M., Duarte, C.M., Meekan, M.G., Thums, M., Calich, H.J., Carmichael, R.H., Costa, D.P., Ferreira, L.C., 2019. The importance of sample size in marine megafauna tagging studies. Ecol. Appl. 29 (6), e01947.
Shoop, C.R., Kenney, R.D., 1992. Seasonal distributions and abundances of loggerhead and Leatherback Sea turtles in waters of the northeastern United States. Herpetol. Monogr. 6 (1992), 43–67. https://doi.org/10.2307/1466961.
Smolowitz, R.J., Patel, S.H., Haas, H.L., Miller, S.A., 2015. Using a remotely operated vehicle (ROV) to observe loggerhead sea turtle (*Caretta caretta*) behavior on foraging grounds off the mid-Atlantic United States. J. Exp. Mar. Biol. Ecol. 471 (2015), 84–91. https://doi.org/10.1016/j.jembe.2015.05.016.
Thomson, J.A., Cooper, A.B., Burkholder, D.A., Heithaus, M.R., Dill, L.M., 2012. Heterogeneous patterns of availability for detection during visual surveys: spatiotemporal variation in sea turtle dive–surfacing behaviour on a feeding ground. Methods Ecol. Evol. 3 (2), 378–387.
Thomson, J.A., Cooper, A.B., Burkholder, D.A., Heithaus, M.R., Dill, L.M., 2013. Correcting for heterogeneous availability bias in surveys of long-diving marine turtles. Biol. Conserv. 165, 154–161.
Wallace, B.P., Tiwari, M., Girondot, M., 2013. *Dermochelys coriacea*. The IUCN Red List of Threatened Species 2013: e.T6494A43526147. https://doi.org/10.2305/IUCN.UK.2013-2.RLTS.T6494A43526147.en.
Wallace, B.P., Zolkewitz, M., James, M.C., 2015. Fine-scale foraging ecology of leatherback turtles. Front. Ecol. Evol. 3, 15.
Wallace, B.P., Zolkewitz, M., James, M.C., 2018. Discrete, high-latitude foraging areas are important to energy budgets and population dynamics of migratory leatherback turtles. Sci. Rep. 8 (1), 1–4.
Warden, M.L., Haas, H.L., Richards, P.M., Rose, K.A., Hatch, J.M., 2017. Monitoring trends in sea turtle populations: walk or fly? Endanger. Species Res. 34, 323–337.
Westgate, A.J., Koopman, H.N., Siders, Z.A., Wong, S.N.P., Ronconi, R.A., 2014. Population density and abundance of basking sharks *Cetorhinus maximus* in the lower bay of Fundy, Canada. Endanger. Species Res. 23, 177–185. https://doi.org/10.3354/esr00567.
Wildlife Computers, 2019. Determine Biofouling Through Data. https://wildlifecomputers.com/blog/determine-biofouling-through-data/.
Wildlife Computers, 2021. SPLASH10 (-F, -BF, -FL, -X, -L, -LX, -FX) TDR10 (-DD, -F, -BF, -X, -L, FL, -FX, -LX, -BX) with host version 1.26.3002 User Guide. https://static.wildlifecomputers.com/SPLASH10-TDR10-User-Guide-3.pdf.
Wildlife Computers, 2023. FASTLOC® GPS. https://wildlifecomputers.com/data/technologies/fastloc/.