

**INTER-AMERICAN TROPICAL TUNA COMMISSION
COMISIÓN INTERAMERICANA DEL ATÚN TROPICAL**

Special Report 24
Informe Especial 24

**TRIAL SURVEY FOR EASTERN TROPICAL PACIFIC DOLPHINS:
PROJECT REPORT**

**ESTUDIO DE PRUEBA DE LOS DELFINES DEL PACÍFICO
ORIENTAL TROPICAL: INFORME DEL PROYECTO**

By-Por

Cornelia S. Oedekoven, Lewis McMillan, Sakina Fatima, David Harris-Birtill, Claudia Faustino,
Laura Marshall, Stephen T. Buckland, Cleridy E. Lennert-Cody

La Jolla, California, USA

2021

The Antigua Convention, which was negotiated to strengthen and replace the 1949 Convention establishing the Inter-American Tropical Tuna Commission (IATTC), entered into force on 27 August 2010. The IATTC is responsible for the conservation and management of the “stocks of tunas and tuna-like species and other species of fish taken by vessels fishing for tunas and tuna-like species” in the eastern Pacific Ocean, and also for the conservation of “species belonging to the same ecosystem and that are affected by fishing for, or dependent on or associated with, the fish stocks covered by [the] Convention.”

The members of the Commission and the Commissioners are listed in the inside back cover of this report.

The IATTC staff's research responsibilities are met with four programs, the Data Collection and Data Base Program, the Biology and Ecosystem Program, the Stock Assessment Program, and the Bycatch Program and International Dolphin Conservation Program.

An important part of the work of the IATTC is the publication and wide distribution of its research results. These results are published in its Bulletin, Special Report, Data Report series, and papers in outside scientific journals and chapters in books, all of which are issued on an irregular basis, and its Stock Assessment Reports and Fishery Status Reports, which are published annually.

The Commission also publishes Annual Reports and Quarterly Reports, which include policy actions of the Commission, information on the fishery, and reviews of the year's or quarter's work carried out by the staff. The Annual Reports also contain financial statements and a roster of the IATTC staff.

Additional information on the IATTC's publications can be found in its web site.

La Convención de Antigua, negociada para fortalecer y reemplazar la Convención de 1949 que estableció la Comisión Interamericana del Atún Tropical (CIAT), entró en vigor el 27 de agosto de 2010. La CIAT es responsable de la conservación y ordenación de las “poblaciones de atunes y especies afines y otras especies de peces capturadas por embarcaciones que pescan atunes y especies afines” en el Océano Pacífico oriental, así como de la conservación de “especies que pertenecen al mismo ecosistema y que son afectadas por la pesca de especies de peces abarcadas por la ... Convención.”

En la contraportada del presente informe se alistan los miembros de la Comisión y los Comisionados.

Las responsabilidades de investigación del personal de la CIAT son realizadas mediante cuatro programas: el programa de recolección de datos y bases de datos, el programa de biología y ecosistemas, el programa de evaluación de poblaciones, y el programa de captura incidental y el Acuerdo sobre el Programa Internacional para la Conservación de los Delfines.

Una parte importante del trabajo de la CIAT es la publicación y amplia distribución de los resultados de sus investigaciones. Se publican los mismos en sus series de Boletines, Informes Especiales, Informes de Datos, y publicaciones en revistas científicas externas y capítulos en libros, todos de los cuales son publicados de forma irregular, y sus Informes de la Condición de las Poblaciones e Informes de la Situación de las Pesquerías, publicados anualmente.

La Comisión publica también informes anuales y trimestrales, los que incluyen acciones de política de la Comisión, información sobre la pesquería, y resúmenes de trabajo realizado por el personal en el año o trimestre correspondiente. Los informes anuales contienen también un estado financiero y una lista del personal de la CIAT.

Se presenta información adicional sobre las publicaciones de la CIAT en su sitio web.

DIRECTOR
Jean-François Pulvenis (Acting)
HEADQUARTERS AND MAIN LABORATORY—OFICINA Y LABORATORIO PRINCIPAL
8901 La Jolla Shores Drive
La Jolla, California 92037-1509, USA
www.iattc.org

**INTER-AMERICAN TROPICAL TUNA COMMISSION
COMISIÓN INTERAMERICANA DEL ATÚN TROPICAL**

Special Report 24
Informe Especial 24

**TRIAL SURVEY FOR EASTERN TROPICAL PACIFIC DOLPHINS:
PROJECT REPORT**

**ESTUDIO DE PRUEBA DE LOS DELFINES DEL PACÍFICO
ORIENTAL TROPICAL: INFORME DEL PROYECTO**

By-Por

Cornelia S. Oedekoven¹, Lewis McMillan², Sakina Fatima², David Harris-Birtill², Claudia Faustino¹,
Laura Marshall¹, Stephen T. Buckland¹, Cleridy E. Lennert-Cody³

La Jolla, California, USA
2021

¹ Centre for Research into Ecological and Environmental Modelling / University of St Andrews

² Computer Science / University of St Andrews

³ Inter-American Tropical Tuna Commission, 8901 La Jolla Shores Drive, La Jolla CA 92037-1509, USA

1. EXECUTIVE SUMMARY

Following recommendations from a recent workshop on methods for estimating marine mammal abundance (Johnson *et al.* 2018), the Inter-American Tropical Tuna Commission (IATTC), in coordination with the *Comision Nacional de Acuacultura y Pesca* (CONAPESCA) of the government of Mexico and with the support of the *Instituto Nacional de Pesca y Acuacultura* of Mexico (INAPESCA), the Mexican tuna industry, and the Centre for Research into Ecological and Environmental Modelling (CREEM) at the University of St Andrews (USTAN), is planning to undertake a ship-based survey, in conjunction with drones, to estimate the abundance of dolphin populations in the eastern tropical Pacific (ETP) Ocean and improve management of those populations and the ETP ecosystem. The project was designed to be carried out in two phases: a short trial survey, and a longer main survey. The purpose of the trial survey was to test the methodology developed by researchers at CREEM (Oedekoven *et al.* 2018) and the suitability of the equipment provided for the main survey. Specifically, in addition to evaluating the overall suitability of the research vessel and marine mammal observers for the project, the trial survey had two goals related to the use of drone equipment provided to the project: 1) test whether the drones can be used to detect dolphin schools directly ahead of the survey vessel that might be missed by ship-based marine mammal observers, data essential for estimating the trackline detection probability of ship-based observers, and 2) test whether the drones provided can be used to collect data on dolphin school size and species composition, data essential for calibrating estimates made by the ship-based observers (school size calibration). The drone equipment, cameras and drone personnel were selected for and provided to the trial survey project by Gtt NetCorp.

To test the research plan, an international research team sailed out from Mazatlán, Mexico, aboard the R/V Dr Jorge Carranza Fraser, provided by the INAPESCA for the project, for a 14-day trial survey, from 17 – 30 November, 2019. The team was led by Dr Cornelia Oedekoven of USTAN and Dr Cleridy Lennert-Cody of IATTC, and composed of scientists, drone pilots and mechanics from four different countries (Mexico, United States, Germany and Chinese Taipei). The area off the Mexican coast between Manzanillo and Acapulco was selected as the study area because it has been shown to be the area with highest density of spotted and spinner dolphins within the ETP, regardless of season.

The research vessel was outfitted with a special observation platform on the level above the bridge, called the flying bridge. A team of six experienced observers in a 2-hourly rotation with three observers on watch at any time during suitable conditions scanned the forward 180° for cetaceans and logged the required information according to the U.S. National Marine Fisheries Service (NMFS) survey protocol that has been consistently used during previous ETP surveys. This protocol prescribes that surveys are conducted in closing mode¹. The protocol also includes a school size calibration component where the school size estimates of the observers are compared against true counts of dolphins in suitable schools (“calibration” schools). During previous surveys, true school size counts were obtained from aerial photography taken from helicopters carried on the survey vessel or shore-based fixed-winged aircraft. During the trial survey, drones were tested for collection of video imagery that could be used for school size calibration.

During the trial survey drones also were tested for collection of video imagery suitable for estimating trackline detection probability. The preferred method for estimating trackline detection probability is

¹ Closing mode means that upon detection of a cetacean school, the ship leaves the survey trackline and approaches the school to gather information on species identification, school composition and school size. The alternative to closing mode is to operate in passing mode where all information is gathered from a distance as the ship continues along the trackline.

mark-recapture distance sampling (MRDS). In contrast with conventional distance sampling where line-transect data are collected from a single platform, MRDS methods require double-observer platform data. For this survey, a drone served as platform 2 and was to survey the area in front of the ship covering a wide corridor across the ship's transect line by flying in either a zigzag pattern or parallel lines while maintaining station at 5 nm ahead of the ship (hereafter referred to as "zigzag" flights). Detections of cetacean schools made via the drone were to serve as trials for the flying bridge (platform 1). Of interest was whether dolphin schools detected by the drone were later detected by the ship-based observers. Video footage captured by the drone during these zigzag flights was to be sent back to the ship for real-time monitoring by the drone observers and recorded on-board the drone for post-survey image analyses.

Due to the extensive experience of the flying bridge observers, the implementation of the NMFS survey protocol on the Jorge Carranza was successful. The flying bridge equipment worked well, although a few fixes and alterations are needed for the main survey. The captain and the other ship officers were very effective and helpful at implementing the survey protocol including quick responses to requests made by flying bridge observers, maneuvering the ship in closing mode so school size and species composition estimates could be obtained. A total of 1,733.06 km of transect lines were surveyed, out of which 766.41 km were conducted in closing mode and 966.65 km in passing mode. All survey effort should have been conducted in closing mode, as that is the required mode for the main survey. However, after unsuccessfully attempting drone zigzag flights with the flying bridge operating in closing mode for two days, passing mode was used to facilitate the further testing of zigzag flights. A total of 215 sightings (205 on effort, 10 off effort) of 26 different species categories were made by the flying bridge observers. A comparison of estimated detection probabilities for spotted and spinner dolphins from the trial survey with previous surveys conducted on smaller research vessels revealed no significant differences between the vessels used in previous surveys and the R/V Dr Jorge Carranza Fraser.

Despite some success with the use of the Seahawk drone for school size calibration (see below), results of the trial survey showed that a different drone system, including better cameras and antennas, will be necessary to evaluate trackline detection probability during the main survey. The drone team was able to launch and land drones in Beaufort sea states up to and including 5. Although one drone was lost at sea due to a sudden loss of satellite coverage, the drone team performed a total of 94 flights, for a total flight time of about 69 hrs. Of these 94 flights, 74 were pure zigzag flights, 15 flights were pure calibration flights and four were initiated as zigzag and then switched to calibration mode. While the drone team safely conducted up to an impressive 13 flights per day, this was not sufficient to cover the entire hours of operation of the flying bridge observers, given the short flight time of the Seahawk drone (< 1 hr) and the presence of only one highly skilled drone pilot on board. Based on the results from this survey, we estimate that, even with a second highly skilled drone pilot, full coverage of all daylight hours would require too many launches and landings during a 120-day main survey, which would be a major safety concern.

Analysis of the video imagery collected by the drone during the trial survey revealed that a better method for archiving video data will be required for the main survey. Contrary to what was specified in the trial survey protocol, continuous recording of the video on-board the drone was not implemented by the drone company. Instead, the drone company elected to use screen capture of the transmitted video as the primary method to store video for post-survey image analyses. This screen recording process resulted in major reductions in video quality due to transmission loss, two layers of compression of the video, various artefacts, frequent pixilation, occasional complete loss of the video and a reduction in the frame rate compared to what was originally recorded by the camera

onboard the drone. Due to the screen recording process, the video footage from the zigzag flights was often of too poor quality to classify objects of interest as dolphins with certainty during real-time monitoring as well as during post-survey review and image analysis from the still frames. During post-survey review by a human observer, large frame context helped in this regard, as well as motion of the animals through the video sequence.

In addition to a better video data archiving procedure, much better resolution for the camera will be required for the zigzag flights during a main survey to cover a larger area and provide better ground resolution. The total length of all zigzag flight legs was 1,013.37 km; with a mean swath width of 154.62 m, the total area covered by the drone during the zigzag flight legs amounted to < 1% of the area covered by the flying bridge observers within their survey strip. The small size of the area covered by the drone was a result of the poor video quality, forcing the drone to fly at low altitudes and at low speeds to be able to detect dolphins. The short flight times of the Seahawk and the rigid method used by the drone company for uploading waypoints onto the drone before each zigzag flight, without being able to change it mid-flight, were the main reasons why implementing the zigzag protocol during closing mode effort was not possible with the Seahawk drone. While the latter issue may be resolved, the short endurance of the Seahawk makes it unsuitable for implementing zigzag flights during closing mode effort. A different drone with longer endurance, and a more flexible flight plan editing procedure, will be required for implementing the zigzag flights during closing mode effort during the main survey.

Despite the video quality issues noted above, three proof-of-concept machine learning models were developed for analysing data generated from the video footage collected during the trial survey. The first machine learning model used convolutional neural networks (CNNs) and the still images from the video footage. The second used clustering algorithms and the video footage from which velocities of objects in the frame were calculated using the optical flow technique. Both these models achieved about 75% balanced accuracy² on their validation datasets. A third model was developed which combined these two approaches and achieved better balanced accuracy of about 83%. This model performance was not sufficient to make useful detections in the video footage for either the zigzag or calibration flights. This is because, while detections of cetaceans could be made via video analysis, review of these detections by a human observer revealed that the false positive detections caused by poor video quality outweighed the cetacean detections.

As a result of careful review of the video imagery by human observers, five detections of cetacean schools by the drone observers during zigzag flights could be confirmed, which demonstrates that detections of cetaceans could be made in real-time by the drone observers and, hence, using drones to collect MRDS data is possible. During real-time monitoring, drone observers logged 92 potential objects of interest; out of these, six were confirmed as detections of cetacean schools. Two of these were of the same school, giving a total of five data points (trials) for the MRDS analyses. For three of these trials the outcome was determined to be failure, i.e. flying bridge observers did not detect the same school, as none of the sightings made by the flying bridge was a potential duplicate. For the other two drone detections a potential match with respective to flying bridge detections could neither be confirmed nor excluded. New MRDS analysis methods need to be developed to incorporate this uncertainty in the duplicate matching as well as to accommodate closing mode effort.

In spite of the poor performance of the Seahawk drone for zigzag flights, the trial survey demonstrated that this type of drone can be used successfully for school size calibration flights. However, a higher resolution camera is needed to identify *all* individuals to species and to ensure animals swimming in

² Balanced accuracy is the mean of the true positive rate and the true negative rate.

close proximity to each other can be distinguished. Both are required to obtain accurate counts by species. Recording multiple sweeps across a given calibration school with slightly varying angles proved important to alleviate potential glare issues. For six schools, all clusters were captured with the drone footage. Manual counts by a human observer were obtained for five of these schools and, hence, the schools are valid calibration schools. However, to reliably make detections from the calibration flights and obtain counts via image analysis, better quality video is needed (see following paragraph).

Given the results of this trial survey, in the next phase of the project, different drone-camera systems with longer endurances and greater video resolution than those provided for the trial survey should be tested. Higher video resolution would also allow the drone to operate at higher altitudes while maintaining the same ground resolution. Increased altitude would also increase the area covered by the drone, and thus increase the sample size for the trackline detection probability assessment. These tests should be conducted in a short sea-trial on a vessel from which drones can be launched under similar conditions as the Jorge Carranza. We recommend that before such a trial, any potential drone provider should provide a detailed assessment of how they can accomplish the project goals. The duration of such a sea trial should be long enough to collect data suitable for improving image analysis algorithms. This requires that schools of dolphins will need to be captured with the video recorded during the zigzag flights flown using the parameters (i.e. the drone altitude and speed as well as the video resolution) required to implement a full coverage of the area 5nm ahead of the ship with a corridor half-width of 3nm. The necessity of conducting a further sea trial and post-trial image analyses needs to be factored into the timing of the main survey.

2. BACKGROUND

Following recommendations from a recent workshop on methods for estimating marine mammal abundance (Johnson *et al.* 2018), the Inter-American Tropical Tuna Commission (IATTC), in coordination with the *Comision Nacional de Acuacultura y Pesca* (CONAPESCA) of the government of Mexico and with the support of the *Instituto Nacional de Pesca y Acuacultura* of Mexico (INAPESCA), the Mexican tuna industry, and the Centre for Research into Ecological and Environmental Modelling (CREEM) at the University of St Andrews (USTAN), is undertaking a ship-based survey, in conjunction with drones, to estimate the abundance of dolphin populations in the eastern tropical Pacific Ocean (ETP) and improve management of those populations and the ETP ecosystem. The project was designed to be carried out in two phases: a short trial survey, and a longer main survey. The purpose of the trial survey was to test the methodology developed by researchers at CREEM (Oedekoven *et al.* 2018) and equipment provided for the main survey. Specifically, in addition to evaluating the overall suitability of the research vessel and marine mammal observers for the project, the trial survey had two goals related to the use of drone equipment provided to the project: 1) test whether the drones can be used to detect dolphin schools directly ahead of the survey vessel that might be missed by ship-based marine mammal observers, data essential for estimating trackline detection probability of ship-based observers, and 2) test whether the drones provided can be used to collect data on dolphin school size and species composition, data essential for calibrating estimates made by the ship-based observers (school size calibration).

To test the research plan, an international research team sailed out from Mazatlán, Mexico, aboard the R/V Dr Jorge Carranza Fraser, provided by the INAPESCA for the project, for a 14-day research trip, from 17 – 30 November, 2019. The team was led by Dr Cornelia Oedekoven of USTAN and Dr Cleridy Lennert-Cody of IATTC, and composed of scientists, drone pilots and mechanics from four different

countries (Mexico, USA, Germany and Chinese Taipei). The area off the Mexican coast between Manzanillo and Acapulco was selected as the study area for the trial survey because it has been shown (Reilly 1990) to be the area with highest density of spotted and spinner dolphins (Figure 1) within the ETP, regardless of season.



FIGURE 1. Eastern spinner dolphins (left) and spotted dolphin (right). Photo credits: Paula Olson.

In this report we present a summary of work undertaken during the trial survey and results based on detailed analysis of all the data collected during the trial survey. We discuss the implications of the outcome of the trial survey with regard to equipment, methods and timing of a main survey.

3. OBJECTIVES TRIAL SURVEY

As proposed by Oedekoven *et al.* (2018), the potential objectives of the main survey are:

1. Estimate relative abundance of priority stocks such that the estimates are comparable as far as possible with past estimates from NMFS surveys.
2. Estimate absolute abundance of the priority stocks.

For this phase of the project it was assumed that the priority species included only the north-eastern offshore spotted dolphin and the eastern spinner dolphin; for a definition of priority stocks see Oedekoven *et al.* (2018; Section 2.2). Previous surveys in the ETP have been conducted on NOAA research vessels following a well-defined survey protocol (Kinzey *et al.* 2000). During these cruises, employing trained observers and high observer consistency within and between cruises contributed to the comparability of estimates from different years, as did repeated use of the same research vessels. Hence, to meet objective 1, it is necessary to maintain this comparability for the main survey with regards to the ship, observers and survey protocol. Furthermore, school size calibration has generally been a component of previous NMFS surveys in the ETP where suitable calibration schools were photographed from helicopters or fixed-winged aircraft. Using drones instead for this purpose, as proposed by Oedekoven *et al.* (2018), needs to be tested before a main survey.

In order to address objective 2, Oedekoven *et al.* (2018) proposed to use drones to collect mark-recapture distance sampling (MRDS) data. For this purpose, a drone should be flown ahead of the ship, preferably during all daylight hours of every survey day when flying bridge observers are on effort, i.e. during suitable viewing conditions with Beaufort sea states < 6, sufficient visibility and no rain. The drone should fly 5nm or more ahead of the ship covering the area widely in a zigzag pattern or with parallel lines. Video footage taken via the drone needs to be transmitted back to the ship for real-time monitoring by the drone observers as well as be recorded aboard the drone for post-survey image analyses. The drone needs to be flown at a suitable altitude and speed to video-capture a sufficiently large area in front of the ship to obtain a large enough sample size (number of detected cetacean

schools) for MRDS analyses. At the same time, the video footage should have sufficient resolution to allow the species of any dolphins detected to be identified.

Hence, for the two main survey objectives above combined, the purpose of the trial survey was to perform the following tests listed below.

- 1 Test the research vessel as a suitable platform to conduct line transect surveys using NMFS protocol:
 - a. Flying bridge setup;
 - b. Observers;
 - c. Ship and crew.
- 2 Test the research vessel as a suitable platform to conduct drone operations:
 - a. Platform for launching and landing during varying sea states;
 - b. Drone operations;
 - c. Drone team.
- 3 Test the drone for collecting trackline detection probability data:
 - a. Fly 5nm or more ahead of the ship in a zig-zag pattern or parallel lines covering a large area during closing mode effort;
 - b. Collect high-resolution video of suitable quality for dolphin species identification and record on-board the drone;
 - c. Real-time transmission of video back to the ship;
 - d. Real-time monitoring by drone observers for cetacean sightings;
 - e. Develop algorithms post-survey for automatic detection of cetacean schools using video recorded on board the drone.
- 4 Test the drone for collecting school size calibration data:
 - a. Navigate the drone over the calibration school;
 - b. Fine scale manoeuvring when with the school;
 - c. Collect high-resolution video of the school of suitable quality for species identification of each animal and record video on-board the drone;
 - d. Real-time transmission of video back to the ship;
 - e. Real-time monitoring by drone observers for cetacean sightings;
 - f. Develop algorithms post-survey for automatically generating counts of cetacean groups using video recorded on board the drone.

4. STUDY AREA AND ITINERARY

Following Oedekoven *et al.* (2018), the study area for the trial survey was selected to be within the area of the ETP in which highest densities of the priority species, the north-eastern offshore spotted dolphin and the eastern spinner dolphin, would be expected (Reilly 1990). There were 16 pre-designed transect lines to be followed by the research vessel during the trial survey (Figure 2, Table 1). Transects 1 and 16 were designed to lead out of and back to Mazatlán, while transects 2-15 were in the high-density area (Figure 2). The scheduled dates for completing the transect lines are shown in Table 1. It is noted that for the purpose of this trial survey, it was not important to follow the transect lines exactly or to cover all lines within the survey. The lines served merely as examples to practice the protocol for the main survey.

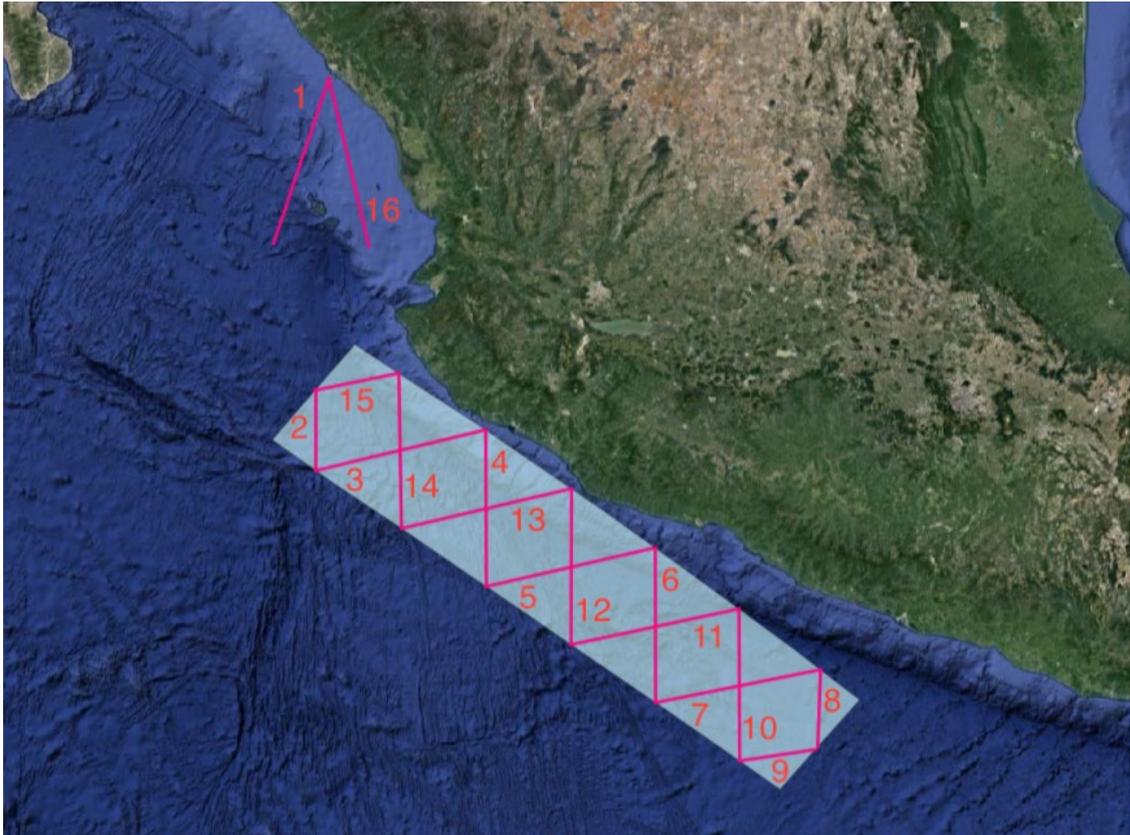


FIGURE 2. Study area (shaded area) and transect lines designed for the trial survey.

TABLE 1. Planned dates in 2019 for surveying each of the transect lines shown in Figure 2.

Planned date	Transect
17 November	1
18 November	2
18 – 19 November	3
19 – 20 November	4
20 – 21 November	5
21 – 22 November	6
22 – 23 November	7
23 November	8
24 November	9
24 – 25 November	10
25 – 26 November	11
26 – 27 November	12
27 – 28 November	13
28 – 29 November	14
29 November	15
30 November	16

5. METHODS: EQUIPMENT AND PROCEDURES

5.1. Research vessel as a suitable platform to conduct line transect surveys using NMFS protocol

5.1.1. Flying bridge installations

To serve as a marine mammal survey vessel, the R/V Dr Jorge Carranza Fraser had to be outfitted with a special observation platform on the level above the bridge, called the flying bridge. This flying bridge included four sets of high-powered binoculars (bigeyes) mounted on pedestals, a data recorder station, a wind dam and a canopy for sun protection (Figure 3). The flying bridge setup included a recorder station with a waterproof box containing the keyboard and monitor connected to a PC located in the drone lab (shown below in Figure 6, Section 5.3.2). At any given time when flying bridge observers were on effort, three observers were simultaneously on watch, rotating at forty minute intervals from the port side bigeyes to the central observer station and then to the starboard bigeyes (more details in the next section). The observer at the center station (“recorder”) entered data on survey effort, viewing conditions and sightings using WinCruz software developed for previous NMFS surveys. Crucial information for initialising WinCruz software was the height of the binoculars mounted on the flying bridge above sea level (20.44m).

The design of this flying bridge was based on instructions and examples provided by CSO with consultancy from JCS (full names corresponding to initials are provided in Table 2). Using these instructions, engineers at SENAV designed a blue-print and worked on the installation of the flying bridge before the survey.

It is noted that the installation of the flying bridge according to the blueprint was never fully completed before the trial survey. This had several consequences. The metal platforms that were supposed to surround the pedestals were left as full rectangular platforms sitting behind the pedestals. This was a major problem as the observers needed to scan not only the area forward of the ship but also to the sides. The platforms were also not high enough to accommodate the observers for viewing through the bigeyes. The observer chairs were built too low so that the central observer sitting in the chair could not scan the waters near the ship (which is one of their main duties) while seated. Temporary fixes were implemented by JCS before the trial survey started (see Figure 3). These included holes cut into the metal platforms so that they could surround the bigeye pedestals, and stacks of wood to raise the height of the platforms. A safer solution to the latter issue will be needed for a main survey.



FIGURE 3. Observers on the flying bridge: three main observers, SY, EV and AB (for initials refer to Table 2) scan the forward 180° with big eyes binoculars and naked eye. Tracker BB, on the right, focuses on the central areas within 15° on either side of the ship. Note the wooden pieces used to increase the height of the platforms were only a temporary fix. This was required as the pedestals did not allow lowering the bigeyes enough for the observers.

5.1.2. NMFS survey protocol: line transect surveys in closing mode

During previous surveys, which were conducted by the NMFS, conventional distance sampling methods (*e.g.*, Buckland *et al.* 2015) were used to estimate abundance of dolphins (Gerrodette *et al.* 2008). During such surveys, the ship travelled sequentially along each transect line placed in the study area, and trained marine mammal observers recorded the location of cetacean schools using measurements of the radial distance to the dolphin schools from the ship and angle in relation to the ship's heading (Figure 4), as well as the size and species composition of the dolphin schools (Kinzey *et al.* 2000). Radial distances were obtained with the use of the reticle scale in the right ocular of the bigeyes, and angles were read off the angle ring. The recorder entered the angle and reticle into WinCruz software which calculates a distance to the sighting using the reticle and height of the observer that made the sighting above sea level.

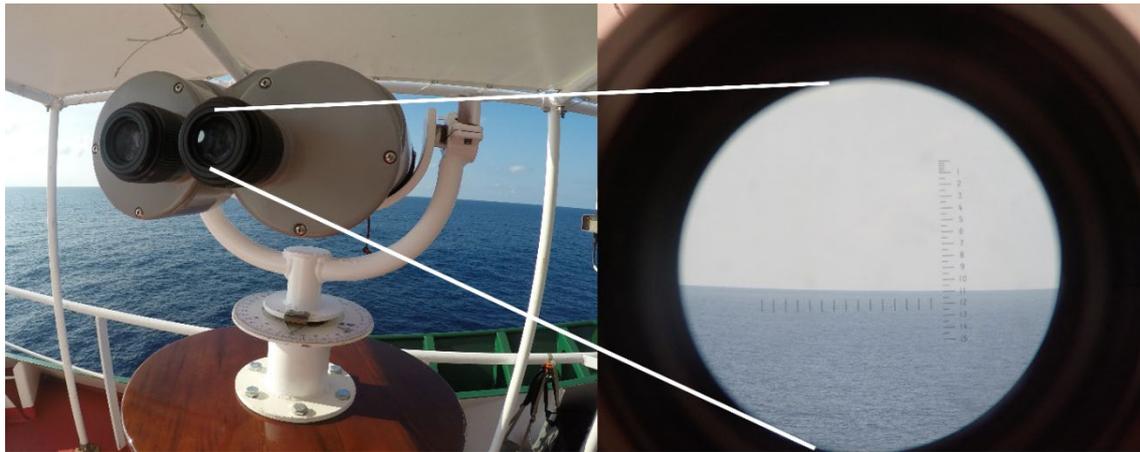


FIGURE 4. Bigeye binoculars (25 x 150 powered): the angle ring and pointer as well as the reticle scale in the right ocular allowed determination of the angle and radial distance of the detected dolphin school in relation to the ship.

Previous ETP surveys were conducted in closing mode where, in the case that the flying bridge observers detected a school within 3nm perpendicular distance from the transect, they stopped search effort when the sighting was made and turned the ship to approach the detected school in order to obtain species identification, group composition and group size estimates. These approaches were directed from the observers on the flying bridge in direct communication via VHF radio to the bridge. When all the necessary data on the school were collected, the flying bridge would resume search effort from their current location while the ship continued along the transect or on a track parallel to the original transect, if the approach had taken them away from the original transect. Further details can be found in (Kinzey *et al.* 2000).

The alternative to closing mode is passing mode where the observers collect all information on the school from a distance while the ship passes the school without changing course or speed. Using the data collected during a survey in the ETP dedicated to reveal potential differences between closing mode and passing mode estimates, Schwartz *et al.* (2010) demonstrated that closing mode encounter rate estimates were on average lower than passing mode estimates. Passing mode, on the other hand, can result in negatively biased school size estimates and poor species identification.

Species are always identified to the lowest taxonomic level possible using NMFS species codes (Kinzey *et al.* 2000). If, for example, a group of dolphins was too far away to be seen well enough, the species code entered for the group could be one of the following options: 077 (unid. dolphin), 177 (unid. small delphinid), 277 (unid. medium delphinid). In some areas, subspecies distributions overlap, which may result in the use of specific codes to indicate that the group could not be identified to any particular subspecies (e.g. 090: *Stenella attenuata* unid. subsp.).

The NMFS survey protocol further includes a special protocol for obtaining group size estimates as there is a considerable amount of uncertainty and potentially bias in the observer estimates. For every school, each observer on watch makes an independent estimate of the group size, including best, high and low estimates, as long as they were confident that they saw the entire school. In the case of multispecies schools, observers also estimate the percentage of the school represented by each species. Each observer records estimates in their own notebook, without exchanging any information on numbers with other observers. The cruise leader transfers the estimates into the WinCruz data files at the end of each day. At the analysis stage, estimates are adjusted by an observer-specific calibration factor with the intent to correct potential biases in each observer's estimates (e.g. Gerrodette *et al.*

2018). This factor is obtained for each observer by comparing their school size estimates for specific schools, called calibration schools, against the true counts for these schools. Calibration schools are those schools for which aerial photographs were taken, capturing the entire school, and thus true counts could be obtained. In this way, the calibration schools are used to calibrate each observer, i.e. compare each observer's estimates with the true counts and estimate an observer-specific calibration factor. Observers are never told what their calibration factor is while they are active observers, i.e. before retirement, in order to maintain consistency across years.

5.1.3. Observers

Out of the 12 scientific personnel who participated in the trial survey, eight have extensive experience conducting marine mammal surveys using the NMFS survey protocol (rows 1 - 8 in Table 2) of which a large proportion took place in the ETP. Six of these (rows 3 - 8 in Table 2) formed the key flying bridge observer team rotating through two-hourly watches (40 min at each position: port side bigeyes, recorder, starboard bigeyes) throughout all daylight hours. The survey coordinator and visiting scientist stood drone observer watches and trained the two junior observers in their duties. Two junior observers (rows 9 - 10 in Table 2) were recruited from the IATTC observer pool and have extensive experience of monitoring marine mammals in the ETP. These were trained in both drone observer and flying bridge duties. For the latter, they stood tracker watches, tracking sightings on the bigeyes, and evaluated in situ if their sightings matched any of those made by the main team of flying bridge observers.

TABLE 2. Scientific personnel on board the trial survey (not including INAPESCA personnel).

#	Initials	Title	Name	Tasks
1	CSO	Chief scientist / cruise leader	Dr Cornelia Oedekoven	Supervision and coordination of all teams, communication with command, drone directing during calibration schools, daily science meetings, daily data editing and backup
2	JCS	Survey coordinator/lead drone observer	Mr Juan Carlos Salinas	Drone station and flying bridge watches, data entry and proofing, logistics, hardware and software maintenance, final flying bridge setup
3	PO	Flying bridge observer	Ms Paula Olson	Flying bridge watches, sighting forms, data entry and proofing, photo-id, schedule
4	SY	Flying bridge observer	Ms Suzanne Yin	Flying bridge watches, sighting forms, data entry and proofing, training new personnel, WinCruz software installation
5	AB	Flying bridge observer	Ms Andrea Bendlin	Flying bridge watches, sighting forms, data entry and proofing, photo-id
6	CH	Flying bridge observer	Mr Christopher Hoefler	Flying bridge watches, sighting forms, data entry and proofing
7	DB	Flying bridge observer	Ms Dawn Breese	Flying bridge watches, sighting forms, data entry and proofing
8	EV	Flying bridge observer	Mr Ernesto Vazquez	Flying bridge watches, sighting forms, data entry and proofing
9	RDL	Jr observer/drone observer	Mr Ramon De Leon	Drone station watches, tracker watches flying bridge
10	BB	Jr observer/drone observer	Mr Braulio Bernal	Drone station watches, tracker watches flying bridge
11	CLC	Visiting scientist/lead drone observer/assistant cruise leader	Dr Cleridy Lennert-Cody	Drone station watches, assisting cruise leading (see tasks above)
12	AS	Host country observer	Dr Alvin Suarez	Logistical coordination

5.1.4. Ship and crew

Closing mode effort required that the ship generally traveled at a constant speed of 10 knots along the transect line; however, when flying bridge observers detected a school of cetaceans they requested changes in course and speed from the bridge officer on duty via VHF radio. An efficient approach of the school depends on the rapid responses of the officers to these potentially frequent requests, and on the general manoeuvrability of the ship, i.e. how well it accelerates/decelerates and turns.

5.1.5 Estimating detection probabilities and comparison with previous surveys

Detection probabilities were estimated by fitting a half normal detection model to the observed perpendicular distances with a truncation distance w (largest distance included in analyses) of 5.5 km. Models were estimated for spotted and spinner dolphins where data for all subspecies were combined for the respective species to increase the sample size for the analyses (Buckland *et al.* 2015). For more robust estimation, we used multiple covariate distance sampling (MCDS) methods where data from spotted and spinner dolphins were combined in a single analysis with species as a covariate for the detection function model (Marques *et al.* 2007). This provided species-specific estimates of the detection function, its parameters and various quantities that can be derived from the detection function, such as the average detection probabilities p within the search strip of half-width w or the probability density function of observed distances evaluated at distance zero $f(0)$ where $f(0) = \frac{1}{pw}$.

We compared these estimates with those from previous surveys (1979-2000) using the estimates of $f(0)$ for the northeastern offshore spotted dolphin and the eastern spinner dolphin from (Gerrodette and Forcada 2005). For a more formal investigation on whether significant differences between ships or between present and previous surveys in general exist, MCDS analyses should be done using all data from previous and present surveys combined with ship or year as a covariate. However, data from previous surveys were not available.

5.2 Research vessel as a suitable platform to conduct drone operations

5.2.1 Drone platform

The vessel was outfitted with a special platform on the stern for launching and landing the drone (Figure 5). This platform was a temporary construction providing a T-shaped flat surface spanning the full width of the ship and elevated from the back deck, thereby reducing the amount of metal obstructions that could be hazardous to launching and landing operations, especially on a moving vessel. Ropes were fitted across the platform that the drone team used to tie themselves to the platform via harnesses for safety.

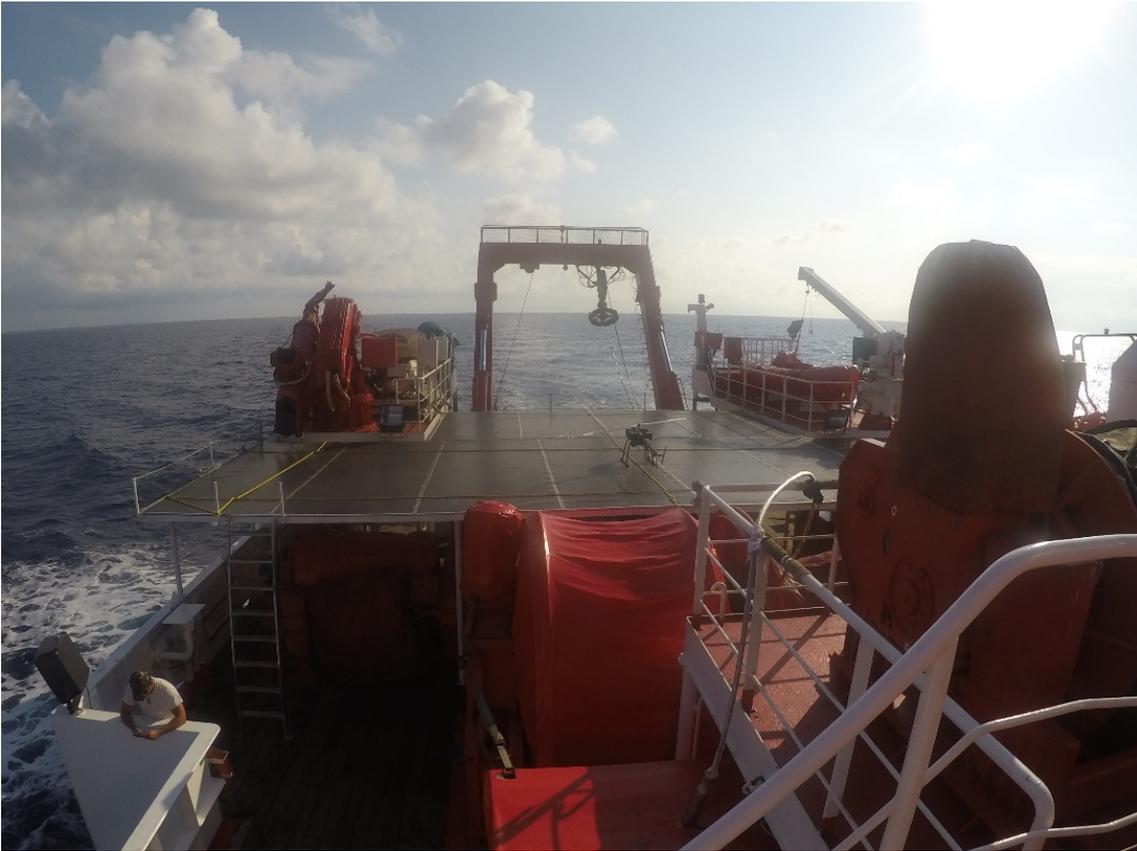


FIGURE 5. Platform for launching and landing drones on the back deck of the R/V Dr Jorge Carranza Fraser with a Seahawk S drone operated by pilots from Gtt NetCorp ready for take-off.

5.2.2 Drone operations

Drones were supposed to be flown during all daylight hours when flying bridge observers were on effort, i.e. during Beaufort sea states up to 5. The feasibility of this was to be tested during the trial survey, as well as successful completion of flights. The two different types of drone operations required for the various objectives (Section 3) posed different challenges which are described in the following sections (Sections 5.3 and 5.4). Hereafter, for the sake of brevity, we will refer to the flights conducted with the purpose of collecting data on trackline detection probability as “zigzag” flights and the flights conducted for school size calibration as “calibration” flights.

5.2.3 Drone equipment and personnel

The equipment and personnel were selected for and provided to the trial survey project by Gtt NetCorp. Three Seahawk S drones (Geodrones, www.geodrones.com, referred to as Seahawk in the following for brevity) were made available, of which two were complete systems with a camera each, and one was without camera (see Appendix for drone specifications). The personnel provided consisted of five drone pilots and engineers and one liaison and logistics support person (Table 3). The primary pilot, LCY, was the only 1st pilot and responsible for all launches and landings. The relief pilots (2nd pilots) took over while the drone was in the air. TW was the team leader and responsible for communications with the scientific personnel. The operations manager, CO, executed fast change-overs between drone flights.

TABLE 3. Drone team members provided by Gtt NetCorp.

Initials	Title	Name	Tasks
LCY	1 st (main) Pilot / Engineer	Mr Lu Cheng-Yuan (Kevin)	Responsible for all take offs and landings from ship; hand over to other pilots cruising mode
TW	2 nd (relief) pilot	Mr Travis Wiginton	Support flying helicopters while cruising; hand over the take-off and landing, communication with chief scientist and at daily science meetings
LYM	Support C3 Engineer & Language; 2 nd (relief) pilot	Mr Lin Yi-Min	Support flying helicopters while cruising; hand over when take-off and landing
LJH	Support C3 Engineer & Language; 2 nd (relief) pilot	Mr Lin Jia-Huei (Jason)	Support flying helicopters while cruising; hand over when take-off and landing
CO	Operations Manager and Support C3 Logistics	Mr Carlos Orellana	Drone maintenance, change-over of drones between flights
VT	Liaison language support; C3 Logistics Coordination support	Ms. Violeta Trigueros	Assist drone team, attend daily science meetings

5.3 Trackline detection probability

5.3.1 Mark-recapture distance sampling (MRDS) with drones as a second observer

During line transect surveys, observers are allowed to miss dolphin schools that pass the ship at some distance from the transect line. However, it is assumed that observers detect all schools directly on the transect line (Buckland *et al.* 2015). If schools on the transect line pass undetected, the resulting abundance estimates will be biased low. As the dive intervals of the dolphin species of interest are relatively short, it has generally been assumed that all schools on the transect lines are detected, regardless of weather conditions (Gerrodette and Forcada 2005). However, a recent analysis of previous survey data (Barlow 2015) provided evidence that this may not be the case (hereafter referred to as the “g(0) issue”), and that the probability of missing schools on the transect line increased when wind speeds increased and viewing conditions deteriorated.

During the trial survey, we tested if the drones provided could be used to collect data for estimating trackline line detection probability. The preferred method for addressing the g(0) issue for the ETP survey is mark-recapture distance sampling (MRDS, e.g. Borchers 2012). In contrast with conventional distance sampling where, e.g., line-transect data are collected from a single platform, MRDS methods require double-observer platform data. Here, detections made from one platform, say platform 2, represent trials for the other platform, say platform 1. In this context, trial outcomes (successes or failures) refer to whether or not platform 1 detects a group of dolphins initially detected by platform 2. Here it is crucial that the two observation platforms are such that platform 2 does not influence the observers on platform 1.

5.3.2 Testing zigzag flights

For this survey, a drone would serve as platform 2 (see previous section) and survey the area in front of the ship by flying a zigzag pattern, covering a wide corridor across the transect line while maintaining station at 5 nm ahead of the ship. Video footage captured by the drone would be sent back to the ship for real-time monitoring by the drone observers (Figure 6) and recorded onboard the drone for post-survey image analyses. The sightings made via the camera equipment aboard the drone (using detections made by both real-time observation and image analyses) would represent the trials for the flying bridge observers (platform 1). The drone should survey sufficiently far ahead of the ship so that dolphins will not have reacted to the ship at the time they are detected by the drone. This distance was to be investigated during the trial survey.



FIGURE 6. Drone observers CLC and JCS monitor the drone video footage transmitted back to the ship in real-time. Drone engineer LJH of Gtt NetCorp, on the left, ensures that video footage and GPS data are recorded.

Detections of dolphin schools and their locations were to be compared with those made by the flying bridge observers. Details of the drone protocol and the equipment provided to the project can be found in the Appendices. Given the drone protocol, the following tests of the use of Seahawk drones and cameras were to be conducted:

- 1) Fly a zigzag pattern 5 nm ahead of ship covering a wide corridor continuously during all daylight hours when the flying bridge observers are on effort (Figure 7a);
- 2) Transmit video in real-time to the ship and provide a real-time feed of drone GPS into computer system used by the drone observers; archive video, GPS and altitude information on board the drone;
- 3) Fly two drones at the same time to provide continuous coverage of the ship's trackline during all hours when flying bridge observers are on effort; and,
- 4) When the ship turns on a sighting made by the flying bridge observers, the drone turns to the location of the sighting and continues the zigzag flight from there on a projected transect line (Figure 7b).

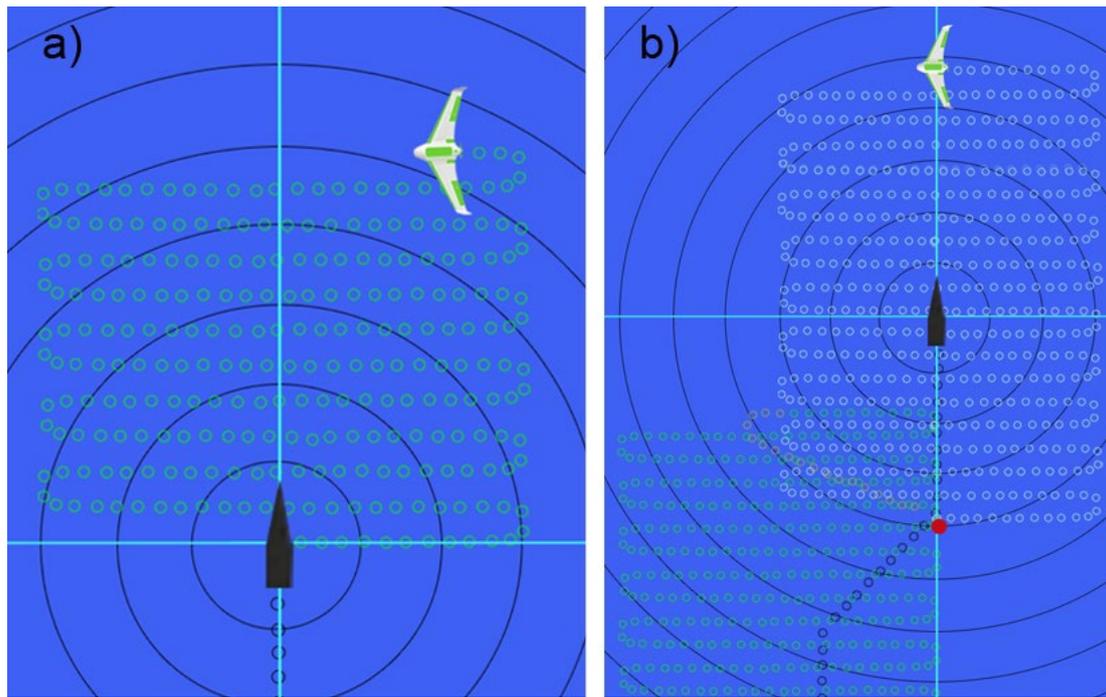


FIGURE 7. Zigzag flight of the drone (green-white symbol) in relation to the ship (black polygon) where green and black bubbles represent the recent track of the drone and ship, respectively. Light blue lines represent 0° , 90° , 180° and 270° from the ship, black circles are concentric distance rings around the ship, each 1nm apart, red dot is a dolphin sighting made by flying bridge observers. a) Drone covers area in front of the ship while maintaining 5nm distance to the ship; b) Drone continues zigzag from the location of the sighting along a projected transect while the ship works the sighting. Once the ship resumes effort, the areas immediately in front of the ship and out to 5nm ahead of the ship were covered already by the drone. Projected transect here is the course of the transect from the location of the sighting.

5.3.3 Real-time monitoring and review

Lead and junior drone observers conducted real-time monitoring of the drone footage in teams of two with at least one lead observer present at any time while the drone was in the air. Drone observers tracked the progression of the drone along its path in the map view of the drone-WinCruz software, which was a version of WinCruz specifically modified for the trial survey to monitor drone and ship tracks simultaneously. This software allowed drone sightings to be entered, as well as flying bridge sightings, to evaluate the spatial proximity of potential duplicate sightings made by the two platforms. Drone observers also kept a written log of notes relating to effort and objects of potential interest detected in real-time in the video, i.e. potential schools of cetaceans. Most of these objects of potential interest were reviewed in the same evening, and all were reviewed again in St Andrews.

5.3.4 Video analysis

During the trial survey, the Seahawk drones recorded about 69 hours of video footage. Due to the amount of footage recorded in these types of studies, it is generally not practical to enlist a human observer to review the entire footage again post-survey in order to make detections of cetacean schools and to count individuals within each detected school recorded in the video footage. Thus, the aim of this project was to develop image and video analysis methods for these purposes. Specifically, a novel two-pronged approach was developed to address this problem. The two-pronged approach used both the still frames that the video is composed of and the video itself to calculate the velocity of objects or shapes within these frames. The image and video analysis component of this project can

be split up into two main sections: 1. dataset generation, and 2. design and implementation of supervised³ and unsupervised⁴ machine learning models.

5.3.4.1 Video Data

Departures by the drone company from the survey protocol on video capture and storage led to suboptimal imagery. The camera on board the drone recorded video at a resolution of 1920 x 1080 pixels. However, the drone then streamed the video data to a computer on the ship, instead of saving it to disk on the drone, contrary to specifications in the trial survey protocol. This process of streaming the video from the drone to the ship caused a vast reduction in video quality due to the video compression required to stream the video. The video was then captured via screen recording on the computer on the ship, instead of saving straight to disk, also contrary to specifications in the protocol. This again vastly decreased the quality of the video because the video was compressed as it was captured.

In addition to these two layers of video compression, the screen capturing process introduced several other artefacts. The first was the toolbar added to the right of every frame in all the footage (Figure 8). This meant that a portion of the video was lost for analysis. Next, occasionally a mouse cursor (Figure 8) or the computers desktop was captured, which again meant a loss of some or all of the image for analysis for some periods of time. Additionally, the computer's screen resolution was 2040 x 1036 pixels, which meant that the video footage was stretched, causing the computer to use interpolation on the video to approximate the extra pixels. Finally, another consequence of screen recording was that the original frame rate of the video captured by the drone was lost, meaning that the motion data was not as accurate as it would have been were the footage captured and saved to disk on board the drone. To illustrate the problem caused by this compression, distortion and interpolation, Figure 9 shows what a dolphin looked like when the section of the still frame containing the dolphin was enlarged. Human expert observers had difficulty identifying dolphins when viewing enlarged image section without the context of the larger image or video footage.

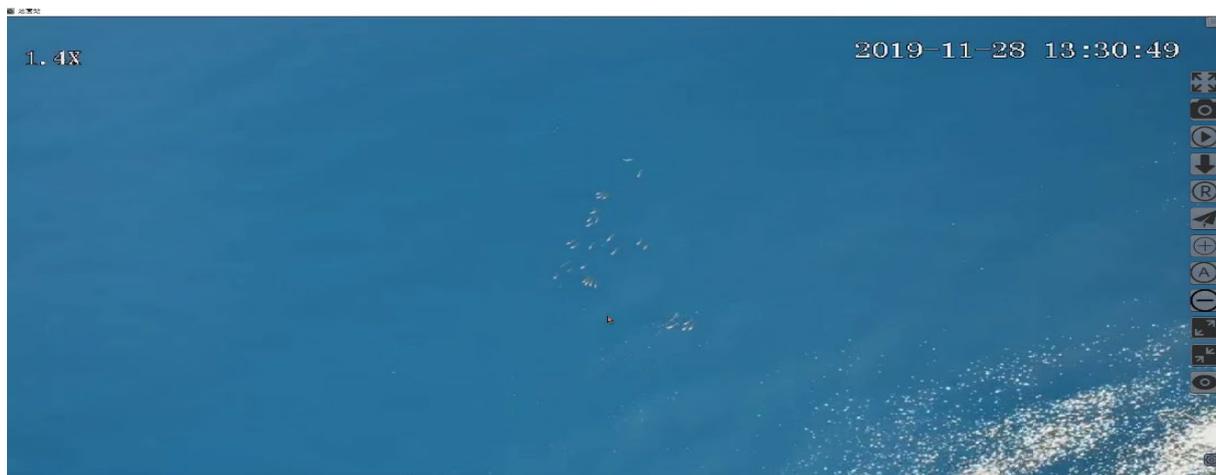


FIGURE 8. Example of still frame from drone video footage. The top left corner shows the current zoom level. Top right shows the date and time. Toolbar on the right side of the image is due to the

³ Supervised learning is a machine learning task where the machine determines a predictive model based on data with known outcomes.

⁴ Unsupervised learning is machine learning task where the model looks for previously unseen relationships in data with little to no human supervision.

video being screen recorded. Below the larger cluster of dolphins in the middle of the image, a mouse cursor is present.



FIGURE 9. Example of the quality of the data demonstrating the poor pixel resolution. The enlarged section on the left shows the dolphin enclosed in the green box within the large frame. Without context of the large frame (or the video, not shown here), the object in the enlarged section is almost impossible for a human observer to identify accurately as a dolphin and it is entirely impossible to identify its species.

5.3.4.2 Dataset Generation

As the supervised portion of the machine learning model required labelled⁵ data, we first needed to generate a dataset that could be labelled by an expert human observer (see Section 5.3.4.2.1 for details). A subset of the overall video footage was chosen based upon the drone observer notes, which detailed the types of animals observed, and when they appeared in the drone footage. Video segments were chosen so that they included some instances of all animals observed (dolphins, whales, turtles, birds), objects (trash, boats, logs), and different sea conditions. This was done to ensure that the final model would not confuse any other object with the target, the dolphin. Still frames were extracted from these video segments, with one frame taken every 25 frames (video frame rate is approximately 25 frames per second, so this equated to one frame every second). This process yielded around 8,000 frames for dataset generation. The frames were then fed into a computer vision algorithm specifically designed for this project. Figure 10 shows the main steps of this algorithm.

The first stage of the algorithm was to open the image and then acquire the magnification and altitude data from the still imagery and other data sources. The video equipment recorded a magnification factor (zoom) in each frame, and the drone team recorded the drone's GPS position and altitude while recording the video footage. A script was created to parse the GPS and altitude information, and assign

⁵ Labelled data is data that has been tagged or classified with labels that identify certain properties or features or contained objects. Here it is used to classify groups of pixels.

a GPS position and an altitude to each frame. Due to the poor resolution and the distortion of the video footage, it was not feasible to use simple Optical Character Recognition (OCR) (Mori *et al.* 1999) to obtain the magnification number (see upper left corner of Figure 8) in each still image. Therefore, a simple neural network was trained on the well-known handwritten numbers dataset, MNIST (LeCun, Bottou *et al.* 1998), and then used to read the magnification number in the video footage.

In order to be able to screen all frames, a relationship between magnification and altitude, and dolphin size was obtained from still images with confirmed dolphins. The measurements of confirmed dolphins were obtained with ImageJ (Rasband 2008) software using the line measurement tool to measure from beak tip to end of the tail. These measurements were carried out on frames which contained several dolphins that were readily observable to ensure that any errors in the measurements were minimised. A straight line was fitted to the data which related altitude and magnification to dolphin length (Figure 11).

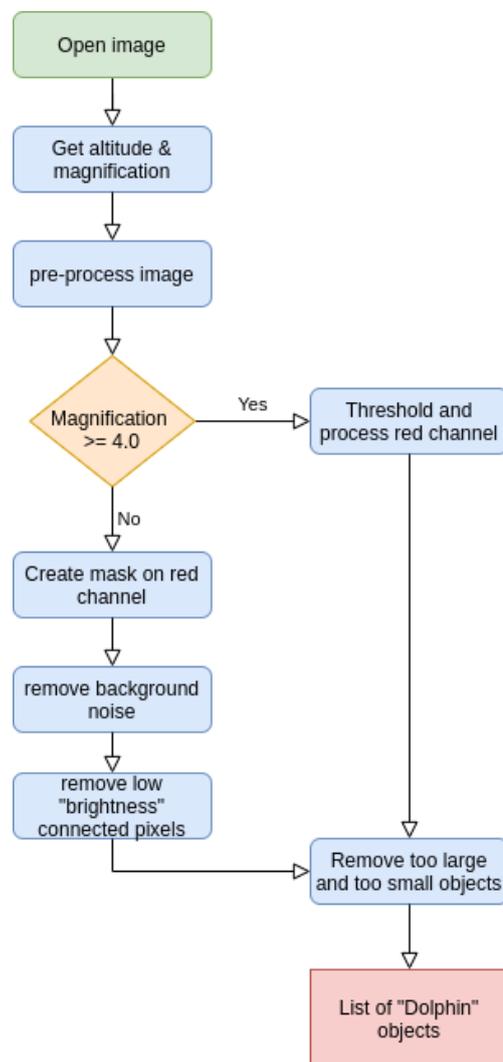


FIGURE 1. Overview of computer vision algorithm for the generation of dolphin objects for labelling by a human expert.

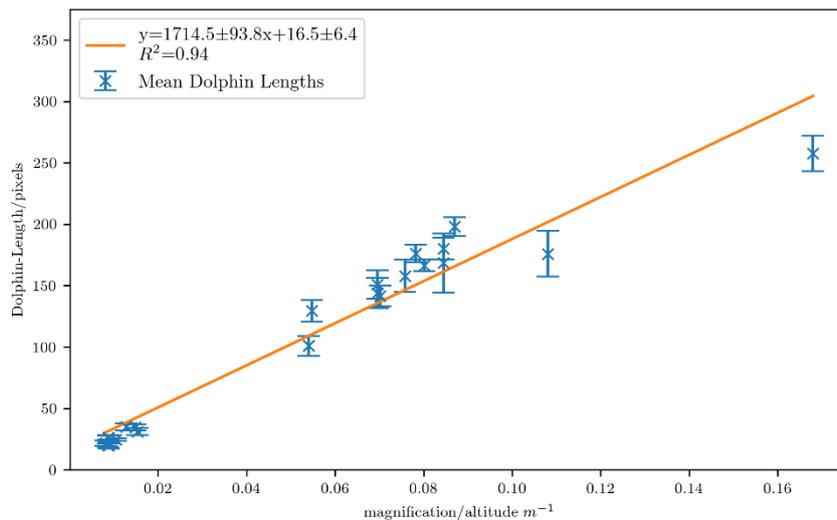


FIGURE 2. Straight line fit for relating dolphin length to magnification and drone altitude. Mean measured dolphin lengths for a given magnification/altitude are shown as stars (mean), whiskers reach out to 1 SD from the respective mean. The straight line fit yielded the equation shown in the legend, where y is the measured dolphin length, and $x = \text{magnification} / \text{altitude}$.

The next step in the algorithm was to pre-process each of the 8,000 frames. This involved cropping the image of each frame to remove the time, magnification and toolbar (Figure 8).

Once the images were pre-processed, they were further analysed to identify *dolphin* objects. The type of analysis depended on the magnification level because it was found through experimentation that for higher levels of magnification a different method performed better than the method used for lower levels of magnification. If the magnification level was greater than or equal to 4x, only the red channel of the frame was analyzed, as cetaceans appear redder when compared to the sea (Marie, Mejias *et al.* 2013). The red channel was thresholded which yielded a set of *possible-dolphin* objects from the video footage with magnification greater or equal to 4x. If the magnification was less than 4x, some further pre-processing was carried out to remove noise, and to mask areas where sun glare was prominent. This yielded a set of *possible-dolphin* objects from the video footage with magnification of less than 4x. These *possible-dolphin* objects were then processed to remove low brightness pixels that were not connected to the object or fell below a certain brightness threshold, thereby removing non-dolphin objects from the dataset. The final step of the algorithm was to reduce the number of objects that would be incorrectly classified as dolphins during analysis of the imagery by taking into account the size of *possible-dolphin* objects. To this end, the expected size of a dolphin within each frame was estimated from the altitude and magnification data. This allowed a comparison of the expected size with the size of the objects that were identified by the algorithm as possible dolphins within these frames. These would then be removed if they were too small or too large to be a dolphin, thereby reducing the number of objects that would be incorrectly classified as dolphins during analysis of the imagery.

This algorithm yielded a set of 248,879 *possible-dolphin* objects for labelling (classification) by a human expert observer, of which about 10% were labelled (see following subsection).

5.3.4.2.1 Manual object classification

A set of 20,195 objects, from the 248,879 *possible-dolphin* objects selected by the computer vision algorithm from the drone videos (Section 5.3.4.2), were reviewed by a human observer. The objects were of different sizes (i.e., varying number of pixels), potentially included one or more dolphins, and

showed the full extent or only a section of the individual(s). An observer manually reviewed the selected objects by viewing them individually using a graphical user interface (GUI) designed specifically for this project. The GUI contained a set of options that facilitated the classification of each object. Specifically, for each *possible-dolphin* object, the GUI presented the observer with three panels. The first panel showed the full still frame in which the *possible-dolphin* object was detected with the object surrounded by a bounding box. This panel was very useful for context, e.g. to see if other dolphins were nearby. In a second panel, the object was enlarged, which was useful for identifying the shape of the object. In the third panel, a short video sequence was provided, which contained the full still frame in the middle of the sequence. The video sequence proved to be very useful for classification as it allowed the user to follow the movement of the object through a series of frames.

Eleven classes were defined for object classification. Classes were based on only one type of wildlife being captured within the box surrounding the object, e.g. *dolphin* (or *multiple dolphins*), or *bird*. In cases where more than one classification was in theory possible, cetaceans took priority over non-cetaceans and dolphins took priority over whales. For example, if one dolphin was captured swimming under a log, the object was classified as *dolphin*. The ability to classify an object depended on the altitude of the drone, the magnification level of the drone camera, the quality of the footage and the environmental conditions during which it was recorded. A conservative classification approach was adopted, i.e. objects were only classified as *dolphin*, or other specific animal or object when it was possible to do so with certainty, and one of the unknown categories was used otherwise. The most important division in view of the machine learning models was between *dolphin*, *multiple dolphins*, *whales* on the one hand and the other classes on the other hand. Classes were defined as follows:

Dolphin - This class was used when one dolphin could be identified as dolphin with certainty within the box. This included all delphinid species. If small sections of other individuals were present within the box they were to be ignored. The object could be an entire animal or a section of it; the dolphin could be underwater, surfacing or completely airborne (e.g. when leaping or spinning). As long as one dolphin could be identified, this dictated the object classification as *dolphin*. This also applied to jumps when the dolphin could be seen among the splash, regardless of the amount of water captured.

Multiple Dolphins - This class was used when two or more dolphins could be clearly identified as dolphins, or the entirety or large sections of more than one dolphin were present within the box. This included all delphinid species. The object could be entire animals or sections of them. The dolphins could be underwater, surfacing or airborne. Dolphins could be close together or apart; they could be overlapping making it look like one continuous animal (either aligned or side by side).

Whale - This class was used when one or more whales could be clearly identified as whales within the box. The object could be entire animals or sections of them. The whale(s) could be underwater, surfacing or airborne (breaching). No dolphins were to be present within the box or else the object would be classified as *dolphin*; other wildlife, on the other hand could be present.

Bird - This class was used when one or more birds could be clearly identified as birds within the box. The birds could be flying, diving or sitting on the water, on other wildlife or on an object floating at the surface. In the case that a dolphin or whale was also visible within the box, the object was to be classified as *dolphin* or *whale* instead, respectively. In the case that a bird was sitting on a turtle, the object was to be classified as *turtle*. In the case that a bird is sitting on a log, the object was to be labelled as *bird*.

Turtle - This class was used when one or more turtles could be clearly identified as turtles within the box. The object to classify was generally the entire animal(s). The turtle(s) could be underwater or surfacing, swimming or resting.

Fish - This class was used when one or more fish (e.g. shark, tuna) could be clearly identified within the box. The object to classify could be entire animal(s) or sections of such. The fish could be underwater or, less likely, surfacing or even airborne. No cetaceans were to be present within the box.

Boat - This class was used when a ship, boat or a section of either, could be seen within the box. No wildlife was to be present.

Rubbish/Log - This class was used when a floating object could be identified (e.g. log, plastic object). No wildlife were to be present.

Unknown - This class was used when nothing in particular could be identified but the presence of a dolphin or other cetacean could not be excluded within the box. This may be the case when, for example, a dolphin had created a splash but it is not clear if the dolphin was still within the box. Other examples include when there was not enough definition in the still frame or video sequence to make a classification (e.g. the image was blurry or pixelated).

Unknown, not cetacean - This class was used when cetaceans were definitely not identifiable within the box (e.g. the image was excessively pixelated or blurry, or too dark).

Water - This class was used when nothing in particular could be identified within the box and only water features were observed, such as glare, waves or whitecaps. This could also include splashes created by dolphin jumps where the animal was no longer associated with the splash, and only water/splash can be seen within the box.

5.3.4.2.2 Regrouping of classes

The majority of objects (15,841) were labelled as *water*, the second most frequently assigned class was *dolphin* (2,704 objects), while five classes were only assigned five times or less (Figure 12), indicating that the labels were greatly unbalanced⁶. This typically results in machine learning models that predict poorly when applied to new (unseen) data. Therefore, the 11 classes were regrouped into two classes: *dolphin*, and *not dolphin*. The *dolphin* class was made up of the *dolphin* and *multiple dolphin* classes. The *not dolphin* class was made up of the remainder of the classes. The binary classification problem was easier than the multi-class problem, as there was more data for each of the labels, allowing the supervised machine learning models to better learn from the data, and therefore better predict on unseen data.

⁶ An unbalanced dataset is a dataset where one or more labels is significantly underrepresented or overrepresented in the dataset. For example, a dataset containing the labels *dog*, *duck*, and *cat* which have 10, 1000, and 34 entries respectively. This dataset would be said to be unbalanced as the *duck* label has significantly more entries than the other labels.

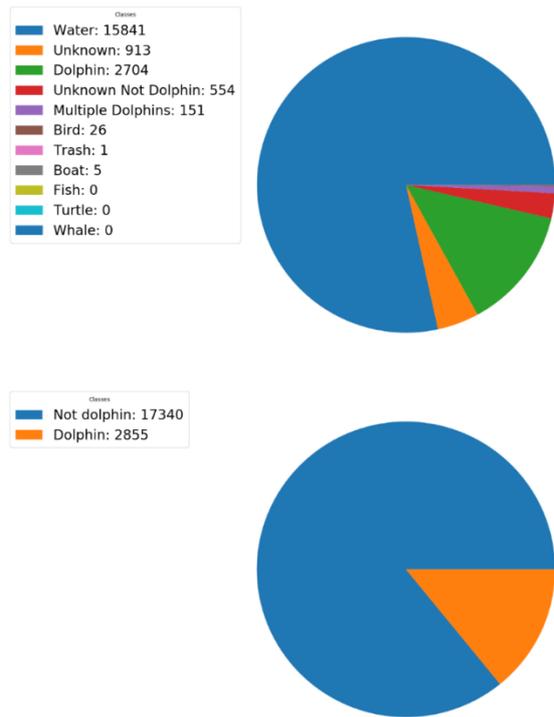


FIGURE 12. Distribution of classes from the labelled dataset: all original 11 classes (top pie chart); and, re-classified to the binary class case (bottom pie chart), where dolphin and multi-dolphin are one class, *dolphin*, and all other classes are in the *not dolphin* class.

For both the supervised and unsupervised machine learning models, the dataset needed to be split into train, validation and test sets (Table 4). This split was required as the supervised machine learning models needed training data to train the models. The validation data was for tuning the model to find the free parameters (hyperparameters) that give the best performance. Finally, the test data would serve to assess the models performance on unseen data at the end of the training/tuning cycle. However, at the time of the report, models were not sufficiently tuned to analyse the test data set. Hence, only preliminary results for the validation data set are presented here. A ratio of 60%/20%/20% was used for splitting the labelled dataset into training/validation/test datasets, respectively.

TABLE 4. Percentage of dolphins and the ratio of dolphin to not dolphin objects, in the training, test and validation data subsets.

Subset	Percentage of <i>dolphins</i>	Ratio <i>dolphins/not dolphins</i>
Train	12.1%	1,602/11,688 \approx 0.14
Validation	16.4%	566/2,885 \approx 0.20
Test	15.5%	536/2,918 \approx 0.18

To ensure that no data was leaked from the training data set into the validation or test datasets (which could cause an artificial performance boost), the full dataset was first divided by assigning the different videos, i.e. from a specific flight on a specific date, to one of the three groups. This meant that the still frames from a single video were found only in one subset, e.g., still frames from video XYZ were only in the training subset whereas still frames from the video ABC were only in the validation subset. For each subset the ratio of *dolphin* to *not dolphin* labels was kept as close as possible to the ratio in the

full data set, i.e. about 13% of the labels were *dolphin* (Table 4). Each of these data subsets were then used to train, tune and test the supervised and unsupervised machine learning algorithms.

5.3.4.3 Design of supervised and unsupervised machine learning models

From an early age, humans are adept at recognising and labelling moving objects. However, despite how simple it may be for humans, training a model to recognise and correctly label an object in an image remains an open problem (Zhang *et al.* 2019). Therefore, to detect and count the number of dolphins in a given frame, a combination of image data and motion data derived from the video footage were used to create an overall model that can detect dolphins from video footage. The aim was to broadly replicate the way that primates detect and identify objects with their visual sense (Hubel and Wiesel 1968).

There are two different problems we can potentially solve with an automatic detection model: classification and object detection. The classification problem essentially consists of showing an image to the model and the model returning a positive or negative response as to whether it thinks the one or more objects of interest (i.e., one or more objects of the “target class”) are present in the image. The classification problem is computationally much easier to solve than the object detection problem. The object detection problem is the same as the image classification problem, but with the added task that the object must be localised within the image (see Figure 13). Therefore, object detection models are naturally more computationally complex and, as a result, harder to train. We focused on the first problem, classification, to assess whether it was feasible to create an automatic dolphin detection model for this dataset. The second of these problems, object detection, was beyond the scope of this report.

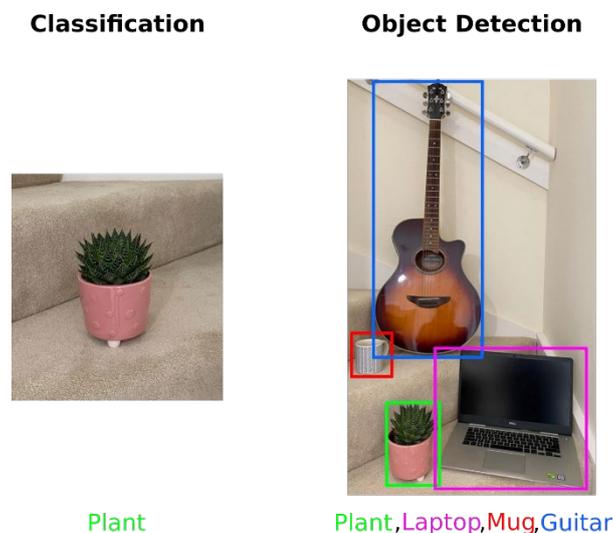


FIGURE 13. Example of classification *versus* object detection. The left image shows an example of a classification task. The right image shows an object detection task. The object detection task involves first identifying the object in the image then drawing a box that fully encapsulates that object.

For the classification problem, Convolutional Neural Networks (CNNs) (Geron 2019) were used for the still image data, and several unsupervised and supervised machine learning models were used for the motion data: Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) (McInnes *et al.* 2017), and Support Vector Machines (SVM) (Geron 2019). For both of these parts of the model, each image was cropped to just the region of interest (ROI). The ROI was defined as the area which included the *possible-dolphin* objects detected by the computer vision algorithm described above and depicted in Figure 10. These ROIs tended to be small, on the order of 350 pixels squared (pixels²), with sides of ≈ 15 pixels. The models were trained on and predicted on these ROIs.

Before discussing the various machine learning models implemented, we must discuss how to measure the performance of these models. Several metrics were used to track the performance of the algorithms as they train, including confusion matrices, accuracy and balanced accuracy. The latter two metrics are derived from entries of the confusion matrix.

A confusion matrix is composed of the number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) for each class, displayed in the form of a matrix. In the two-class classification problem, dolphins (positives) *versus* not dolphins (negatives), the diagonal elements are the TP and TN entries, which are the objects that were correctly classified either as *dolphins* or *not dolphins*. FP are those objects that were falsely identified by the models as *dolphins* when they are *not dolphins*, and FN are those objects that were falsely identified by the algorithms as *not dolphins* when they were *dolphins*. If the classification models worked perfectly, the off-diagonal entries, the FP and FN, would be 0.

Accuracy was defined as the ratio of TP to the total number of predictions for all classes, be this within the *dolphin* class or within the *not dolphin* class. Balanced accuracy is the mean of the TP rate ($TP/(TP+FN)$) and the TN rate ($TN/(TN+FP)$).

5.3.4.3.1 Image analysis - Supervised Learning

CNNs were first developed for studies of the brain's visual cortex and have been used in image recognition since the 1980's (Geron 2019) (Hubel and Torsten 2004). CNNs have been used to great effect in medical image analysis to detect tumours (Moshen *et al.* 2018) for facial recognition (Bartlett *et al.* 2004), and for animal detection (Maire *et al.* 2014) (Norouzzadeh *et al.* 2018).

In brief, CNNs work by learning what filters to apply to an image to generate feature maps of the image that best describe the object within the image. These feature maps represent the features of the object. For instance, when a human identifies an animal such as a dolphin they may look for the fins, beak, tail, etc. A CNN will learn these and possibly other features by using various filters. The CNN then feeds these features to the neural network portion of the model. A neural network is a network of artificial neurons that were originally designed to try to mimic the brain (Geron 2019). The different artificial neurons in the neural network then “fire” depending on the input features. The network as a whole will learn which neurons fire for which inputs and eventually can classify the input image (Geron 2019). Figure 14 shows an example of a CNN.

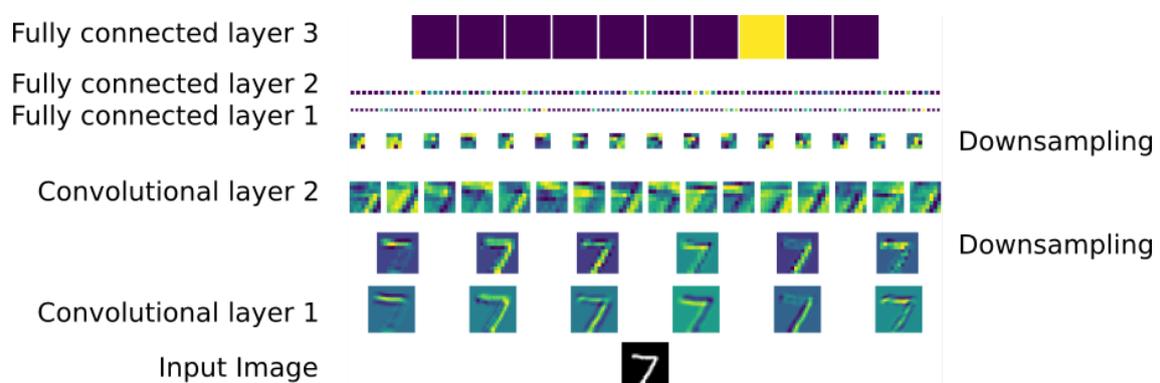


FIGURE 14. Example of a CNN classifying a hand drawn number. The image of the number is fed into the CNN (bottom layer). The convolution layers generate the feature maps which are then shrunk by the max pooling layers before the features are fed into the fully connected layers (neural network). The fully connected layers then classify the image correctly (result layer, not shown).

Several state-of-the-art CNN models, which are readily available and can come pre-trained⁷ on ImageNet (Deng *et al.* 2009)⁸, were trialled for this study: VGG16 (Simonyan and Zisserman, 2014), Densenet (Huang *et al.* 2017), and Resnet (He *et al.* 2016). The performance of these CNNs was compared using the metrics described above and it was found that Densenet gave the best accuracy on the validation dataset. The hyperparameters of DenseNet were then tuned to further improve the performance.

Despite the further tuning of hyperparameters of the DenseNet model, the model's performance on the test data set was relatively poor (Figure 15). In particular, whilst almost all the dolphins were identified (only about 12%, 67/(507+67), missed), there were too many FP; approximately 68% of *dolphin* detections (= 1104/(1104+507)) were FP. This was likely due to how the model was trained with certain hyperparameters, and the poor quality of the video footage.

To increase the overall accuracy of the model, data augmentation and biasing of the weights in the loss functions was carried out. Data augmentation was achieved by using a weighted sampler and several transforms on the frames. The weighted sampler preferentially chose the *dolphin* class over the *not dolphin* class when choosing samples to train the model. In practice this meant that the unbalanced dataset became balanced by oversampling the underrepresented class (*dolphin*). Several transforms were randomly applied to each image before it was fed through the CNN. These transforms included colour jittering, random rotation (0 - 180°), and random horizontal flipping. Finally, the weights in the loss function⁹ were changed to more heavily penalise misclassification of *dolphins*.

Using data augmentation and biasing the weights of the loss function, a balanced accuracy of approximately 75% was achieved on the validation dataset (equal to the mean of 507/(507+67) and 1762/(1104+1762), Figure 15). By comparison, without these modifications, the model only achieved a balanced accuracy of 50% (not shown).

⁷ Due to the complexity of some CNN models, pre-training them allows better performance when it comes to using the CNN on a new problem (He *et al.* 2019).

⁸ ImageNet is a large database of over 14 million images with more than 20,000 categories.

⁹ The loss function (or cost function) is the function that is minimized by the CNN, whilst training.

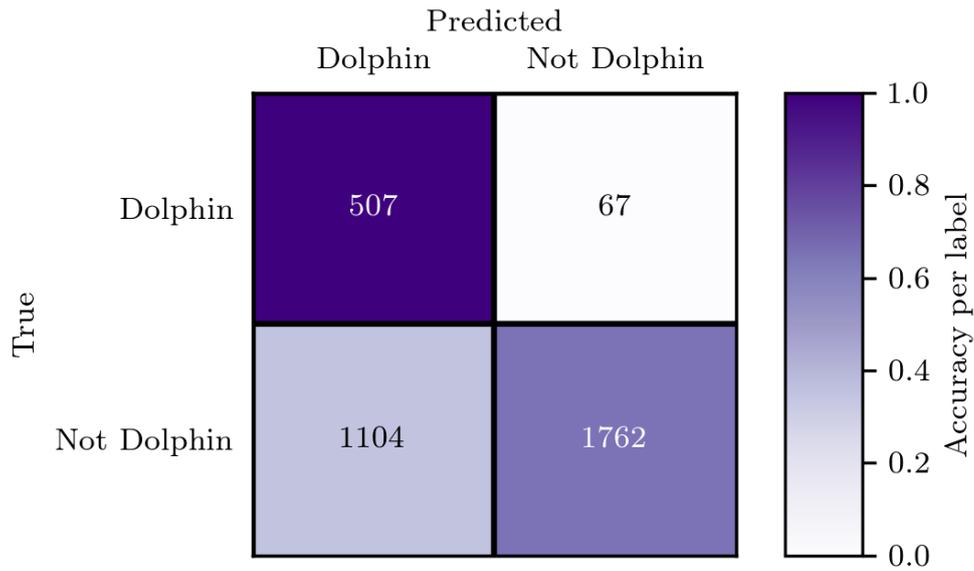


FIGURE 15. Confusion matrix for Densenet CNN model run on the validation data. The TP (507) and TN (1,762) are the diagonal entries. The off diagonals are the FP (1,104) and FN (67). The Densenet model was biased towards classifying the images it received as input as *dolphins*.

5.3.4.3.2 Motion analysis

For the motion analysis model (see Figure 16), it was first necessary to convert the motion of objects in the video into a measure of velocity. This was achieved by using the optical flow technique which calculates the velocity of objects by tracking the movement of their boundaries (e.g., corners, edges) between consecutive frames (see Figure 17). More specifically, the Lucas-Kanade (Lucas and Kanade 1981) optical flow technique was used. This method is more efficient at calculating the optical flow of objects as it includes tracking of pixels. Applying the optical flow technique yielded a velocity in pixels per second (p/sec), which was converted to meters per second (m/sec) using the previously described length scale from the GPS and magnification data (see Section 5.3.4.2).

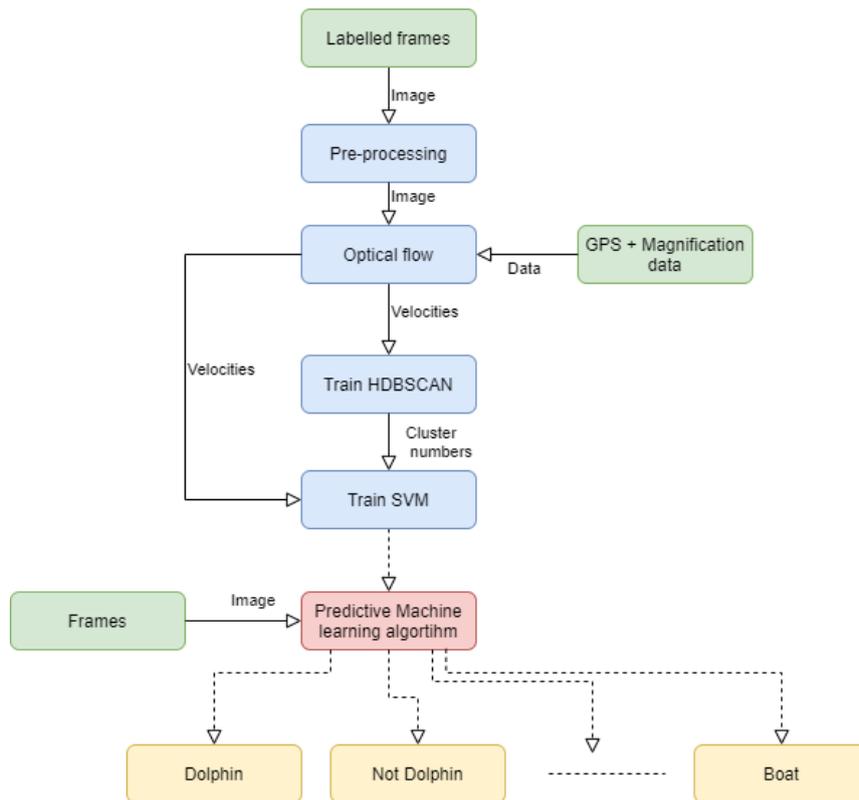


FIGURE 16. Overview of the motion analysis model. Green boxes indicate input data, blue boxes relate to training, the red box indicates the final predictive model, and the yellow boxes are the output labels.

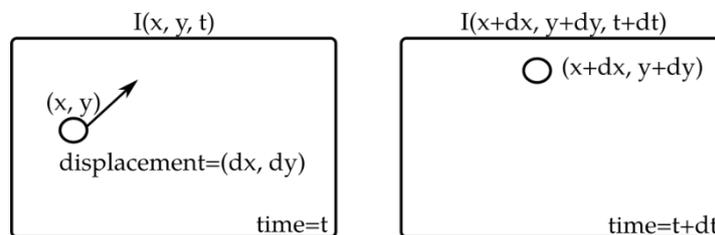


FIGURE 17. Example of optical flow between two still frames. The optical flow tracks the moving object (circle) by tracking its pixel intensity, I , as it moves. The pixel intensity of the object in the left image $I(x, y, t)$ is the same in the right image $I(x+dx, y+dy, t+dt)$. Thus, we can calculate the velocity of that pixel, knowing the time step between the two frames.

Before applying optical flow to the video data, it was necessary to pre-process the video to remove noise and format the data correctly. The first step in this process was noise removal, which consisted of a background subtraction and then applying a threshold (Figure 18) which only captured those pixels which had significant motion compared to the rest of the frame (Shaikh *et al.* 2014). In the next step, contours were drawn around pixels with similar intensities (colour); contours were the required input for the optical flow algorithm.



FIGURE 18. Example of the pre-processing of one still frame before it was fed into the optical flow algorithm. The bottom image shows the results of background subtraction and thresholding applied to the top image.

The velocities calculated by the optical flow algorithm were unclassified and could belong to dolphins or other objects such as, e.g. waves or sun glare, or could represent apparent motion caused by the motion of the drone itself. To aid the machine learning model, we used an unsupervised clustering technique to cluster similar instances of velocities. The technique employed to this end, was Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) (McInnes *et al.* 2017). Here, all velocities from all analysed frames were first separated into high and low density clusters, where density referred to how densely points were distributed in velocity space. Next, HDBSCAN built a hierarchy of the initial density clusters to determine which density peaks merged or separated based on a given local threshold λ (see right panel in Figure 19). Finally, HDBSCAN used this hierarchical structure to determine the optimal way to cluster the input data. This hierarchical method makes HDBSCAN particularly tolerant to noisy data when compared to other clustering techniques such as K-Means (Gujunoori and Oruganti 2017).

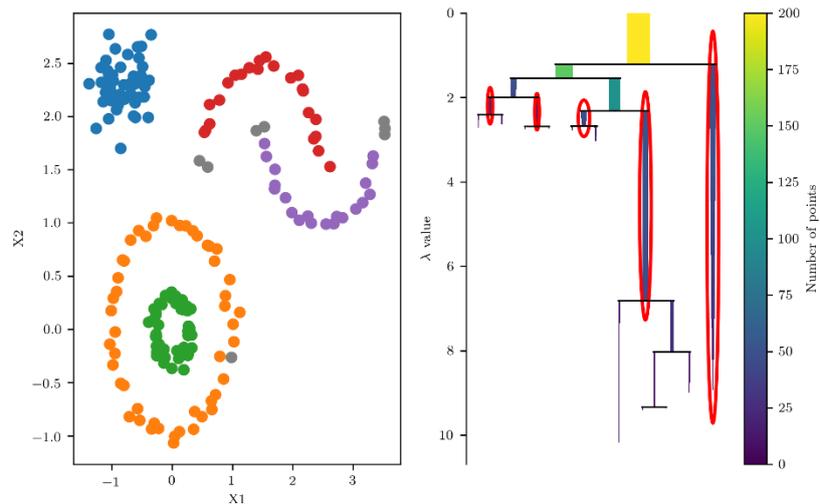


FIGURE 19. Example of the HDBSCAN algorithm clustering data. The left image shows the data where each cluster was given a different colour based on which cluster it belonged to. The right image shows the dendrogram for the data on the left. The five clusters were taken from parts of the dendrogram with the largest leaves. The largest leaves in this example are circled in red. Here, lambda (λ) is the local threshold value.

The HDBSCAN algorithm assigned a cluster number to each velocity measurement. However, the cluster number does not carry any readily accessible information, and it was necessary to train a supervised machine learning model to assign a label (*dolphin* or *not dolphin*) to these cluster numbers. Several methods including SVM (Geron 2019), random forests (Geron 2019), K-nearest neighbours (Altman 1992) were trialed to label the cluster numbers. It was found that SVMs gave the best performance, after extensive hyperparameter tuning (see below). To achieve this performance, a method which oversamples the classes which are underrepresented, and under-samples the abundant classes, was used in conjunction with SVM. The oversampling method also generated some synthetic samples as part of this process (Batista and Monard 2004).

SVM is a method capable of linear or nonlinear classification, regression, and outlier detection. The SVM achieves this by fitting a hyperplane in N-dimensional space (where, e.g. in two dimensional space, a hyperplane is a line) that distinctly classifies the data points and maximises the margin between the hyperplane and the data points. Figure 20 shows an example of an optimal hyperplane (H1) which separates the two classes of points with the largest margin, represented by the small lines perpendicular to H1. By comparison, while hyperplane H2 also separates the two classes, it has a small margin represented by the small lines perpendicular to H2. H3 does not separate the two classes.

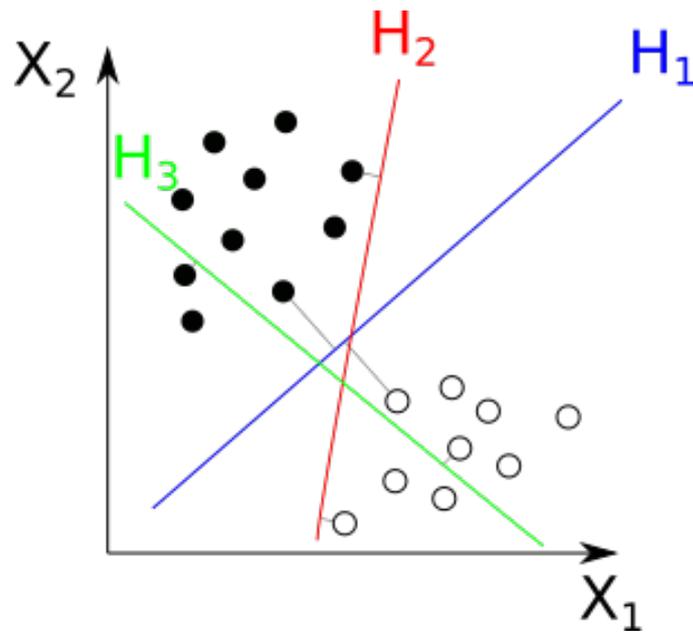


FIGURE 20. Example of classification using SVM including three hyperplanes, H1, H2, and H3, that could be used to define the two classes of points (black and white). H3 does not separate the classes. H2 does separate the classes but with only a small margin. H1 separates the classes with the optimal margin.

The confusion matrix from application of the HDBSCAN and SVM methods to the validation data (Figure 21) showed that the majority of dolphin objects were correctly classified as *dolphins*, only misclassifying about 14% ($80/(80+493)$). However, as for the image analysis (Figure 15), a large proportion, about 68% ($1053/(493+1,053)$), of the predicted *dolphins* were FP. Overall, the model achieved a balanced accuracy of about 75% (equal to the mean of $493/(493+80)$ and $1828/(1,053+1,828)$, Figure 21).

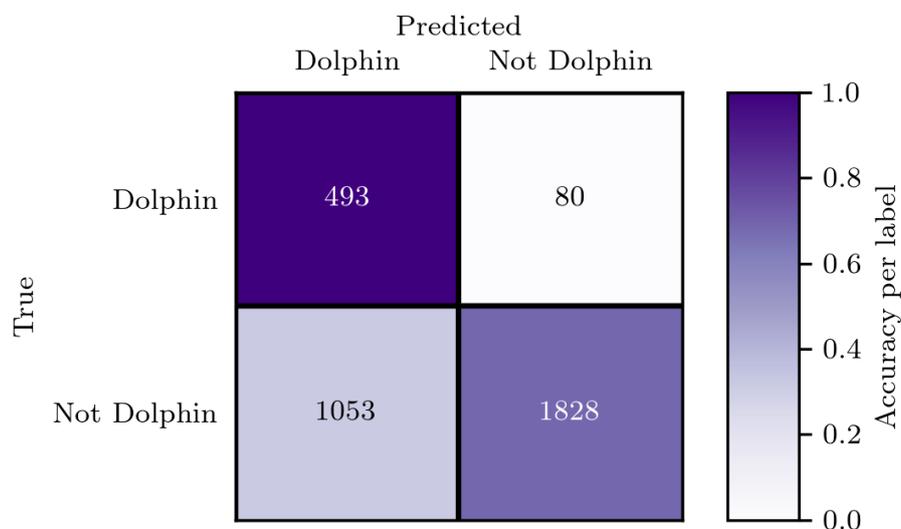


FIGURE 21. The confusion matrix for the motion model run on the validation data, with TP (493) and TN (1,828) on the diagonal, and FP (1,053) and FN (80) on the off-diagonal.

5.3.4.3.3 Combined model: “Triton”

As each of the two methods, the image and motion models, achieved a balanced accuracy of about 75% (Table 5), they were combined into one model that is referred to hereafter as ‘Triton’, after the

half-human half-fish Greek demigod of the sea (Britannica 2020). Triton follows the paradigm of multi-input and mixed data. Here the multi-input refers to the images that are fed into the CNN model and the velocity data that is fed into the motion model. The advantage of using this paradigm was that the overall model had more data to learn from, and thus was expected to have increased performance in comparison to that of the constituent models by themselves.

An overview of the Triton model is presented in Figure 22. The data pipeline for the motion analysis portion of the model remained unchanged. The image analysis part of the model was slightly changed in that the classification stage of the CNN was removed, and instead the model outputted several features (as described in Section 5.3.4.3.1). The features from the CNN model and the cluster label from the motion analysis part of the pipeline were concatenated (joined) together. The concatenated output was then fed into a small neural network (the fully connected layers in Figure 22). As described in Section 5.3.4.3.1, the neural network then classified each of the inputs into a class, in this case either *dolphin* or *not dolphin*.

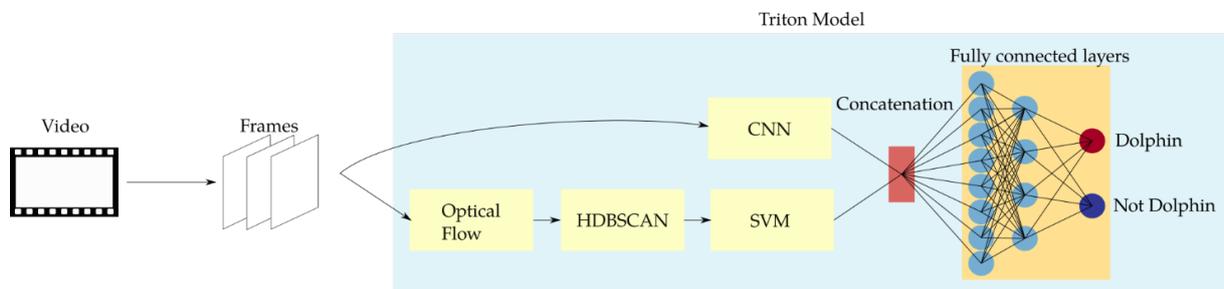


FIGURE 22. Diagram showing the Triton model, including the data pipeline. The video was split up into individual frames and fed into the model. The frames took two distinct paths through the model: the motion data analysis is shown in the lower path and the image analysis is shown in the upper path. These paths were joined at the concatenation stage, and then fed through the fully connected layers for classification.

The confusion matrix for Triton (Figure 23) showed a vast improvement over the individual models. Triton achieved an 8% increase in performance, giving a balanced accuracy of about 83% (equal to the mean of $454/(454+120)$ and $2,493/(373+2,493)$). This was accomplished via Triton’s vast improvement in the FP rate compared to the image model and the motion model (Table 5). However, this improvement came at a cost; Triton had a greater FN rate as compared to that of either of its constituent models. Thus, the Triton model was slightly biased towards classifying the images it receives as *dolphin*. Given that the purpose of the model was to classify dolphins in the imagery, this increase in the FN was acceptable.

TABLE 5. Balanced accuracy, given as a percentage, FP and FN for the three machine learning models: image, motion and the Triton model which combines image and motion analysis.

Model	Balanced accuracy	FP	FN
Image	≈75	1104	67
Motion	≈75	1053	80
Triton	≈83	373	120

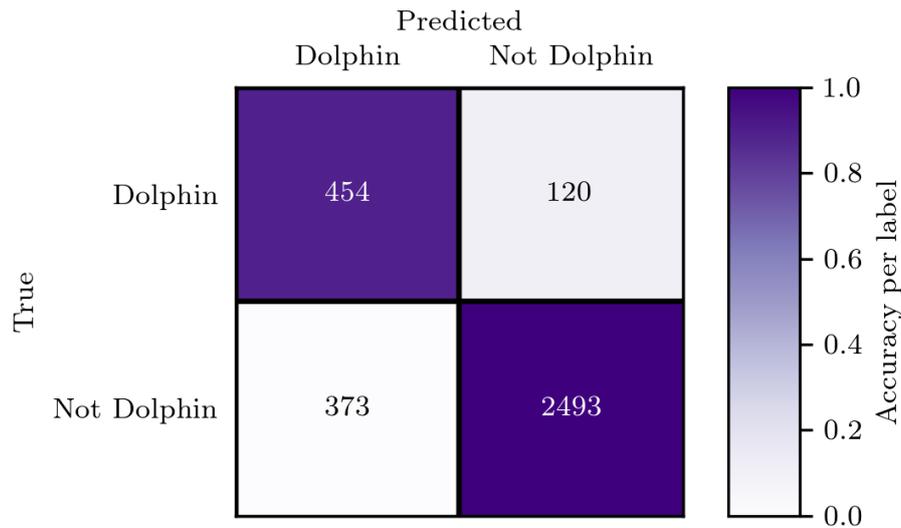


FIGURE 23. The confusion matrix for the combined model run on the validation data, Triton. The TP (454) and TN (2,493) are the diagonal entries and the FP (373) and FN (120) are on the off-diagonal.

5.3.5 Comparing real-time with image analysis detections

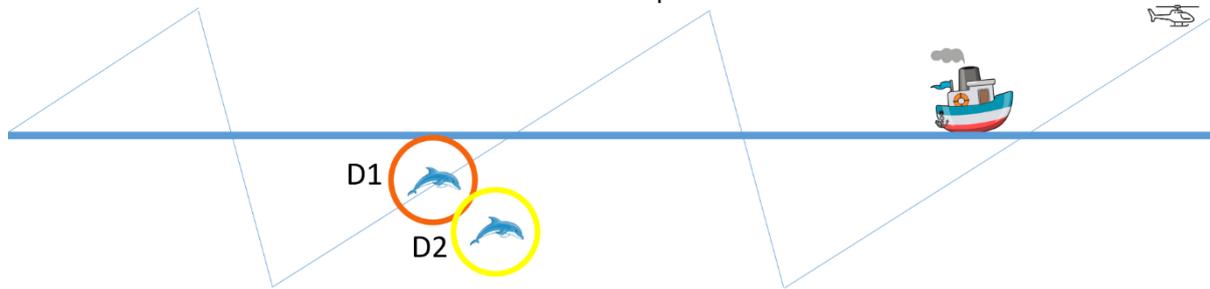
The efficiency of detecting schools of cetaceans via real-time observations, as compared with video analysis, can be explored by comparing the detections made by each method. As a first step, real-time observations were confirmed in a post-hoc review. This was necessary as drone observers often only logged in real-time approximate times, i.e. not to the exact second, and vague statements, e.g. object of possible interest (see Section 5.3.3). Using the video time of the frames with these detections and the location of the detected objects within the frames, would allow for a direct comparison between the two methods; however, due to the poor video quality, the video analysis models were not sufficient to make detections from the zigzag flight videos. Hence, a list of drone detections was compiled only using the detections confirmed as cetaceans during post-hoc review.

5.3.6 Matching sightings between platforms: flying bridge sightings with detections made via drone

This step of the analysis served to determine whether duplicate detections were made between the flying bridge and the drone. Duplicate detections were identified as such when it could be determined with a high degree of certainty that a given detection made via the drone was the same school as one of the flying bridge detections. For this purpose, each detection from the list of detections from the previous step (Section 5.3.5) was matched against the flying bridge detections. Information used for this purpose was primarily the separation in space and time between two candidate detections, one detection from the drone and one from the flying bridge, that were potential duplicates (Figure 24). However, the species id, group composition, behaviour and, if applicable, direction and speed of travel were also used, whenever available.

Matching drone detections with ship-board detections

Is D1 the same dolphin as D2?



Key information available:

D1: dolphin detected by drone at location X_1, Y_1 at time T_1

D2: dolphin detected by ship at location X_2, Y_2 at time T_2

FIGURE 24. Schematic representation of ship and drone survey tracks with two dolphin detections that are a potential duplicate detection between the two observation platforms.

5.3.7 MRDS analysis methods

As described in Section 5.3.1, the key for collecting MRDS data were the trials set up by observer 2 (drone) for observer 1 (flying bridge) (see Figure 25 for a schematic representation). The trials represented all detections made by via the drone. Out of these trials, the successes were those schools that were identified as duplicate detections, as described in Section 5.3.6. Those trials that did not match any detections made by the flying bridge were the failures. These data (trials with successes and failures), along with the perpendicular distances from the transect line completed by the ship, were the data required for MRDS analysis (Borchers *et al.* 1998). In comparison to conventional distance sampling methods where only the perpendicular distances are used for fitting a detection function (the DS model) and to estimate average detection probabilities, the trial data are also used to fit a mark-recapture (MR) model. The MR model is a binomial generalised linear model (GLM) fitted to the trial data using the perpendicular distances as a covariate. This model estimates the conditional detection probabilities of observer 1 detecting a school given that observer 2 also detected the same school. Depending on the level of independence between the two platforms, either the MR model or a combination of the MR and DS models is used (Burt *et al.* 2014). AIC can be used to determine the appropriate level of independence (Rankin *et al.* 2020).

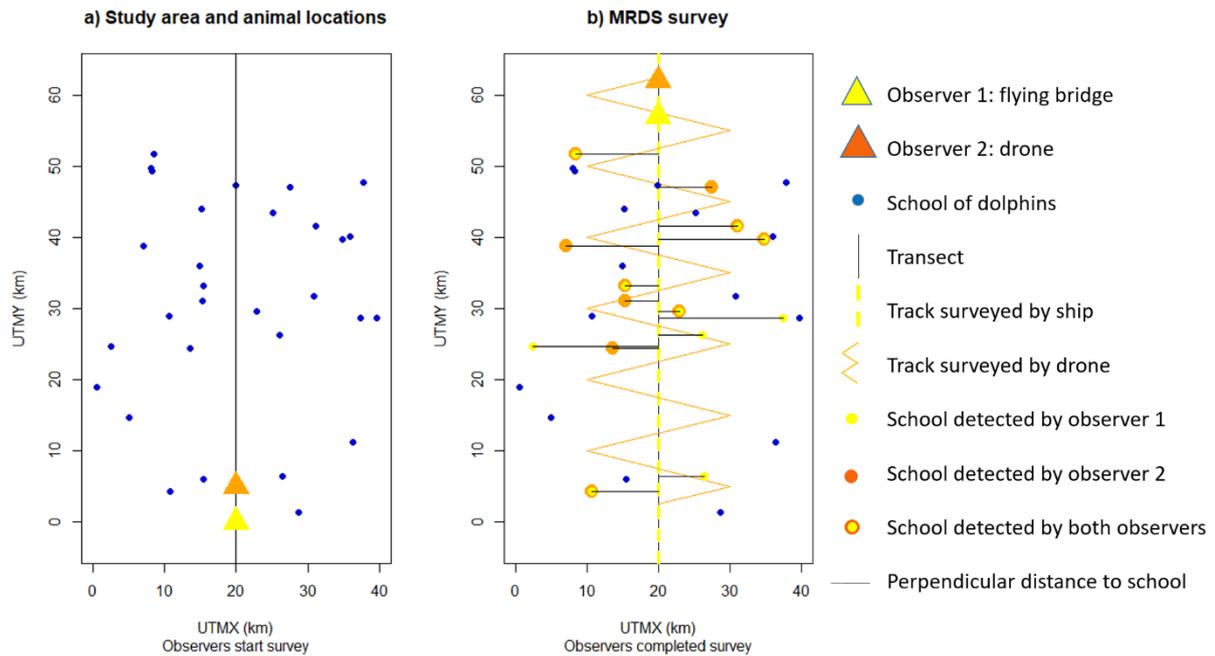


FIGURE 25. Schematic representation of an MRDS survey along a transect.

The additional difficulty for analysing MRDS data that was collected during closing mode effort (Section 5.1.2) was the disruption of simultaneous search that occurred when the flying bridge interrupted general search effort to approach a school before the observers had fully exploited all their chances to detect a different school that was detected by the drone. In the example shown in Figure 26, the ship turned to the right to approach sighting FB1 made by the flying bridge observers. At that point, the drone had already made two detections (D1 and D2). D1 was a potential duplicate with FB1 and would count as a full trial as the observers were on effort the entire time the school was within visible range (from when it entered the horizon until it passed the ship's beam). However, as the search effort of the flying bridge was interrupted to approach FB1 while D2 was within visible range, the chances of the flying bridge to detect D2 were reduced which needed to be accounted for in the analysis. Furthermore, when the ship resumed effort from the location of FB1, the perpendicular distance to D2 was different than before the turn. Methods to account for these issues do not currently exist and will need to be developed.

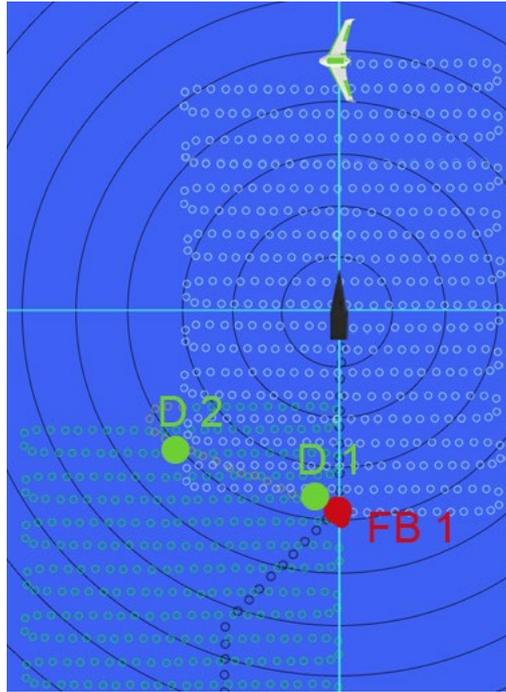


FIGURE 26. Schematic representation of ship track (black bubbles) and drone track (green bubbles) along with locations of schools detected by flying bridge (red) and drone (green). Black concentric circles represent distance rings around the ship at 1nm increments.

5.4 School size calibration

Estimating the number of individuals in a school of dolphins accurately poses a challenge when estimating dolphin abundance. Dolphin schools seen during ETP surveys vary in size from a few dolphins to several hundred animals (Gerrodette *et al.* 2019). For observers who are tasked with estimating school size, it is unknown how many animals are below the surface at any given time, making it difficult to obtain accurate estimates of school size. Marine mammal observers that have participated in previous NMFS surveys in the ETP have been calibrated by comparing their estimates against the true number of dolphins for a subset of schools (e.g., Gerrodette *et al.* 2019). This true number can be obtained from counts of dolphins in high resolution imagery (video or photo) of the school taken from above. In order to be able to capture the whole school and to identify each dolphin to species, it is required that the imagery is taken from a relatively high altitude that allows the whole dolphin school to be observed for some time while ensuring adequate ground resolution. To this end, the purpose of the calibration flights during the trial survey was to test if it was possible to use the Seahawk drone to obtain high resolution imagery of calibration schools from which true counts by species could be obtained.

5.4.1 Testing calibration flights

Flying bridge observers alerted the cruise leader whenever a potential calibration school was detected, i.e. a school of dolphins whose formation was a compact, single cluster; schools that consisted of several clusters or were spread out over a large area were not suitable calibration schools. To obtain imagery of a potential calibration school, the drone had to be flown out to the school, and either try to capture the whole school in a single frame or, if the school was too large, by slowly sweeping across the school.

To test the school size calibration protocol (Appendix 1), the following tasks were to be implemented by the drone pilots and engineers in coordination with the drone and flying bridge observers:

- 1) manoeuvre the drone to directly above a dolphin school seen by the flying bridge observers;

- 2) collect video of the entire dolphin school to be used for obtaining the true number of the school size and species identification of each dolphin within the school.

5.4.2 Real-time monitoring

The real-time monitoring for the calibration flights served a different purpose compared to the real-time monitoring conducted during the zigzag flights. Here, observers needed to both observe and direct the drone pilots where to fly, using the information received from the flying bridge observers. After a few test runs, it was found that the best constellation was if CSO and JCS served as the drone observers placed at the drone observer station in the lab, CLC communicated directly with the drone pilots at the drone pilot station outside, and 2nd pilot LYM was stationed on the flying bridge and communicated in Chinese with flying bridge observer SY (which helped bridge the language barriers between the drone and science teams). Once the school was located by the drone, the drone hovered above the school for several minutes in order to ensure each individual would be visible at the surface at least once in the imagery. In the case that the school was too large to be captured in one frame, several slow sweeps were flown across the school or, if the school was travelling, the drone hovered while the school passed underneath the drone. It should be noted that even if the real-time monitoring for the zigzag flights is to be dropped from the next phases of the project, real-time monitoring is essential for the calibration flights for finding the schools and ensuring suitable footage is collected.

5.4.3 Video analysis

To get a count of the dolphins in each calibration school, the video will be fed through the Triton model (once the object detection capability of this model is fully operational). The output of the model is a count of dolphins per frame. It is assumed that the model will not give 100% accurate counts, with counts fluctuating from frame to frame. Therefore, noise removal in the form of sigma-clipping will be employed to remove outlying counts. After the frames with the outlying counts are removed, the model output contains counts as a function of time for the entire video of the school. It will be possible to assess the size of the school from the distribution of these counts, and to provide a best estimate as well as upper and lower bounds on the estimate.

5.4.4 Manual counts

For the manual counts, we focused our efforts on those schools for which all clusters were captured by the drone footage and, hence, were potentially valid calibration schools, if true counts could be obtained. Manual counts were done by first identifying sweeps: sections of the video during which the drone did not double back over the school. Most calibration flights contained multiple sweeps for which counts could be compared, as long as the entire school was captured within the sweep. The aim of the manual counts was to obtain true counts of the school. Hence, these counts were done by identifying individual dolphins or small groups of dolphins swimming together and tracking these through the sweep, including those times when they were subsurface and out of sight.

5.4.5 Comparison of image analysis counts with manual counts

Once counts of dolphins are obtained with both methods, i.e. manually and video analysis, they can be compared in two ways:

1. Comparison of counts of dolphins in single frames
2. Comparison of best counts for a cluster or entire school across a frame sequence

The first method is straightforward as it only requires counting the visible dolphins in the respective frames. The second method is more complex as it requires accounting for those dolphins that are not visible at or near the surface throughout the entire frame sequence. This requires tracking individual dolphins or small groups of dolphins swimming together while they are submerged out of sight and

estimating the fraction of time that dolphins are at the surface. While this could be done manually, at least with some degree of certainty, models need to be developed to do this via image analysis. Using the estimated fraction of time dolphins are at the surface, image analysis or manual counts from the first method can then be adjusted to obtain estimates of true counts. Developing these methods is beyond the scope of this report.

6 RESULTS

6.1 Research vessel as a suitable platform to conduct line transect surveys using NMFS protocol

6.1.1 Flying bridge setup

All the required equipment for implementing the NMFS survey protocol was present, i.e. four sets of bigeye binoculars mounted on pedestals with reticle scales and angle rings, observer chairs, recorder station with canopy and wind dam. However, as the installation was never completed before the trial survey (see Section 5.1.1), several modifications will be required for the main survey to ensure the safety of the observers and successful completion of the survey. A full list of recommendations and requirements will be made available for the next phase of the project. Nonetheless, observers were able to implement the survey protocol as outlined in Section 5.1.2. We note that the binocular height on the flying bridge of the Jorge Carranza, at 20.44m, is much higher than those of previous ETP surveys. Binocular heights ranged between 10.4 and 10.7m (RV David Starr Jordan, Townsend Cromwell, McArthur I, and Endeavor) before the RV McArthur II with 15.2m was used as the second survey vessel in 2003 and 2006. A height of 10.7m results in a maximum ship-to-horizon viewing distance of 6.3nm, 15.2m in 7.5nm and 20.4m in 8.7nm. Whether this had an influence on detection probabilities was investigated by comparing detection probabilities of the trial survey with those from previous survey years (see below Section 6.1.7).

6.1.2 Observers

Due to the extensive experience of the flying bridge observers, the implementation of the NMFS survey protocol on the Jorge Carranza was successful. Details on the completed survey effort and flying bridge sightings can be found in Sections 6.1.4 and 6.1.5. The lead roles for the drone observer team were taken by CLC and JCS (and CSO for directing calibration flights) as a deep understanding of the requirements was essential during the testing phase. The IATTC observers were trained in both drone and flying bridge observing. For taking the lead in drone observing or taking a full position as a flying bridge observer, further training will be necessary. However, both IATTC observers were valuable team members which we recommend for the main survey.

6.1.3 Ship and crew

The captain and the other ship officers were very effective and helpful at implementing the survey protocol, including quick responses to requests made by flying bridge observers, maneuvering the ship in closing mode so school size and species composition estimates could be obtained, and assisting with planning and implementing changes to the transect lines due to weather and changes in data collection needs. In particular, the captain stood out in terms of his ship handling skills and willingness to accommodate the science and drone teams. We note that the Jorge Carranza has diesel-electric engines with three generators. For implementing the NMFS survey protocol with a cruising speed of 10 nm and rapid speed and direction changes, more than one generator was required. However, while maintaining a cruising speed of about 9.5 nm, only one generator is required, which is more

economical. Overall, the R/V Dr Jorge Carranza Fraser performed very well as a marine mammal survey vessel.

6.1.4 Survey effort

A total of 1,733.06 km of transect lines were surveyed during the 14-day trial survey, out of which 766.41 km were conducted in closing mode and 966.65 km in passing mode (Table 6). It was not the goal of the trial survey to complete all 16 transects; for conducting the required tests, it was more important to remain in high density areas. Hence, completed tracklines (Figure 27) deviated from the plan (Table 1).

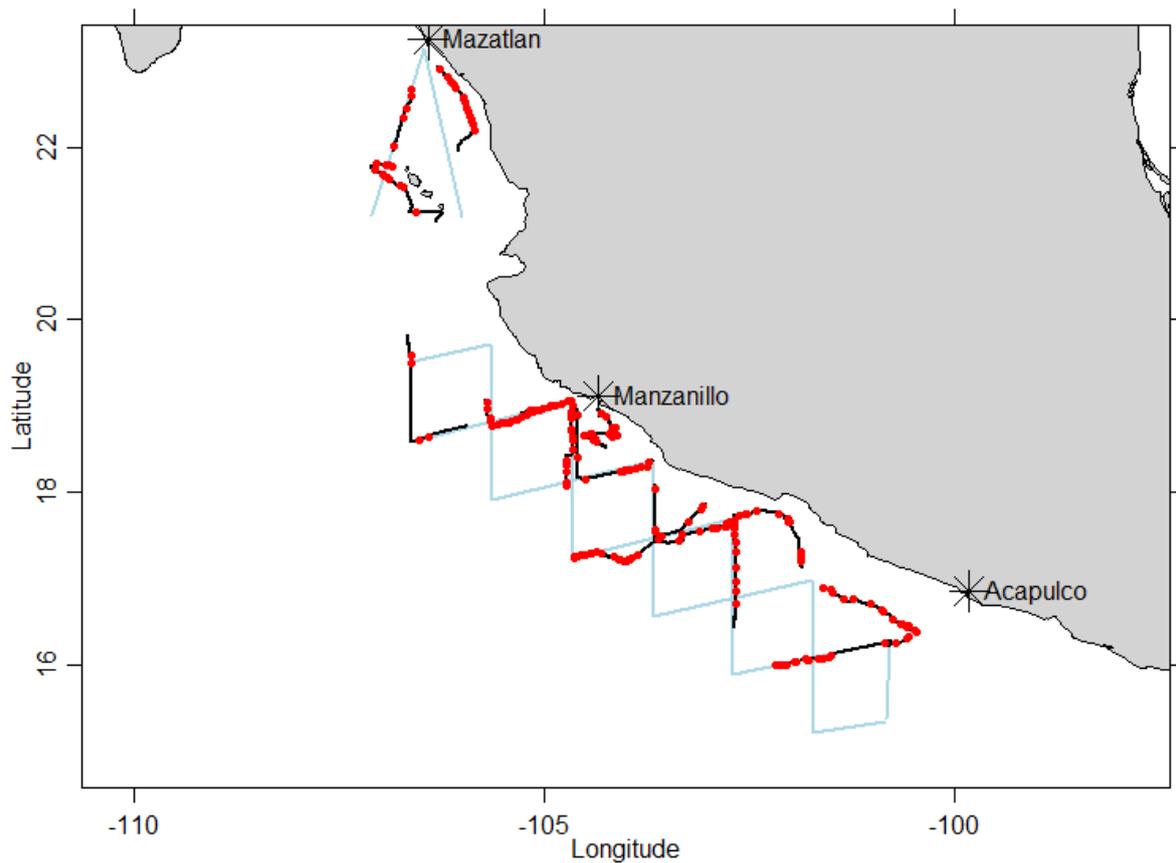


FIGURE 27. Completed survey effort (black) with flying bridge sightings (red), and planned tracklines (blue).

All survey effort should have been conducted in closing mode as that is the required mode for the main survey. However, after unsuccessfully attempting drone zigzag flights with the flying bridge operating in closing mode for two days (17-18 November), it was decided to switch to passing mode to facilitate the testing of the zigzag flights (see Section 6.2.2 for more details). It was not possible to implement the zigzag flights during closing mode effort as the flight duration of the Seahawk was too short. However, for testing calibration flights it was necessary to find good calibration schools, which was easier when the flying bridge was operating in closing mode. Hence, depending on the priority of a given day, the effort mode changed.

TABLE 6. Kilometers of survey effort conducted by the flying bridge per day. Effort only includes times when observers were actively searching and logged as on effort.

Date	Closing	Passing	Combined
17 November	46.98	0.00	46.98
18 November	189.86	0.00	189.86
19 November	0.00	130.28	130.28
20 November	8.03	176.00	184.03
21 November	11.00	112.02	123.02
22 November	128.89	0.00	128.89
23 November	108.51	0.00	108.51
24 November	59.53	0.00	59.53
25 November	44.28	45.69	89.97
26 November	67.65	95.71	163.36
27 November	70.11	99.52	169.63
28 November	5.93	97.16	103.08
29 November	14.39	135.79	150.19
30 November	11.26	74.47	85.73
Total	766.41	966.65	1,733.06

Viewing conditions during the 14-day trial survey varied with Beaufort sea states ranging between 1 and 5 and swell heights ranging between 1 and 6 feet (Table 7). Sighting rates of the flying bridge observers generally decreased with increasing Beaufort sea state or swell height.

TABLE 7. Flying bridge survey effort and sighting rate, in number of sightings per km effort, for each Beaufort sea state (left) and swell height (right) encountered during the trial survey.

Beaufort	Effort (km)	Sighting rate	Swell (feet)	Effort (km)	Sighting rate
0	0.00	.-	0	0.00	.-
1	158.94	0.28	1	30.06	0.23
2	471.32	0.14	2	257.04	0.15
3	424.86	0.10	3	707.84	0.14
4	564.55	0.07	4	378.24	0.08
5	113.39	0.08	5	272.88	0.07
6	0.00	.-	6	87.01	0.08

6.1.5 Sightings

During the 14-day trial survey, a total of 215 sightings (205 on effort, 10 off effort) were made by the flying bridge observers (Table 8, Figure 28-Figure 33).

TABLE 8. Summary of marine mammal sightings made by flying bridge observers: number of pure and mixed schools, where mixed schools were included once for each species or species category in the mixed school; size is the mean school size across observers of their best estimates of total school size for pure schools, and in the case of mixed schools, their best estimates of subgroup size. Summaries are provided for closing and passing mode, as well as for both modes combined. The column “SWFSC Code” provides the NMFS species code (see Kinzey et al. 2000).

SWFSC Code	Species / species category	Pure		Mixed		Total	Survey mode
		Number	Size	Number	Size		
001	<i>Mesoplodon peruvianus</i>	3	3	0	--	3	combined
001		0	--	0	--	0	closing
001		3	3	0	--	3	passing
002	<i>Stenella attenuata</i> (offshore)	22	30	4	68	26	combined
002		7	13	3	42	10	closing
002		15	38	1	148	16	passing
006	<i>Stenella attenuata graffmani</i>	3	21	0	--	3	combined
006		2	26	0	--	2	closing
006		1	10	0	--	1	passing
010	<i>Stenella longirostris orientalis</i>	6	140	13	108	19	combined
010		4	181	6	102	10	closing
010		2	56	7	113	9	passing
015	<i>Steno bredanensis</i>	9	7	1	7	10	combined
015		2	10	1	7	3	closing
015		7	7	0	--	7	passing
017	<i>Delphinus delphis</i>	9	126	0	--	9	combined
017		3	128	0	--	3	closing
017		6	125	0	--	6	passing
018	<i>Tursiops truncatus</i>	8	8	2	3	10	combined
018		4	10	2	3	6	closing
018		4	7	0	--	4	passing
021	<i>Grampus griseus</i>	1	33	0	--	1	combined
021		0	--	0	--	0	closing
021		1	33	0	--	1	passing
032	<i>Feresa attenuata</i>	4	42	0	--	4	combined
032		2	33	0	--	2	closing
032		2	52	0	--	2	passing
033	<i>Pseudorca crassidens</i>	4	19	1	16	5	combined
033		3	13	1	16	4	closing
033		1	36	0	--	1	passing
037	<i>Orcinus orca</i>	1	16	0	--	1	combined
037		0	--	0	--	0	closing
037		1	16	0	--	1	passing
048	<i>Kogia sima</i>	3	1	0	--	3	combined
048		1	1	0	--	1	closing
048		2	1	0	--	2	passing
049	ziphiid whale	7	4	0	--	7	combined
049		3	4	0	--	3	closing
049		4	4	0	--	4	passing
051	<i>Mesoplodon sp.</i>	11	3	0	--	11	combined
051		4	4	0	--	4	closing

051		7	2	0	--	7	passing
061	<i>Ziphius cavirostris</i>	1	2	0	--	1	combined
061		0	--	0	--	0	closing
061		1	2	0	--	1	passing
070	<i>Balaenoptera sp.</i>	1	1	0	--	1	combined
070		1	1	0	--	1	closing
070		0	--	0	--	0	passing
075	<i>Balaenoptera musculus</i>	2	2	0	--	2	combined
075		0	--	0	--	0	closing
075		2	2	0	--	2	passing
076	<i>Megaptera novaeangliae</i>	9	1	0	--	9	combined
076		2	1	0	--	2	closing
076		7	2	0	--	7	passing
077	unid. dolphin	13	13	2	2	15	combined
077		4	12	2	2	6	closing
077		9	14	0	--	9	passing
078	unid. small whale	3	1	0	--	3	combined
078		2	1	0	--	2	closing
078		1	1	0	--	1	passing
079	unid. large whale	5	1	0	--	5	combined
079		2	2	0	--	2	closing
079		3	1	0	--	3	passing
090	<i>Stenella attenuata</i> (unid. subsp.)	26	26	6	30	32	combined
090		7	41	2	24	9	closing
090		19	21	4	33	23	passing
096	unid. cetacean	1	2	0	--	1	combined
096		1	2	0	--	1	closing
096		0	--	0	--	0	passing
099	<i>Balaenoptera borealis/edeni</i>	4	1	0	--	4	combined
099		2	1	0	--	2	closing
099		2	1	0	--	2	passing
177	unid. small delphinid	31	15	6	18	37	combined
177		9	22	4	6	13	closing
177		22	13	2	41	24	passing
277	unid. medium delphinid	11	8	0	--	11	combined
277		3	13	0	--	3	closing
277		8	6	0	--	8	passing

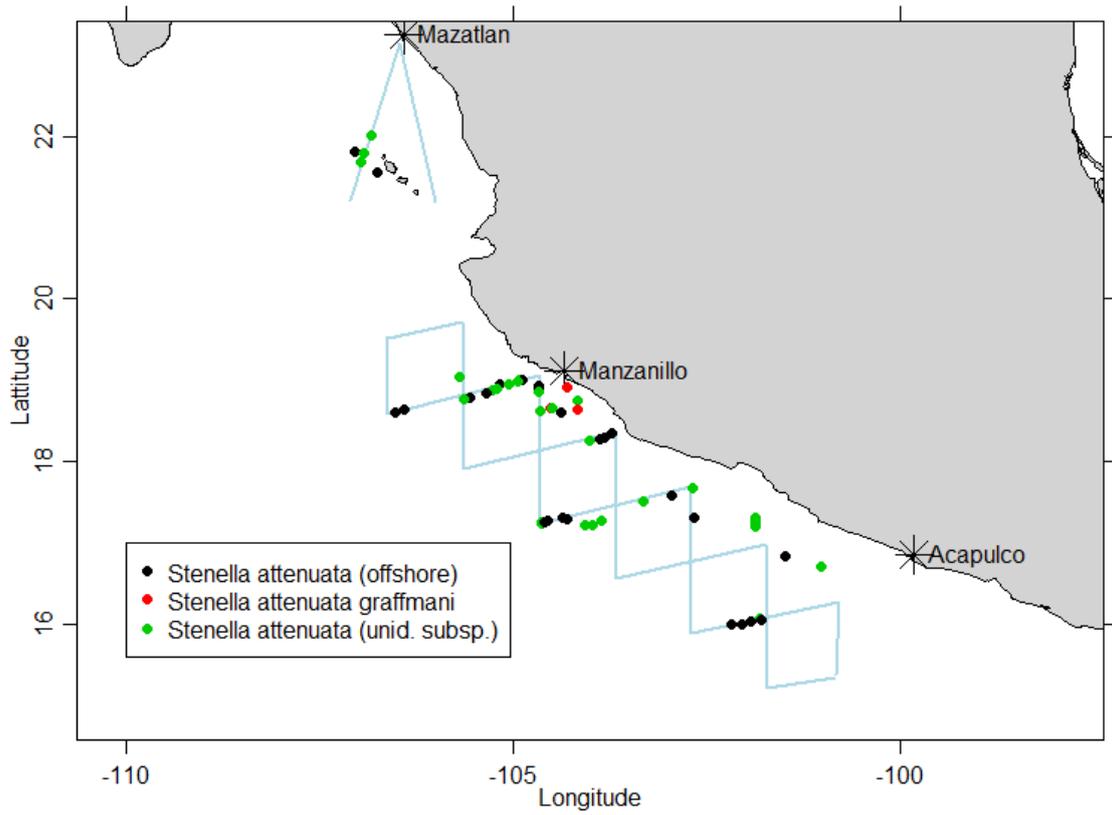


FIGURE 28. Spotted dolphin sighting locations.

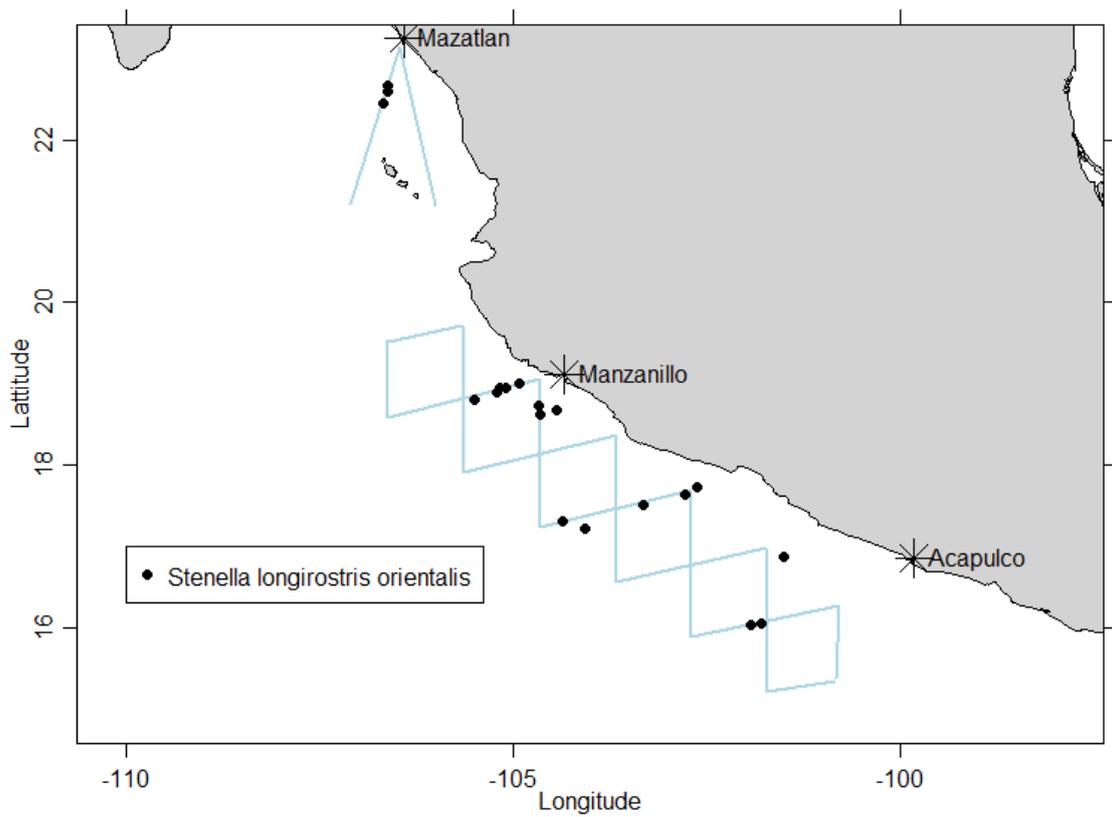


FIGURE 29. Eastern spinner dolphin sighting locations.

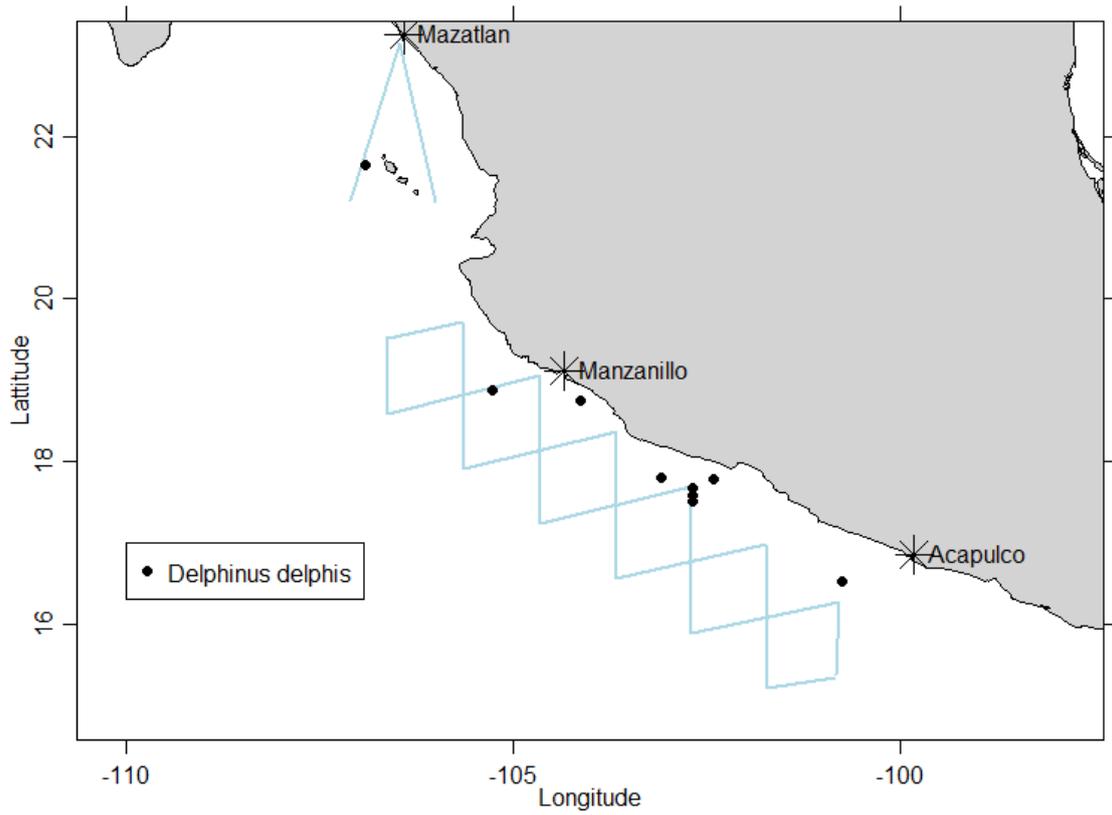


FIGURE 30. Short-beaked common dolphin sighting locations.

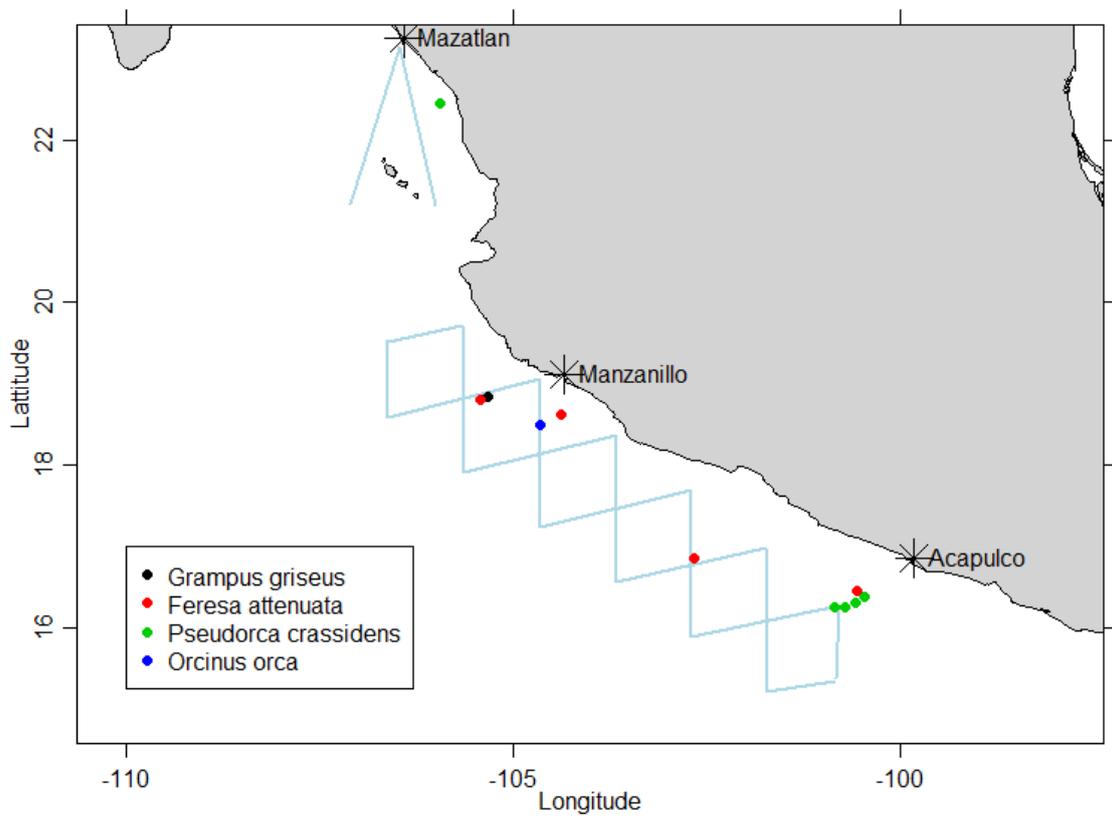


FIGURE 31. Sighting locations of species belonging to the blackfish family.

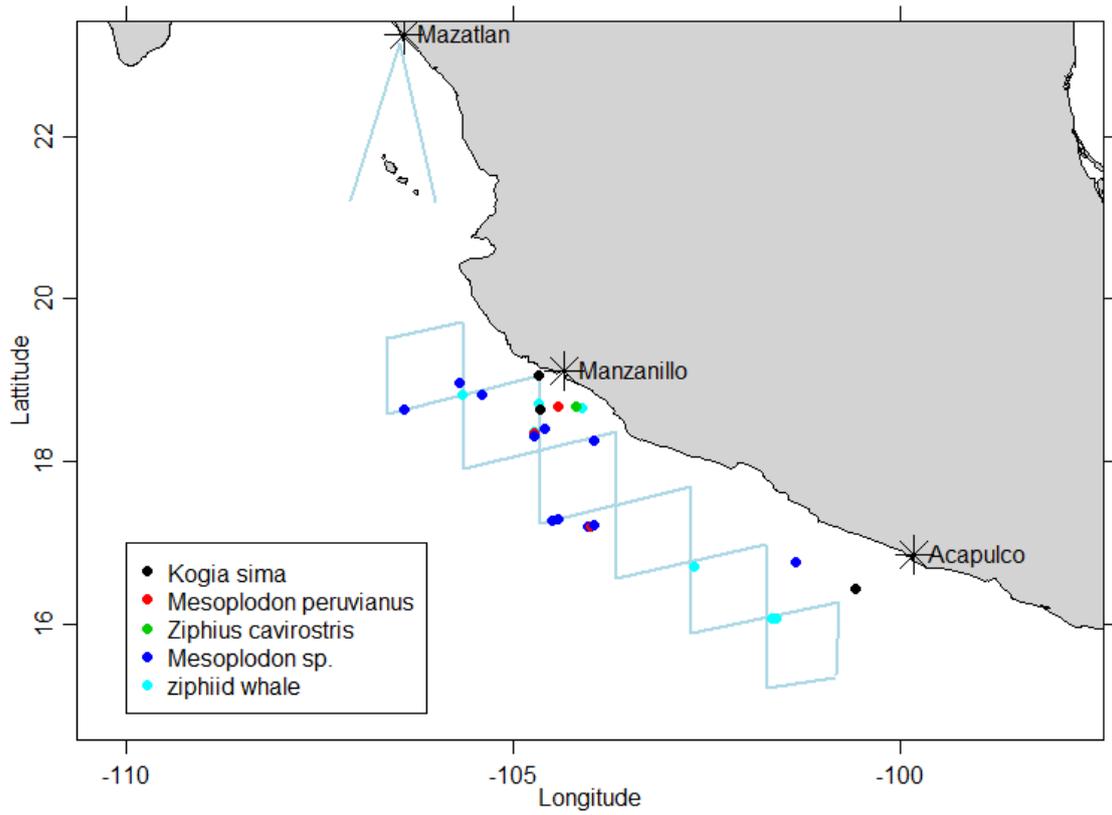


FIGURE 32. Dwarf sperm whale (*Kogia sima*) and beaked whale sighting location.

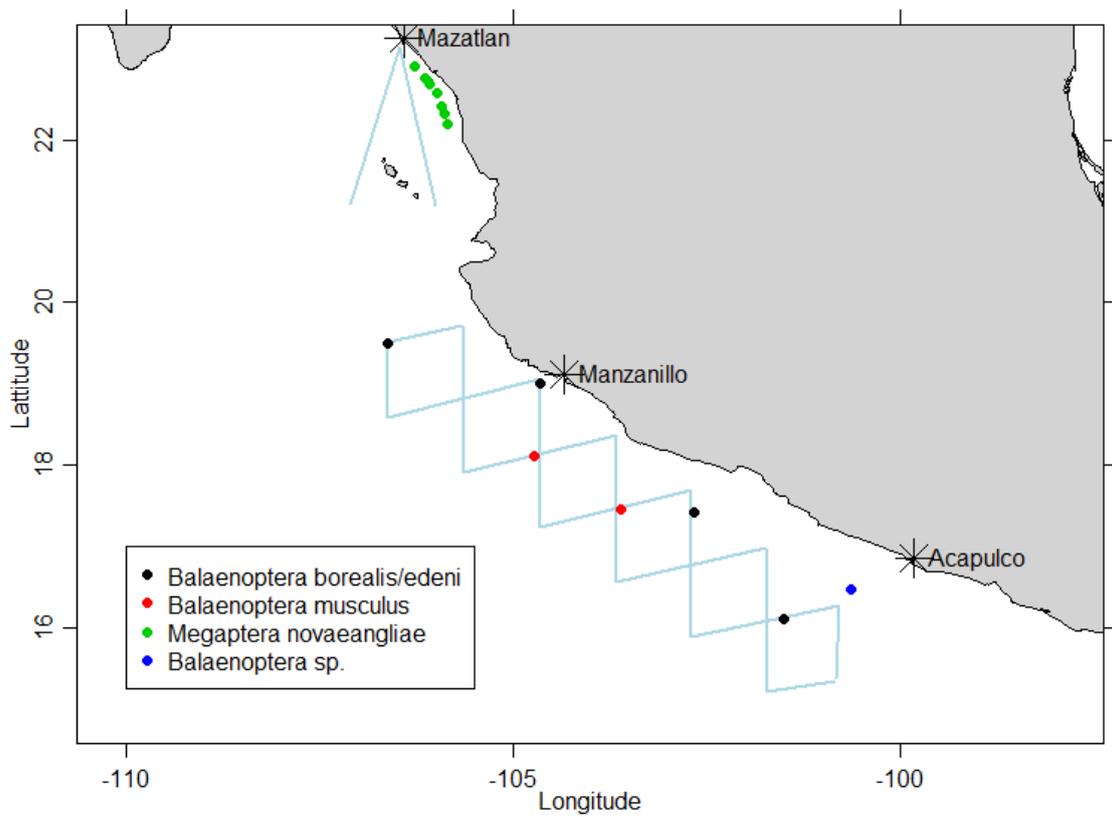


FIGURE 33. Baleen whale sighting locations.

Out of the 215 sightings, 198 were single species schools and 17 were mixed (Table 9). Nine of the mixed schools were composed of spotted and spinner dolphins, one was composed of spotted and rough-toothed dolphins. The latter, as well as the remaining seven mixed schools, contained clusters or individuals that were not seen well enough to identify to species and, hence, these sightings include a fraction of unidentified delphinids.

TABLE 9. Species composition of mixed species schools detected by the flying bridge observers. SWFSC codes refer to the following species: 002: *Stenella attenuata* (offshore); 010: *Stenella longirostris orientalis*; 015: *Steno bredanensis*; 018: *Tursiops truncatus*; 033: *Pseudorca crassidens*; 077: unidentified dolphin; 090: *Stenella attenuata* (unidentified subsp.); 177: unidentified small delphinid.

Species 1	Species 2	Species 3	Number of Sightings
002	010	--	4
010	090	--	5
010	177	--	4
015	090	177	1
018	077	--	1
018	177	--	1
033	077	--	1

6.1.6 Potential biases in closing versus passing mode

Schwarz *et al.* (2010) provided evidence that surveys conducted in closing and passing mode may have different biases with regards to school size estimates (negative bias for passing mode) and encounter rate estimates (negative bias for closing mode), as well as species identification (more unidentified schools during passing mode). For the 14-day trial survey, we found no evidence of negative bias in school size estimates made during passing mode (Table 8). However, the number of sightings per 1,000 km of effort was higher for passing mode (134.49) than for closing mode (97.86). There were more schools detected in passing mode for which at least a fraction of the individuals were unidentifiable to species or subspecies (Table 10).

TABLE 10. Number of sightings recorded as an unidentified single-species school, by effort mode (closing or passing mode). In the case of the closing mode sightings, only four of the SWFSC code 177 sightings were outside the turning range of 3 nm perpendicular distance from the trackline and thus not approached to obtain species identification; all others were closed on for species identification.

SWFSC Code	Taxonomic categories	Closing		Passing	
		Sightings	Sightings/1,000km	Sightings	Sightings/1,000km
077	<i>unid. dolphin</i>	4	5.22	9	9.31
078	<i>unid. small whale</i>	2	2.61	1	1.03
079	<i>unid. large whale</i>	2	2.61	3	3.10
090	<i>Stenella attenuata</i> (<i>unid. subsp.</i>)	7	9.13	19	19.66
096	<i>unid. cetacean</i>	1	1.30	0	0.00
177	<i>unid. small delphinid</i>	9	11.74	22	22.76
277	<i>unid. medium delphinid</i>	3	3.91	8	8.28

6.1.7 Comparison of detection functions with previous surveys

Estimates of $f(0)$ for the trial survey were smaller for eastern spinner dolphins compared to spotted dolphins; however, 95% confidence intervals overlapped (Table 11). Accordingly, the shape of the estimated detection function, was generally wider for eastern spinner dolphins compared to spotted dolphins (Figure 34)

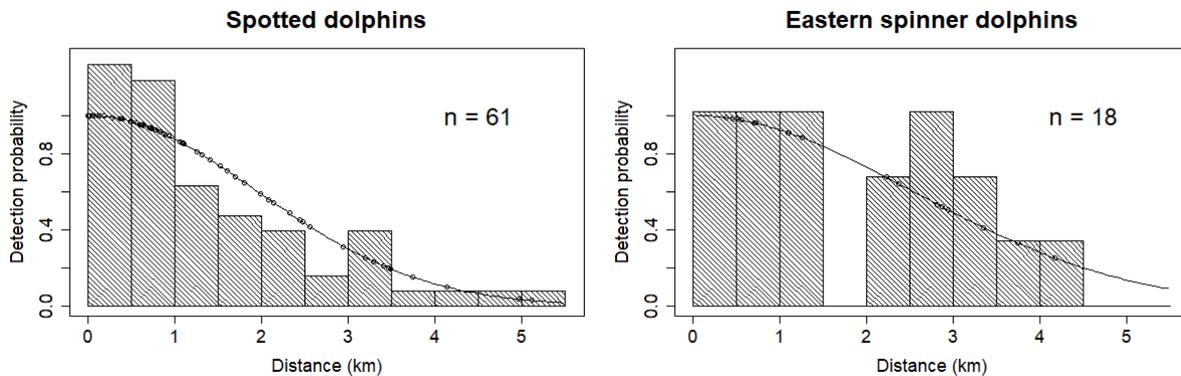


FIGURE 34. Histogram of perpendicular distance and half normal detection function with species (spotted versus spinner) as a covariate shown for spotted and spinner dolphins where n is the sample size and dots represent the estimated detection probabilities for the observed distances.

TABLE 11. Estimated $f(0)$ for spotted and spinner dolphins from the MCDS model with species as a covariate fitted to trial survey data shown with coefficient of variation (CV), lower (LCI) and upper (UCI) bounds of 95% confidence intervals and sample size (n).

Species	$f(0)$	CV	LCI	UCI	n
Spotted	0.42	9.66	0.34	0.50	61
Spinner	0.33	16.64	0.23	0.46	18

Estimated detection probabilities within the 5.5km strip half-width were similar between the trial survey and previous surveys 1997-2000 (Gerrodette and Forcada 2005) (Figure 35). Confidence intervals for estimates from the trial survey overlapped with confidence intervals from previous surveys in all but three years for spotted dolphins and in all but one year for spinner dolphins.

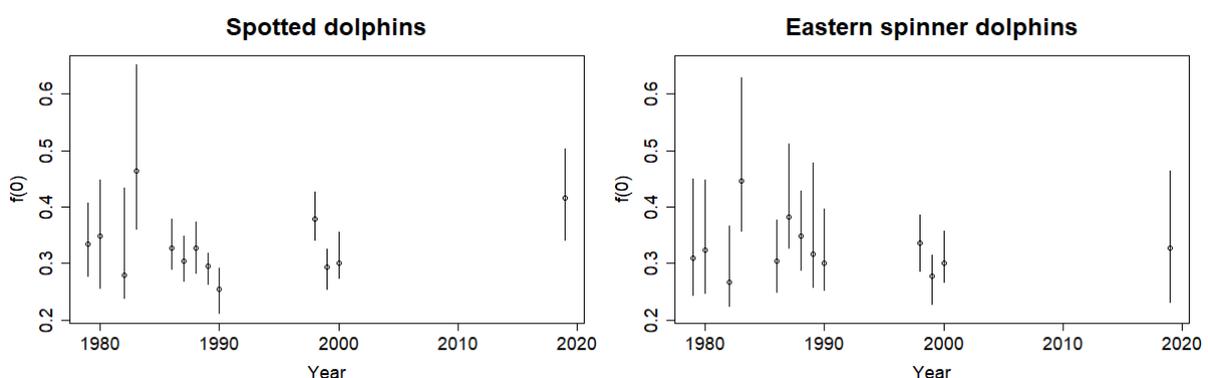


FIGURE 35. Estimates of $f(0)$ and 95% confidence intervals, for spotted and spinner dolphins. Estimates from 1979-2000 were sourced from Gerrodette and Forcada (2005), 2019 estimate from model fitted to trial survey data (Table 11).

6.2 Research vessel as a suitable platform to conduct drone operations

6.2.1 Drone platform: launches and landings in relation to Beaufort sea state

During the 14-day trial survey, we conducted 94 drone flights. Launches and landings occurred within Beaufort sea states ranging from 1 to 5 (Table 12). The number of landings was less than the number of launches due to flight 9 on 21 November, which ended in a drone crash after 22min of flight time due to a loss of satellite coverage. The drone platform was suitable for launching and landing the Seahawk drone. Launches and landings were performed without major incidents, although one minor incident occurred when a 2nd pilot, who was practicing landing the drone, landed the Seahawk with one rail of the landing gear on a safety rope, causing the drone to tilt over to one side and with the blades hitting the drone platform. The resulting damage to the drone was fixed by replacing the blade(s). Most launches and landings were done while the ship was cruising at survey speed of 10 knots. While it was at the discretion of the drone pilots to ask for a change in course and speed of the ship to ensure safe operations, only during a few occasions during higher Beaufort sea states did they ask for a reduction in speed.

TABLE 12. Number of drone launches and landings by Beaufort sea state.

Beaufort	Launches	Landings
0	0	0
1	9	6
2	33	33
3	23	24
4	26	25
5	3	5
6	0	0
Total	94	93

6.2.2 Drone team and equipment

The drone team members were good to work with and worked hard towards the goals of the project with the equipment provided. In particular, 1st pilot, LCY (see Table 3 for reference), stood out in terms of his drone handling skills, launching and landing the drone with ease while the ship was underway (see Section 6.2.1). TW was a good team leader who was proactive in trying to find solutions to problems. The drone team was able to turn around a drone between consecutive flights within about 5min. However, with only one 1st pilot on board, it proved impossible to fly two drones simultaneously, which was contrary to what was specified in the drone protocol (Appendix 1). As the mean flight time of the drone was under 1hr (Table 13), this meant that the goal of flying drones during all daylight hours or during closing mode effort was not accomplished. While the drone team safely conducted up to an impressive 13 flights per day (Table 14), this was not sufficient to cover the entire hours of operation on the flying bridge.

TABLE 13. Summary of total duration (hh:mm:ss) of flights by flight mode.

Mode	Min	1. Quartile	Median	Mean	3. Quartile	Max
Zigzag	00:08:35	00:41:10	00:49:43	00:45:41	00:55:19	01:04:31
Calibration	00:13:27	00:31:28	00:33:43	00:36:09	00:44:07	00:54:00
Mixed	00:46:44	00:49:37	00:52:06	00:51:17	00:53:47	00:54:14
Combined	00:08:35	00:36:57	00:48:25	00:44:24	00:54:29	01:04:31

TABLE 14. Comparison of daily hours of operation (hh:mm:ss) between the flying bridge and the drones.

Date	Flying bridge	Number of flights	Total flight duration
17	04:48:40	3	01:26:09
18	12:00:46	13	08:15:20
19	10:55:11	10	08:54:03
20	10:53:59	10	07:11:10
21	10:59:44	9	06:45:25
22	11:05:11	0	00:00:00
23	11:04:43	5	02:38:22
24	09:03:37	1	00:46:10
25	10:40:47	8	05:49:56
26	10:51:01	6	05:29:59
27	11:04:18	7	05:43:14
28	11:09:02	11	07:51:28
29	10:15:16	6	04:52:42
30	07:13:47	5	03:51:06

Collection of high-quality drone imagery proved problematic because of the equipment provided to the project. Only one of the two drones in operation had a camera that was capable of recording the video footage on board the drone. This was the drone that was lost into the sea on 21 November. Also, recording on board the drone had to be manually operated by one of the drone team members and restarted at regular intervals of about 3min. When this was forgotten, no recording on board the drone took place. This resulted in only very few short video clips actually being recorded on board the drone before the drone crash and none after. As a result, no on board recorded video footage capturing cetaceans exists for any of the calibration flights and only during one zigzag flight during which a drone detection was made was footage recorded on board capturing a single cetacean swimming underwater (drone detection 4, see Section 6.3.7). Hence, what was originally thought of as the backup mechanism for recording video, the screen capture of the transmitted video, became the main source of the video available for post-survey image analyses for the entire duration of the trial survey. Naturally, this video suffered from transmission loss and compression issues which caused frequent pixilation and complete loss of the video (see Section 5.3.4.1). The antennas that the drone team installed on the ship were also not sufficient to receive transmissions from great distances. Whenever zigzag flights at 5 nm ahead of the ship were attempted, transmission loss increased substantially.

6.2.3 Drone operations: effort conducted during zigzag and calibration flights

The total flight time of the 94 flights combined was 69 hr 26 min and 32 sec (Table 15). Out of the 94 flights, 74 were pure zigzag flights, 15 flights were pure calibration flights, and four were initiated as zigzag and then switched to calibration mode during which it was at least attempted to find the school using the drone. It is worth noting that during two of the flights in the latter category, the potential calibration school was found with the drone; however, the effort did not result in a valid calibration school as the entire school could not be captured by video. During 69 pure zigzag flights at least one zigzag leg and up to nine zigzag legs were flown. During three occasions, the zigzag effort on a given leg was interrupted and the drone stopped and redirected to investigate a potential sighting. Out of all 19 flights that were at least in part dedicated to calibration, the potential calibration school was found in 15 occasions (see Section 6.4). We further divided the flights into different effort modes, e.g. whether the drone was in outbound transit to the first waypoint during a zigzag flight, on effort on a zigzag leg or on inbound transit back to the ship. During any given flight, more than one effort mode

was usually conducted. A typical calibration flight would consist of ‘outbound transit’, ‘with school’ and ‘inbound transit’. In Table 15 we summarise the different effort modes for the drone flights, along with their average times and Beaufort sea states encountered. These are explained in more detail in the following sections on zigzag and calibration flights (Sections 6.3 and 6.4). The effort mode ‘other’ was logged during three flights (Table 15) and represents: 1. the random flight path that the drone took on 21 November before it was lost into the sea; 2. the test flight that followed the crash on 23 November; and, 3. the one occasion during which the drone took public relations footage of the ship.

TABLE 15. Effort modes, total number of flights containing each mode of effort, mean and total duration of each effort mode across flights (hh:mm:ss), as well as mean, minimum and maximum Beaufort (Bf) sea states encountered. *: The calculation of the mean only applies to the flights during which the respective effort mode occurred.

Effort mode	Total	Mean time per flight*	Total time	Mean Bf	Min Bf	Max Bf
Zigzag						
Transit outbound	78	00:13:50	17:59:50	2.96	1	5
On effort	69	00:24:02	27:38:19	2.98	1	5
Transit inbound	75	00:09:09	11:27:02	2.94	1	5
Checking	3	00:09:03	00:27:10	2.52	2	3
Total time			57:32:21			
Calibration						
Transit outbound	19	00:15:57	05:03:04	1.90	1	3
With school	15	00:18:12	04:33:09	2.01	1	3
Transit inbound	19	00:04:15	01:21:01	2.17	1	4
Total time			10:57:14			
Other						
Other	3	00:18:59	00:56:57	2.56	2	4
Total time			00:56:57			
Total time			69:26:32			

6.3 Test drone for collecting trackline detection probability data

6.3.1 Zigzag flight effort modes and coverage

The three main effort modes (transit outbound, on effort and transit inbound; Table 15) refer to the transit out to the first zigzag leg, the time on effort in search mode while on the zigzag legs (e.g. Figure 36) and the inbound transit back to the ship, respectively. During either transit, the camera was generally pointing forward at an angle of approximately 50°, where 0° would indicate parallel to the ocean surface and 90° would indicate pointing directly down. While on effort, the camera was generally pointing directly down (but see below in this section) and the drone was moving along the zigzag leg at a steady pace. The effort mode checking refers to time spent by the drone investigating an object of interest, i.e. a potential sighting that was spotted by the drone observers during the real-time observation; it represents “off effort” time for the drone. Transit outbound and inbound durations were, on average, 14 min and 9 min, respectively, while the average on effort time per flight was 24min (5.5min per leg, 300 legs total, Table 15); i.e., only 1 min longer than the average combined transit time. Checking on a potential object of interest occurred only three times, with average duration of 9 min.

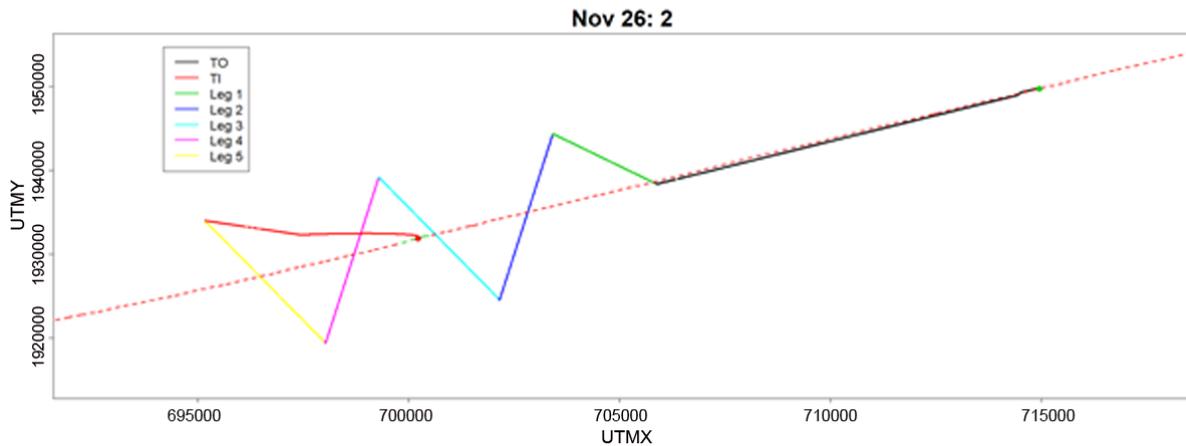


FIGURE 36. Example zigzag flight showing the different effort modes of the drone. TO: outbound transit, TI: inbound transit and on effort in search mode during 5 zigzag legs. Red dashed line indicates the ship track while flying bridge was on effort in passing mode. Green and red dots indicate the location of the drone during take-off and landing, respectively. Date and flight number given in the title. Units for UTM coordinates are meters.

As noted above, it was not possible to test the operation of two drones simultaneously during the trial survey. However, flight statistics from the trial survey make it possible to put forward theoretical considerations about requirements for continuous drone zigzag effort by a Seahawk drone during all daylight hours while the flying bridge observers were on effort, which was a requirement specified in the drone protocol (Appendix 1). Consider a survey day with 12-hours of daylight during which the sunrise and sunset are, say, at 06:00 and 18:00, respectively. The first drone, drone 1, would need to be launched at 05:46 in order to reach the first waypoint of leg 1 at sunrise after 14 min of outbound transit (Table 15), which, in this example, is ~ 3.52 nm ahead of the ship, as that was the mean distance ahead of the ship that was accomplished within the mean transit times (see below in Section 6.3.3. **Target 1**). Drone 1 would then spend 26 min surveying the zigzag legs before it had to start its return journey of 9 min (Table 15) inbound transit at 06:26 to be safely back on board after 49 min of flight time (Table 13) at 06:35. In order to have uninterrupted drone coverage ahead of the ship, the second drone, drone 2, would have to be launched at 06:12 in order to be 3.52 nm ahead of the ship when drone 1 had to return at 06:26. Drone 2 would then survey zigzag legs for 26 min until 06:52 before it had to return to the ship and arrive there at 07:01. Drone 1 would have to be launched again at 06:38 in order to reach 3.52 nm ahead of the ship at 06:52 to relieve drone 2, and so on. This means that each drone would have to be launched again 3 min after it had returned from its last flight. It also means that during all daylight hours, a drone would have to be launched and landed every 26 min. With an average of 12 hr of daylight during the main survey, this would result in approximately 28 launches and landings per day, and for a 120 sea-day survey (Oedekoven *et al.* 2018), $120 \times 28 = 3,360$ launches and landings would be required. Such an undertaking would require sufficient personnel to operate the drones safely, and raises major concerns about safety while being far offshore. Furthermore, while the above example may in theory be feasible (and in theory only), it is noted that with a target distance of 5 nm ahead of the ship (instead of the 3.52 nm used in the example), even two drones would not be sufficient as the outbound and inbound transit times would increase by over 40% each, making the total transit times per flight longer than the time the drone could spend on the zigzag legs.

The GPS locations and altitude of the drone were recorded by the drone team on a system on board the ship in high resolution both in time and space, with about four records per second (altitude in mm and latitude and longitude to the 7th decimal degree). Zigzag legs were flown at a height of 133.82 m

on average, ranging between 66.70 m and 224.18 m (Table 16). With a camera aperture angle of 60°, and assuming the camera was pointing straight down, this resulted in an estimated swath width of 154.62 m on average (range: 77.02 – 258.86 m) captured by the drone video footage. The lengths of the individual zigzag legs were 3.38 km on average (range: 0.08 – 4.31 km), with a total length of 1,013.37 km of zigzag legs surveyed during the two weeks. The total area covered was 157 km² with an average of 0.53 km² per leg (range: 0.01-1.05 km²). It is noted that these calculations were based on the assumption that while on effort, the camera was generally pointing directly down. However, camera angles were not recorded automatically and sometimes this angle was adjusted to improve viewing conditions (e.g., to reduce glare). The drone observers logged the camera angles when these were reported by the drone pilots. As there was no strict protocol in place during which the drone pilots reported *all* angle adjustments immediately (which would for both drone pilots and observers be extremely impractical and subject to a high error rate), exact angles for the entire duration of each flight were not available. In addition, as the absolute viewing angle of the camera in relation to the ocean surface also depends on the pitch and roll angle of the drone, it is not sufficient to record a single camera angle but the orientation of the camera in three dimensions would be required, in addition to the pitch and roll of the drone. Hence, the estimates of swath width and area covered by the drone shown in Table 16 represent approximations.

TABLE 16. Summary of the altitude, swath width and length, and area covered, for the 300 zigzag legs. For altitude and swath width, the summaries presented in the table are based on the average of those quantities for each leg, and it was assumed that the camera was pointing directly down (see text for details).

Flight parameter	Mean	Min	Max	Total
Altitude (m)	133.82	66.70	224.18	NA
Swath width (m)	154.62	77.02	258.86	NA
Length (km)	3.38	0.08	4.31	1,013.37
Area covered (km ²)	0.53	0.01	1.05	157.50
Time (hh:mm:ss)	00:05:31	00:00:23	00:12:46	27:38:19

It is estimated that the area searched by the drone during zigzag flights represents only 0.83% of the area searched by the flying bridge observers. The length of all transect lines covered by the flying bridge was 1,733.06 km (Table 6). Observers were able to make detections out to the horizon at 8.3nm (see Section 6.1.1). However, data are usually truncated at 5.5 km for analysis, which corresponds roughly to the turning radius specified in the NMFS protocol (Section 5.1.2). Using this truncation distance, the area covered by the flying bridge was 19,063.66 km². By comparison, the drone covered an area of 157.50 km² or 0.83% of the area covered by the flying bridge.

6.3.2 Zigzag flights: altitude vs speed

Several flight altitudes and speeds were tested, but only at lower altitudes and slower speeds was it possible for drone observers to review the video in real-time. On the first day of the trial survey test flights were flown with zigzag legs at relatively high speed and low altitude (Figure 37). As this combination resulted in the video rushing by too quickly for real-time reviewing, a wider range of both altitudes and speeds was tested on the second and third days (18 and 19 November). It was concluded that due to the poor quality of the transmitted video, the altitude should be kept at about 100 m and the drone speed at about 30 km/hr. Only at this combination was it feasible for the drone observers to conduct real-time monitoring. At higher speeds the objects of interest were rushing by too quickly, which, in theory, could be alleviated by flying at higher altitudes; however, at higher altitudes the dolphins were too small to detect with any confidence.

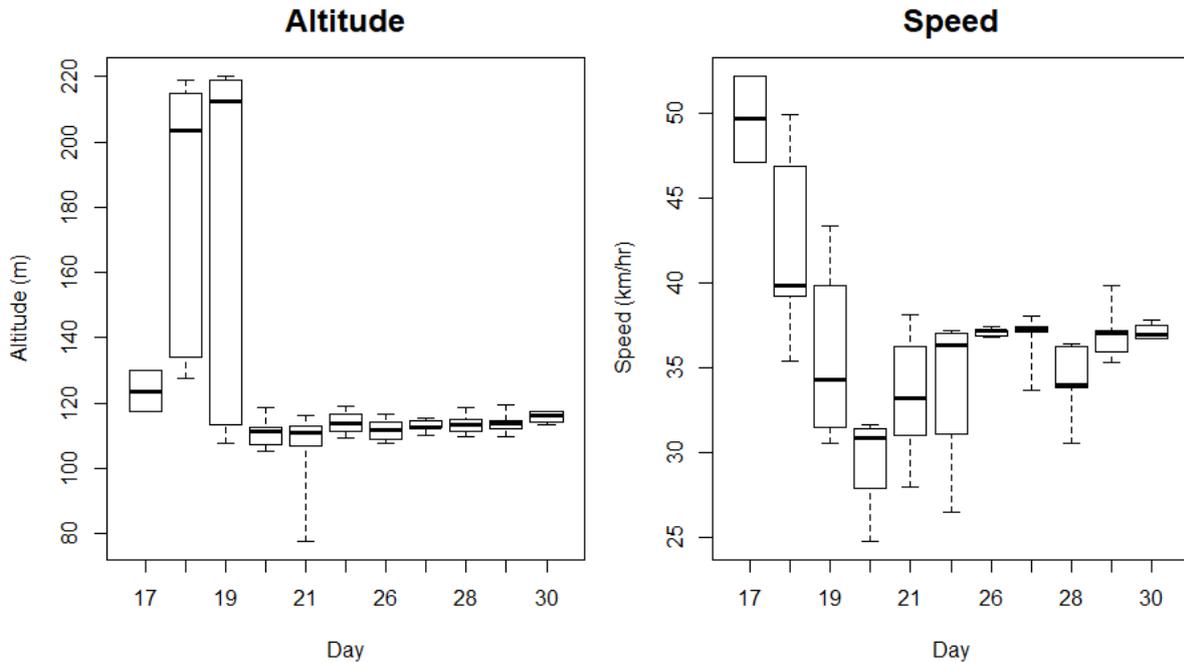


FIGURE 37. Drone altitude and speed distributions, by day, for the 300 zigzag legs. Horizontal bars represent medians, boxes the interquartile ranges and whiskers reach to the extreme values.

6.3.3 Zigzag flights: flight path assessment

The success of the drone at achieving the pre-determined flight path was evaluated using four metrics. During the zigzag flights for the main survey, the drone is supposed to maintain station ahead of the ship at a distance, D_{ship} , of 5nm or more (see Section 5.3.2). The target for these flights is to cover the trackline surveyed by the ship out to a distance $w = 3$ nm on either side of the trackline (Figure 38), i.e. matching the maximum perpendicular distance at which cetaceans detected by the flying bridge observers are within the turning radius of the ship for approaching the animals (see Section 5.1.2). Hence, the targeted width of the corridor covered by the drone w_{cor} should be $w_{cor} = 2w = 6$ nm. For a perfectly symmetrical flight path, this would further entail that the mid-points of the zigzag legs, except of the first and last legs per flight, would be on the ship's trackline and that the perpendicular distance of the drone to the trackline, D_{mid} , should be zero (as shown in Figure 38) where the **Xs** fall directly on the transect line). In order to conduct a zigzag flight according to a given combination of w and D_{ship} , the drone team estimated the locations of the ship along a projected trackline based on the course and speed of the ship at the time of launch and the time until the drone would reach D_{ship} . Using this projected trackline, they determined the waypoints of the zigzag flights and uploaded these onto the drone before launch. How successfully the targets were met can be measured in the following ways, considering both the projected trackline and the trackline that was actually completed by the ship. It is noted that these tracklines (including the timely progression along the respective trackline) often varied due to slight variations in course and speed of the ship. They varied substantially during closing mode when the ship turned on sightings.

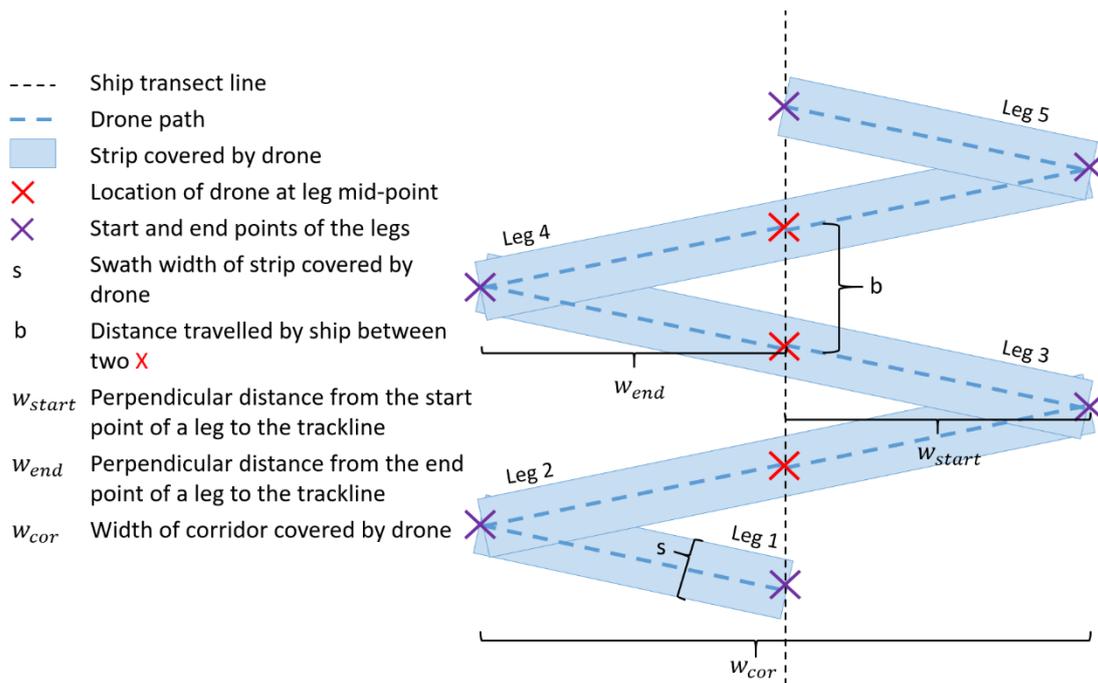


FIGURE 38. Diagram of theoretical drone flight path consisting of five zigzag legs in relation to the ship transect line. Not shown are the outbound and inbound transits from and to the ship.

1. Target: at the mid-points of the zigzag legs the drone should be 5nm ahead of the ship ($D_{ship} = 5\text{nm}$)

Two problems were encountered when attempting to achieve a D_{ship} of 5 nm. To estimate the actual values of D_{ship} during each flights, we measured the distances between the drone's position at its leg mid-points and the concurrent positions of the ship (Figure 39). We excluded the first and last legs of each flight from these calculations as these were not expected to cross the ship's trackline. The first problem was that the collection of video imagery at a D_{ship} of 5nm for was not possible. Video transmission errors increased substantially at values of D_{ship} larger than about 4 nm, and hence, for the majority of flights, shorter distances, e.g. 3 nm, 3.5 nm or 4 nm, were actually targeted resulting in a mean D_{ship} across all flights of 3.52nm (SD=0.90). The second problem was that these shorter values for D_{ship} generally could not be met with precision. If these were met with precision, the distribution of D_{ship} would show distinct peaks at the targeted distances of e.g. 3 nm, 3.5 nm or 4 nm in Figure 40. For example, the mid-points of legs 2-5 in the left example of Figure 39 were at $D_{ship} = 2.50, 2.66, 2.84, 3.02$ nm, i.e. increasing during the flight instead of being constant at the targeted value for that flight. Similarly, the mid-points of legs 2-4 in the example on the right were also increasing with $D_{ship} = 3.92, 4.17, 4.40$ nm. The more striking feature of this example, however, was that the drone did not cover the area of the ship's proposed trackline during this flight, which we address below. This was likely due to a combination of factors, foremost the rigidity of using uploaded, and hence fixed zigzag waypoints, which could not be adjusted during the flight. Slight variations in the ship's course and speed may have also contributed to this (compare, e.g. the projected with completed trackline in Figure 39).

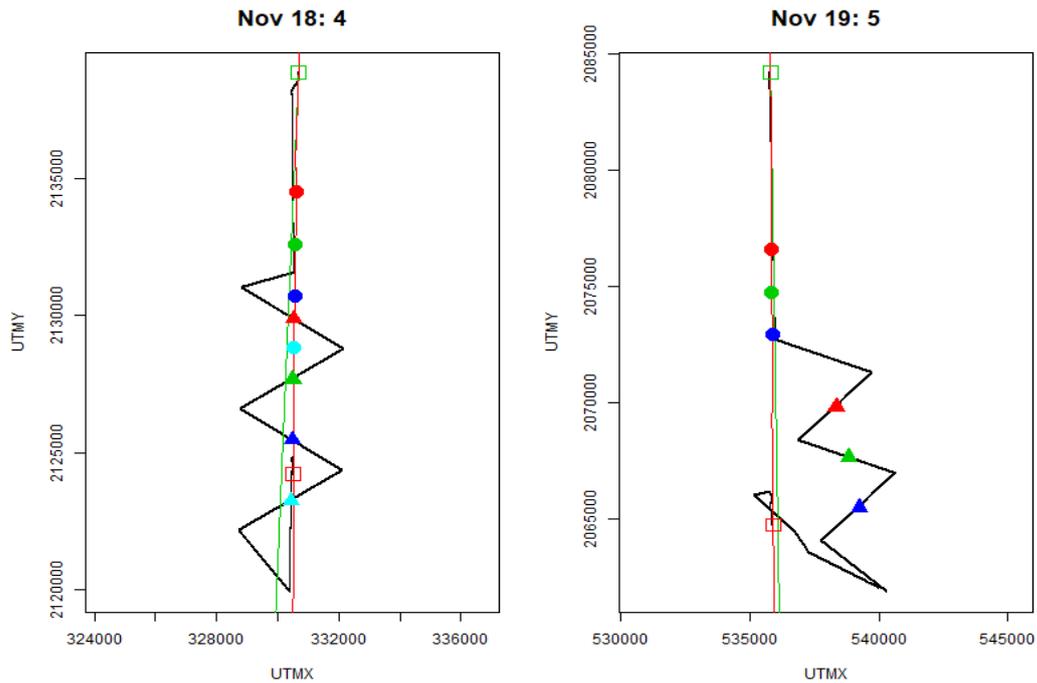


FIGURE 39. Two example flight paths of the drone (black lines) in relation to the completed ship track surveyed by the flying bridge observers (red line) and projected trackline of the ship given the ship's course just prior to launching the drone (green line). Open green and red squares indicate the launch and land locations of the drone, respectively. Coloured triangles indicate the mid-leg locations of the drone for each leg except the first and last of the flight; coloured dots indicate the location of the ship when the drone was at the respective mid-point. Date and flight number are given in the title. Units for UTM coordinates are meters.

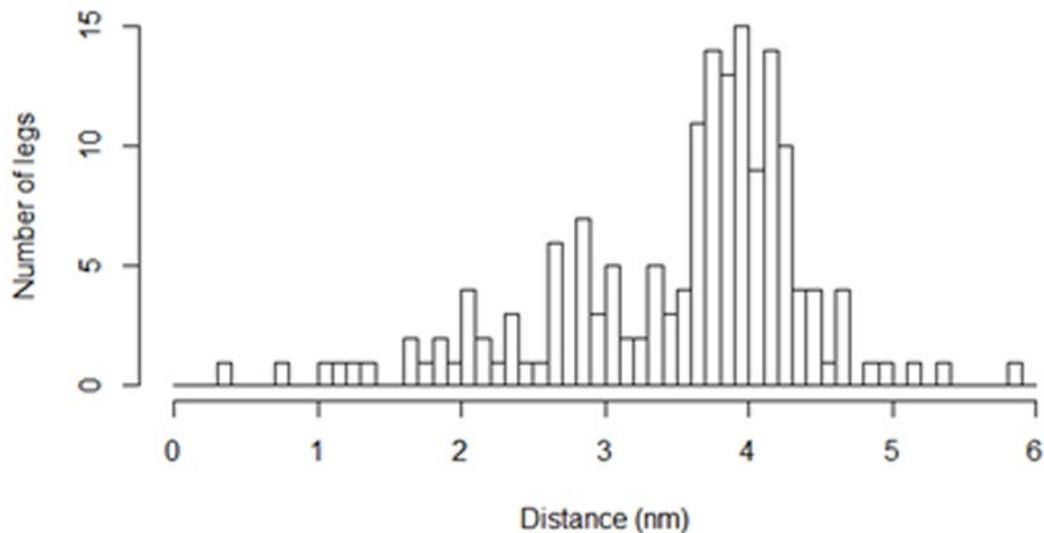


FIGURE 40. Frequency distribution of estimates of D_{ship} .

2. Target: the width of the corridor surveyed by the drone ahead of the ship is 6nm

The achieved width of the corridor surveyed by the drone in front of the ship, w_{cor} , was much less than 6 nm, with a mean of 1.93 nm when measured with respect to the completed trackline, and a

mean of 1.71 nm when measured with respect to the projected trackline (Table 17). w_{cor} was calculated as follows:

$$w_{cor} = |m \times w_{start} - m \times w_{end}| \quad (1),$$

where w_{start} and w_{end} are the perpendicular distances from the start and end points of a given leg, respectively; $m = -1$ if the location of the start/end point was to the left of the trackline and $m = 1$ if it was to the right.

The reduction in w_{cor} was mostly due to a decision made at the start of the trial survey: a reduced target for w_{cor} of 3 nm was necessary to facilitate the initial testing of the drones, with a potential to widen it if tests were successful. As even this reduced target distance was never achieved, w_{cor} could never be increased to 6nm during the trial survey. Hence, for all legs, except the first and last of each flight, w_{cor} was targeted to be 3 nm, i.e. 1.5 nm to either side of the trackline (see Figure 38). Despite the reduced target for w_{cor} , the achieved width of the corridor was much less with a mean of 1.93 nm when measured with respect to the completed trackline and a mean of 1.71 when measured with respect to the projected trackline (Table 17). The most likely reason that w_{cor} was less than 3 nm was that a reduction in w_{cor} was required due to the poor video quality. To try to improve the video quality, the drone had to fly at low altitudes and, therefore, at low speed (see Section 6.3.2). As a result of this, the width of the corridor had to be reduced in order for the drone to maintain position ahead of the ship.

TABLE 17. Summary statistics for w_{cor} (equation 1).

w_{start} and w_{end} measured to:	w_{cor} (nm)			
	1st Quartile	Median	Mean	3rd Quartile
Completed trackline	1.72	1.88	1.93	2.22
Projected Trackline	1.70	1.81	1.71	1.88

3. Target: the drone should be on the trackline at the leg mid-points ($D_{mid} = 0$)

Overall, the estimates of D_{mid} varied considerably within and among flights (Table 18), ranging from 10 – 6,018 m and 2 – 6,002 m for the completed and projected trackline, respectively. D_{mid} was estimated by measuring the shortest distances of the leg mid-points to the completed and projected tracklines, excluding the first and last leg of each flight. The projected trackline is the course of the ship at the time the drone was launched. The two examples shown in Figure 39 are on opposite ends of the observed spectrum of meeting this target. In the example on the left, the mid-points of legs 2-5 were at $D_{mid} = 18, 28, 46$ and 74 m from the completed trackline and at $D_{mid} = 187, 240, 304$ and 358 m from the projected trackline. In the example on the right, the waypoints seem to have been set up based on the wrong course of the ship. Here, the mid-points of legs 2-4 were at $D_{mid} = 2,470, 3,088$ and $3,336$ m from the completed trackline and at $D_{mid} = 2,331, 2,781$ and $3,159$ m from the projected trackline.

TABLE 18. Distances D_{mid} (m) from the drone at leg mid-points to the completed and projected tracklines of the ship.

Distance to:	Min	1. Quartile	Median	Mean	3. Quartile	Max
Completed trackline	10	208	534	952	1,220	6,018
Projected Trackline	2	187	535	817	1,036	6,002

4. Target: the start and end points of the zigzag legs (excluding the first and last of a flight) should be on opposite sides of the trackline

Out of all 61 flights with more than two zigzag legs, the target was achieved for 45 flights with regards to the completed trackline and for 46 flights with regards to the projected trackline (Table 19). However, for 11 flights, this target was missed completely. In the example on the left in Figure 39 this target was achieved, when the start and end points of legs 2-5 were on opposite sides of the trackline (both completed and projected). By contrast, for the example on the right, the start and end points of legs 2-4 were on the same side of the trackline (both completed and projected).

TABLE 19. Number of flights by the percentage of legs per flight (excluding first and last legs) that met the target, relative to completed and projected tracklines.

Percentage of legs covered	0	1-19	20-39	40-59	60-79	80-99	100
	Number of flights						
Completed trackline	11	0	1	0	3	1	45
Projected trackline	11	0	1	0	2	1	46

6.3.4 Real-time transmission of video back to the ship and monitoring by drone observers for cetacean sightings

As mentioned in Section 5.3.4.1, the imagery obtained from the real-time transmission of the video was often of poor quality, with frequent pixilation (Figure 41) and video freezes or complete transmission loss. These issues were likely, at least in part, a result of the antenna that the drone team mounted on the ship for receiving information from the drone. The drone team tried to improve transmission by mounting the antenna in different locations on the ship during the survey, without much improvement. The quality of the transmission was a function of how far the drone was away from the ship, where 5 nm was deemed to be too far, with too many issues to make real-time monitoring viable. Hence, for the main survey, where the drone will be required to operate at least 5 nm ahead of the ship, a better antenna system will be necessary.

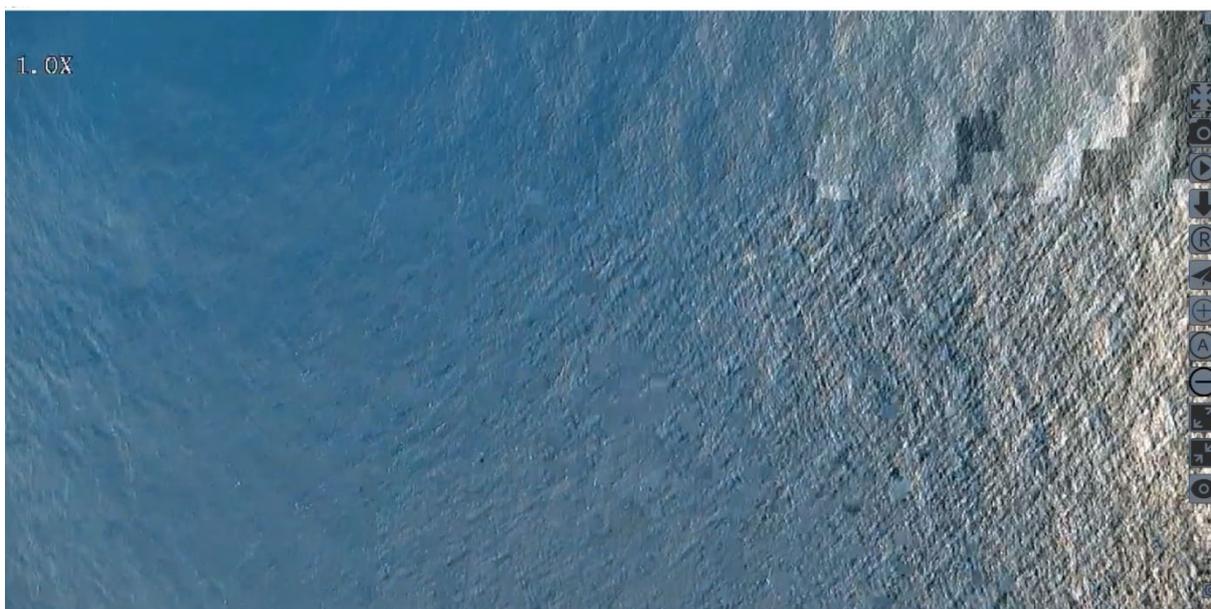


FIGURE 41. Screenshot of video footage with pixilation and compression issues throughout most of the frame.

6.3.5 Video analysis

6.3.5.1 Manual classification

Most of the 20,195 objects of potential interest that underwent manual classification (Section 5.3.4.2.1) were assigned the class *water*, with *dolphin* the second most common class (Table 20); the dataset in general was very inhomogeneous. The majority of classes had a relatively large standard deviation in all measurements, indicating that the size of the labelled objects had a large variability. The dolphin class objects were relatively square which was unexpected, and may indicate that either dolphins were predominantly picked up swimming diagonally relative to the camera or that the computer vision algorithm padded the objects size.

During all effort modes of the zigzag flights combined, drone observers logged objects that could be of at least potential interest during 92 occasions. During post-hoc review, six of these were confirmed as cetacean detections.

TABLE 20. Summary of the dataset classified by the human observer: frequency, mean area, width and height, along with the standard deviation (SD) of each (except for the count). The mean area refers to the mean area of the boxes that fully enclosed each object within each class. Similarly, the mean width and mean height are the mean of the widths and heights of the box that encloses the objects in each class. Each measurement (except for the count) is also presented with its standard deviation.

Label	Count	Mean area	Area SD	Mean width	Width SD	Mean height	Height SD
Dolphin	2,704	350.8	1,243.7	14.2	14.5	14.9	15.0
Bird	26	99.7	50.3	9.1	2.7	10.7	3.1
Multiple Dolphins	151	873.2	2,193.1	23.8	22.8	22.9	20.3
Whale	0	0	0	0	0	0	0
Turtle	0	0	0	0	0	0	0
Unknown	913	215.3	963.5	13.6	7.7	11.6	7.2
Unknown not cetacean	554	183.3	281.4	13.3	7.7	11.6	7.2
Boat	5	388.0	426.1	11.8	7.9	25.0	14.0
Fish	0	0	0	0	0	0	0
Trash	1	3795	0.0	69.0	0.0	55.0	0.0
Water	15,841	250.4	425.3	16.8	11.1	12.6	7.3

6.3.5.2 Machine learning models

As mentioned in Section 5.3.4.3, we first investigated the problem of image classification, as this was the easier problem to solve. The image classification problem consisted of presenting an image to the trained model and the model labelling the image based on the object or objects the model thinks are contained in the image. The model achieved this by assigning a likelihood value to each object class (Section 5.3.4.3.1) based on how confident the model was that the class was in the image. The class with the highest likelihood was thus the class chosen by the model.

This simplistic process had to be modified in order to run the image analysis model over the video footage either as the stand alone model described in Section 5.3.4.3.1 or as part of Triton described in Section 5.3.4.3.3. First, a still image was taken from the video footage. This image was then broken down into patches of size 25 x 25 pixels. This patch size was chosen as this was larger than the mean size of the *dolphin* labels in the manual classification stage (see Section 5.3.4.2) to ensure that any

dolphin in the image would be fully enclosed by the patch. A patch was taken from the image every 25 pixels in each direction, with constant padding added to each side of the image as required to fit integer patches in each dimension. Each patch would then be passed to the model and the model would predict to which class the patch belonged. At the time of writing this report, the motion and Triton models required additional fine-tuning before they can be used for this purpose; hence, only results from the image model are included in this report.

Due to the issues mentioned in the sections above relating to the quality of the video data, we were not able to make reliable detections of dolphins with the video analysis models. Three different likelihood thresholds were chosen to illustrate the point that the current model cannot be used to detect dolphins due to the poor video quality (Figure 41). All dolphins were detected at the 50% and 65% likelihood threshold (upper panels in Figure 41). However, the rate of false positives was too high, rendering the usefulness of this to nil. Even at 90% likelihood (bottom left panel in Figure 42), which was meant to signal that the model was certain that the current patch is a dolphin, this situation did not really improve. Although there were less false positives, none of the dolphins were detected.

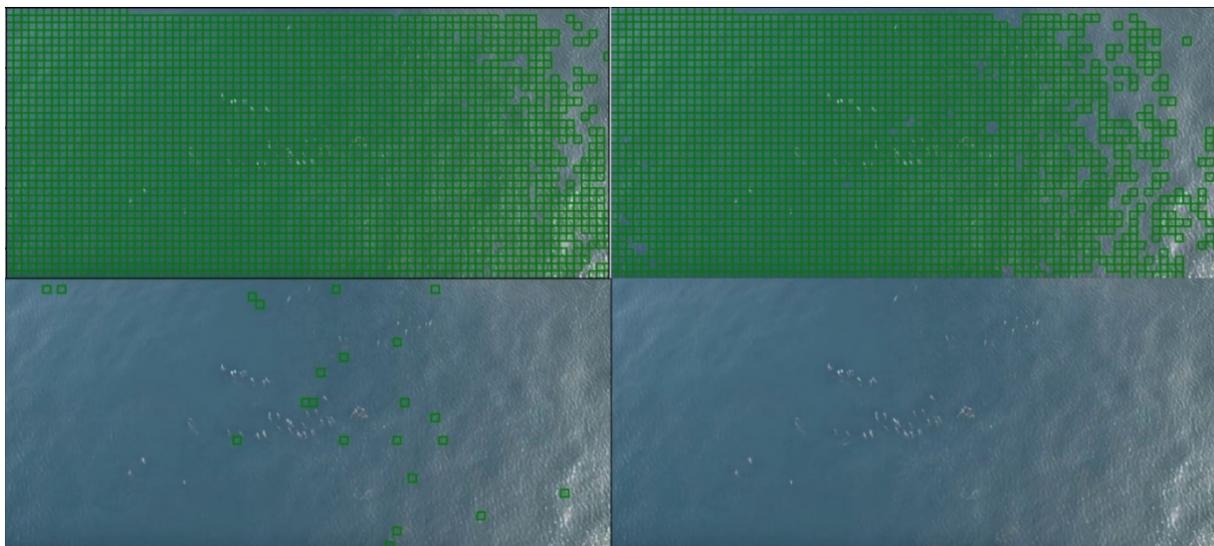


FIGURE 42. Example output from the image model. The bottom right panel shows the input image before patches were created. Top left (threshold of 50%), top right (threshold of 65%) and bottom left (threshold of 90%) panels summarize the output from the model using different likelihood thresholds. The green boxes indicated the patches that were labelled by the model as *dolphins*.

6.3.6 Comparison real-time (including review) with image analysis detections

At the time of writing this report, the limits of what can be achieved with the models on the existing footage was reached. However, new footage with better resolution would allow fine-tuning existing models. Then, we expect to achieve better results and to reliably obtain dolphin detections from the zigzag flights.

6.3.7 Matching sightings between platforms: flying bridge sightings versus drone detections

Of the six confirmed drone detections, two detections were of the same school (see below this section), reducing the number of trials for the flying bridge to five. Detections 1 (Figure 43) and 2 (Figure 44) were too far from the nearest flying bridge sightings to be a potential match. Even though detection 4 (Figure 45) was within less than 3km of flying bridge sighting 058, it was concluded that these were not a potential match based on auxiliary information: 1. drone detection 4 was a single dolphin (species undetermined) swimming underwater while changing directions and flying bridge detection 058 was a group of about 10 bottlenose dolphins (species code 018) that approached the

ship to bow ride; 2. It would have required the group of bottlenose dolphins to travel at a speed of about 10 knots for about 8 min to get from the location they were detected by the flying bridge to the location they were detected at drone considering the time lapse between the detections. It was deemed very unlikely behaviour for this group of bottlenose dolphins to be travelling away from the ship at high speed after first approaching and bow riding. Hence, these detections were not candidates for duplicate detections, and, thus, it was inferred that detections 1, 2, and 4 represent trials that were failures (see Section 5.3.1).

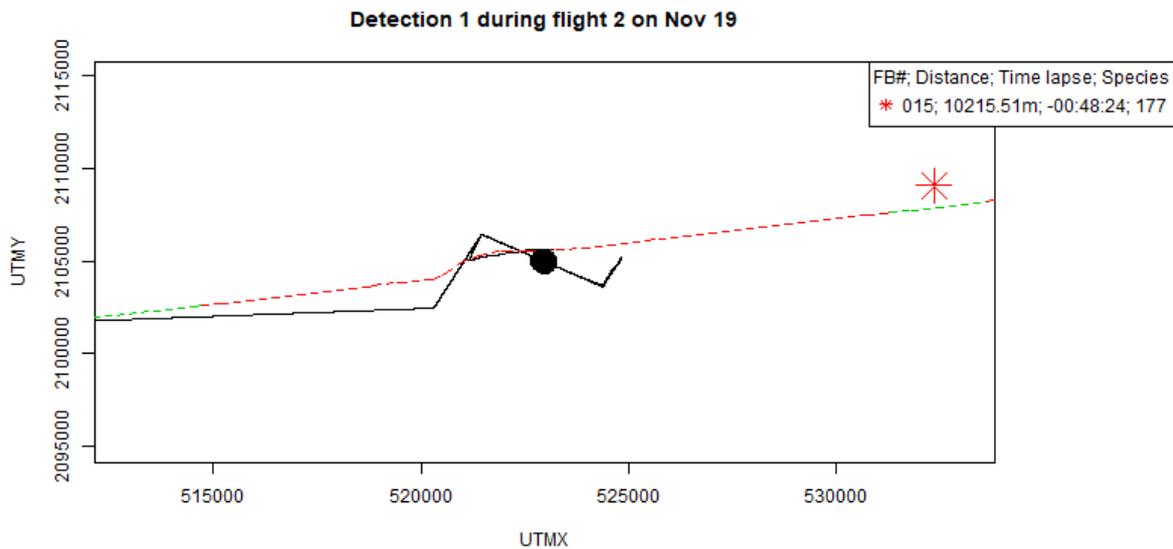


FIGURE 43. Location of drone detection 1 (black dot) in relation to flying bridge detections (stars) within a 10km radius, here only FB# 15 (177: unidentified small delphinids). Distance and time lapse between detections made by drone and flying bridge are given in the legend, where negative time means drone detection occurred before flying bridge detection. Drone track in black (with gaps in the GPS log), ship track as dashed line (red: on effort in passing mode, green: off effort). Units for UTM coordinates are meters.

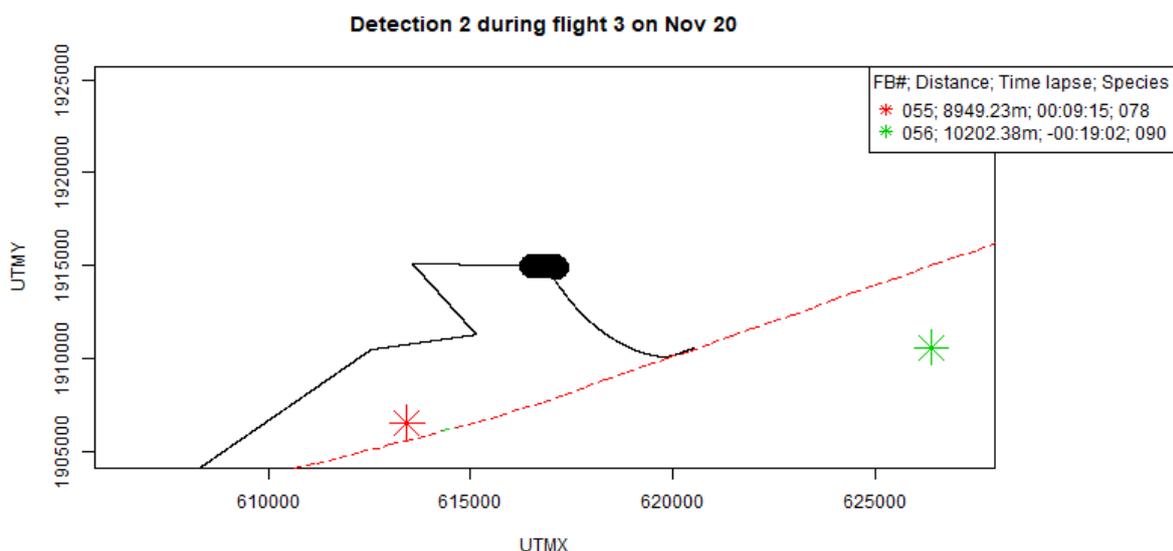


FIGURE 44. Location of drone detection 2 (black dot) in relation to flying bridge detections (stars) within a 10km radius, here FB# 55 (078: small whale) and 56 (090: spotted dolphin, unidentified)

subspecies). Distance and time lapse between detections made by drone and flying bridge are given in the legend, where positive/negative time means drone detection occurred after/before flying bridge detection. Drone track in black, ship track as dashed line (red: on effort in passing mode). Units for UTM coordinates are meters.

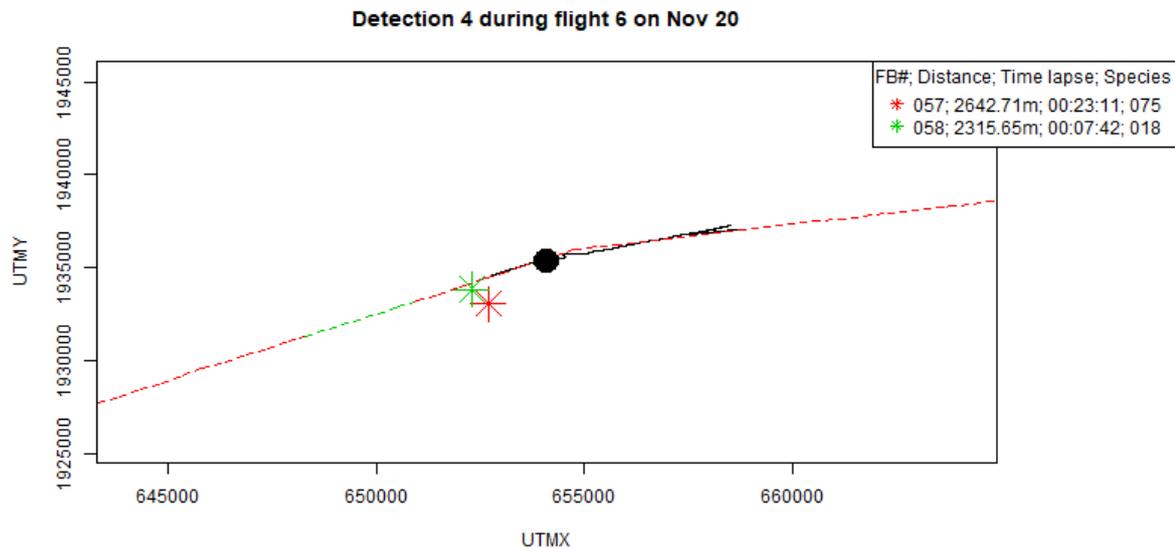


FIGURE 45. Location of drone detection 4 (black dot) in relation to flying bridge detections (stars) within a 10km radius, here FB# 57 (075: blue whales) and 58 (018: bottlenose dolphins). Distance and time lapse between detections made by drone and flying bridge are given in the legend, where positive time means drone detection occurred after flying bridge detection. Drone track in black, ship track as dashed line (red: on effort in passing mode, green: off effort). Units for UTM coordinates are meters.

For the remaining two detections, no determination could be made about their status. Drone detection 12 made during flight 3 on 21 November was detected about 12 min before the flying bridge sighting 71 and their respective locations were approximately 4.6 km apart (Figure 46). To cross a distance of 4.6 km in 12 min requires travelling at a speed of about 12.6 knots. Hence, it seems very unlikely that these detections represent a duplicate unless the dolphins detected by the drone were travelling at high speed directly towards the ship. The species identification made by the flying bridge observers for sighting 071, offshore spotted dolphin, would also be possible for drone detection 12 based on approximate size and behaviour (but impossible to confirm due to poor video quality). However, the travel direction of the dolphins sighted from the drone (Figure 47) was not in the direction towards the location of the flying bridge sighting. The dolphins sighted from the drone were travelling at slow to moderate speed with some porpoising behaviour, and could loosely be described as two subgroups with a total of five individuals. The dolphins spotted from the flying bridge were described as two tight subgroups with a mean best estimate of five individuals across the three on-watch observers, and were approaching the ship from the right of the ship to ride the bow (note that at the time of these detections, the heading of the ship was to the South, putting drone detection 12 on the right side of the ship; Figure 46). During the approach their behaviour was described as low and slowly swimming at about 2knots, with some porpoising. In conclusion, there is some probability that flying bridge sighting 71 was the same as drone detection 12 (based on numbers and aggregation), although spatial separation and direction of travel provided strong evidence against this. Species identification could provide the key information to make a better judgment. The analysis of this type of data where detections made by the two platforms cannot be matched with certainty would require new methods for incorporating uncertainty about duplicate identification.

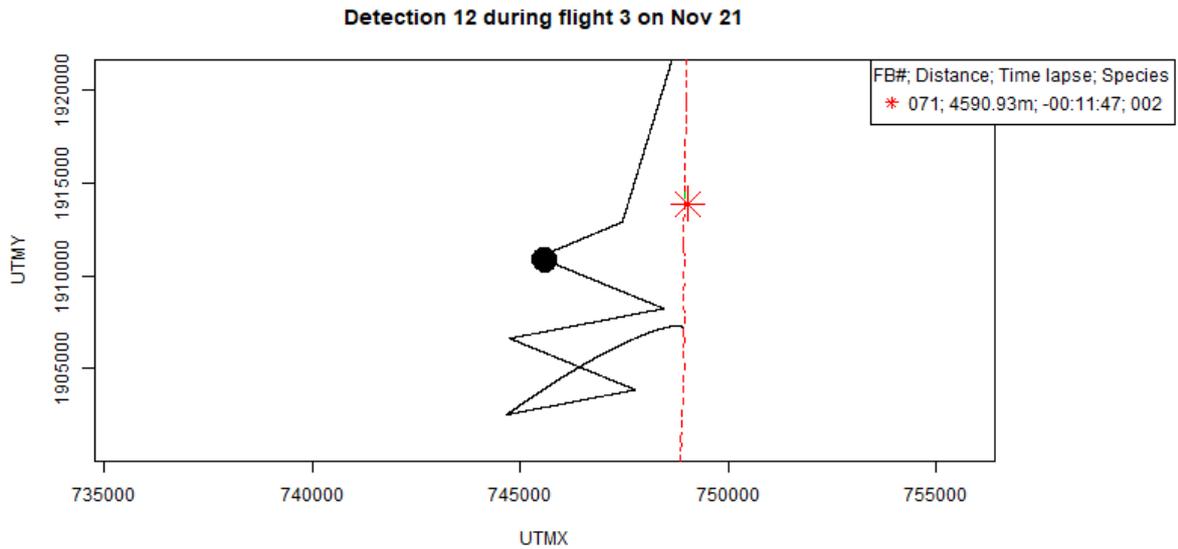


FIGURE 46. Location of drone detection 12 (black dot) in relation to flying bridge detection (star) within a 10km radius, here FB# 71 (002: offshore spotted dolphins). Distance and time lapse between detections made by drone and flying bridge are given in the legend, where negative time means drone detection occurred before flying bridge detection. Note that the drone detection was made during the second zigzag leg while the drone flew southeast. Drone track in black, ship track as dashed line (red: on effort in passing mode, green: off effort). Units for UTM coordinates are meters.



FIGURE 47. Screenshot of video footage taken during flight 3 on 21 November during drone detection 12. Red circle outlines the location of the detected dolphins, red arrow the direction of travel of the dolphins where an arrow pointing up would indicate the same direction as the drone was flying. Left panel shows the same image rotated to match the direction that the drone was travelling at the time of the detection (see Figure 46).

Lastly, from Figure 48 we infer that drone detections 71 and 72 were of the same school as their locations overlapped. Detection 71 was a large school with a loose formation, spread over a few hundred meters, and only two minutes had passed between the two detections during which time the drone had essentially double backed over the path it took while observing 71. Hence, detections 71 and 72 represent a single trial and will be treated as such below. Detection 71 was also displaying aerial behaviour, some of which may have been spinning (but poor video quality did not allow classifying the behaviour or the species identification) and had a few associated birds. Out of all the flying bridge sightings made within a 10 km radius, drone detections 71 and 72 may have also been of

the same school as flying bridge detection 170 based on the location and time lapse information. However, flying bridge sighting 170 was a mixed school of spotted and eastern spinner dolphins in a tight and clumped formation with spinner dolphins at the centre surrounded by the spotted dolphins. They were described as travelling at slow speed, initially of ~2 knots with a direction of 280° relative to the ship's bow, showing no reaction to the ship. No aerial behaviour was observed or any associated animals (e.g. birds) detected. Using all information available, it does not seem likely that drone detections 71 and 72, which are assumed to be the same school, were a match with flying bridge sighting 170, although this cannot be determined with enough certainty as key information (species identification and classification of aerial behaviour) was not available.

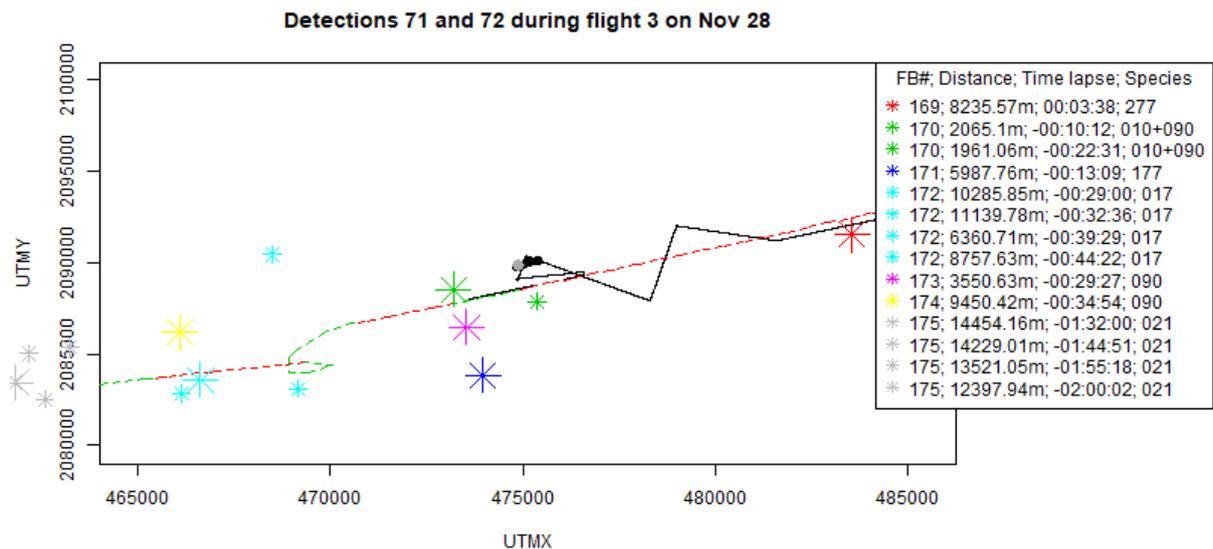


FIGURE 48. Location of drone detections 71 (black dots) and 72 (grey dots) in relation to flying bridge detections (stars) within a 10 km radius, here FB# 169-175 (large stars) and resights (small stars). Distance and time lapse between detections made by drone and flying bridge are given in the legend, where negative time means drone detection occurred after flying bridge detection. Species codes refer to unid. medium delphind (277), eastern spinner dolphins (010), unid. small delphinid (177), spotted dolphins (unid. subspecies, 090), short-beaked common dolphin (017) and Risso's dolphin (021). Units for UTM coordinates are meters.

6.3.8 MRDS data

Out of the five confirmed trials, three were failures and two could not be confirmed as successes or failures due to lack of sufficient information (see previous section). A sample size of five trials was not large enough for an MRDS analysis. Nonetheless, the perpendicular distances of the trials and their corresponding observation conditions are summarised in Table 21. The Beaufort sea state and swell height were taken from the information. The three failures ranged in perpendicular distance from 134 m to 5961 m.

TABLE 21. Summary of MRDS trials: drone detection number, trial outcome, perpendicular distance to the trackline completed by the ship, Beaufort sea state and swell height recorded by the flying bridge observers at the time of the drone detections at the time of the drone detections.

Detection	Trial outcome	Perpendicular distance (m)	Beaufort	Swell (feet)
1	Failure	654.26	2	4
2	Failure	5960.53	2	3
4	Failure	134.23	4	3
10	Uncertain	1274.80	2	4
71	Uncertain	1247.90	3	3

6.4 Test drone for collecting school size calibration data

6.4.1 Calibration flight effort modes and altitude

The calibration flights can be broken down into three effort modes: 1. outbound transit during which the drone was searching for the school; 2. with the school, when the drone was taking footage of the school; 3. inbound transit, when the drone was returning to the ship after all necessary video footage was taken (Figure 49). The average outbound transit time was about 16min (Table 15) while it took about 18 min on average to take calibration video footage. Inbound transit times were short (about 4 min) for calibration flights as the drone was kept at relatively short distances to the ship. The altitude of the drone when with a school was between 60 and 240 m, with an average altitude of 112 m. Most of the flight time when with a school was spent at 100 m.

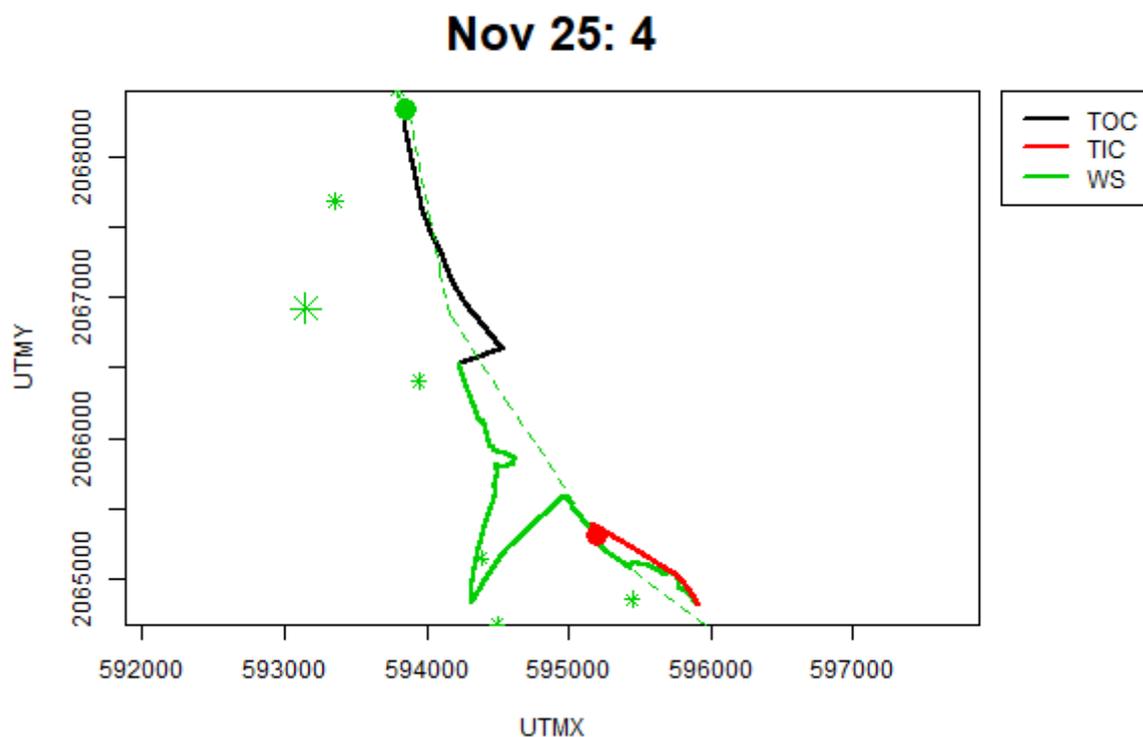


FIGURE 49. Example path of a calibration flight showing different modes with TOC: outbound transit (searching for school), WS: with school, and TIC: inbound transit (returning to the ship). Green and red dots indicate the locations of the drone during launch and landing, respectively. Green stars indicate the location of the initial sighting (big star) and resights (small stars). Date and flight number are given in the title. Units for UTM coordinates are meters.

6.4.2 Calibration flights: assessment of approach method

There were multiple challenges with regards to the calibration flights. The objective for the calibration flights was to obtain suitable video footage that would allow obtaining true counts of the calibration schools. In practice, the first challenge to meet this objective was to position the drone over the school. This required the direct coordination between the drone observers, flying bridge observers, drone pilots and the bridge. As the systems of flying bridge and drone pilots were not connected, this was not simply solved by using the GPS location of the flying bridge sighting as a waypoint for the drone. When this was attempted, it was found that by the time this location was relayed to the drone pilots, the schools had moved too far. Instead, the flying bridge observers continuously tracked the school and relayed information on bearing and distance from the ship to the drone observers and pilots. Using this information, the drone observers then directed the drone pilots as to which course and speed the drone should take and which angle the camera should be at. Flying bridge observers also directed the ship to stay at relatively close distance to the school without causing the school to react to the ship. We found that, given the equipment we had, the most effective method to get the school within view from the drone was to start with the ship pointing at the school with the school at a distance of 1nm. We then aligned the drone with the ship and flew out at slow speed in the direction of the school with the camera at an angle of 50° until the drone reached the school. Given the poor resolution of the video, it was not possible to detect dolphins at a great distance. With this in mind, we kept the altitude between 60 and 130m. For example, at an altitude of 60m above sea level, the ocean surface in the centre of the camera at 50° is at a distance of 93m. This increases to 155m if the drone is at 100m and to 202m if the drone is at 130m.

The flight depicted in Figure 50 shows our first attempt of trying to get the drone over a school, in this case killer whales. We note that in this example, we did not use the method described above immediately when the drone started the search of the school after realigning with the ship. The school was found but lost again as we only saw the school from a distance with a camera at an angle. After that, we used the method described above to relocate them with the drone.

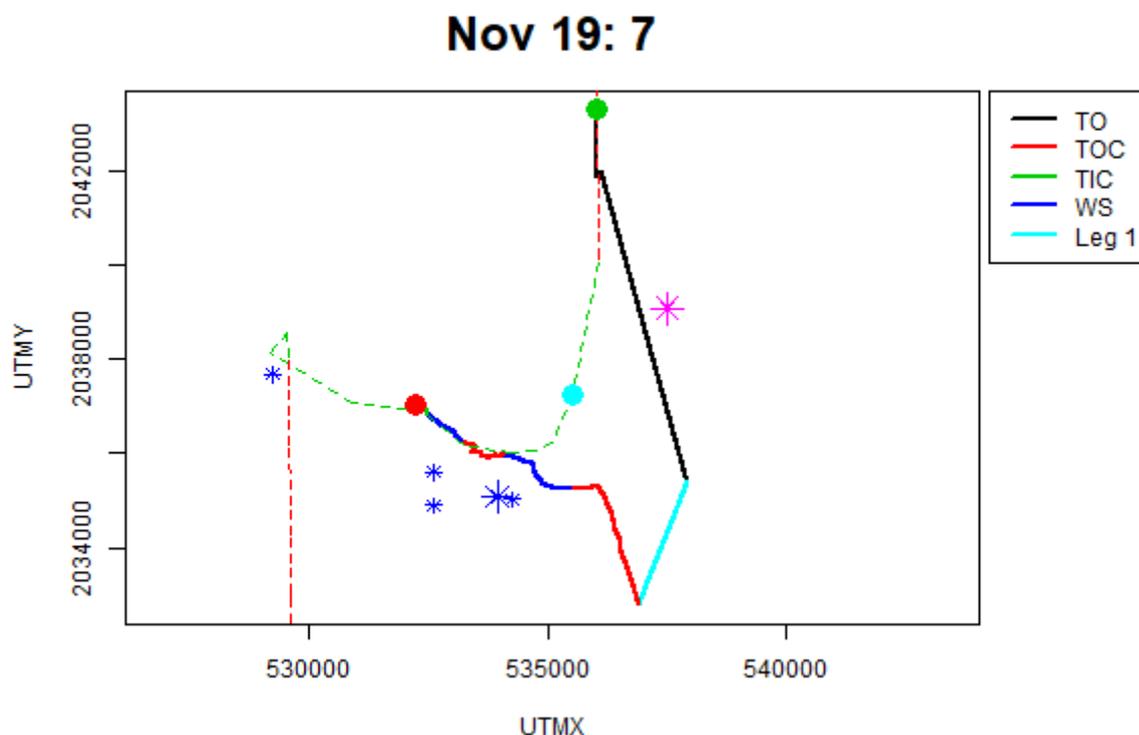


FIGURE 50. Path of example flight with initial zigzag effort which changed to calibration effort. TO: outbound transit to the first zigzag waypoint; TOC/TIC: outbound/inbound transit to/from calibration school; WS: with school; Leg 1 of zigzag flight component. Stars: locations of initial sightings (large) and resights (small) of flying bridge sightings 027 (purple, rough-toothed dolphins) and 028 (blue, killer whales). Green and red dot indicate the launch and land locations of the drone, turquoise dot the location of the ship when drone started the outbound transit towards the calibration school. Dashed line: ship track (red: flying bridge on effort in passing mode, green: off effort). Date and flight number given in title. Units for UTM coordinates are meters.

6.4.3 Video analysis

As stated in Section 6.3.5, at the time of writing this report, it has not been possible to obtain accurate counts for any of the video footage. This was mostly due to the poor quality of the video footage (see Figure 9 and Figure 41). The video quality recorded by the camera was recorded in full HD (1920 x 1080 pixels). Thus, for a camera aperture of 60 degrees, and a drone altitude of 100 m, the width of the ocean surface covered was 115m. Hence, we expected a ground resolution of about 6cm per pixel ($11500 \text{ cm}/1920 \text{ pixels}=5.99 \text{ cm/pixel}$), which would in principle be adequate to classify objects of potential interest as dolphins and count these. However, this resolution was much reduced by the video data collection process which included transmission, compression and screen recording (see Section 5.3.4.1). A ground resolution of 6cm/pixel would have allowed for detection of dolphins with a much higher success rate, while a ground resolution of 2 cm/pixel would generally allow for the identification of the species of the individual dolphins.

6.4.3.1 Manual counts

For six schools we were able to capture all clusters with the drone footage (Table 22). Manual counts were obtained for five of these schools and, hence, are valid calibration schools. The sixth school was too large to be counted manually with too many animals moving in and out of sight to keep track of individuals with a sufficient degree of confidence. The sweeps counted were less than the sweeps recorded (see Section 5.4.4) for two schools as it was uncertain if the entire school was visible within

the sweep due to glare issues, thus, emphasizing the need for recording footage during multiple sweeps.

TABLE 22. Summary of counts for the six schools for which the entire school was captured with the drone video and, hence, would be valid calibration schools if true counts could be obtained. Sweeps refer to the parts of a drone flight during which the entire school was captured at least once. A final count of NA indicates that the school could not be counted manually. Date is shown as day.month.

Date	Flight	Species	Sighting number	Could manual counts be obtained?	Sweeps recorded	Sweeps counted	Final count
24.11	1	010	119	No	4	3	NA
25.11	2	090	122	Yes	5	5	55
27.11	4	015	157	Yes	1	1	10
27.11	7	015	159	Yes	1	1	9
28.11	6	021	175	Yes	6	3	56
28.11	7	032	178	Yes	5	5	36

6.4.3.2 Comparison of image analysis counts with manual counts

At the time of writing this report, the limits of what can be achieved with the models on the existing footage was reached. However, new footage with better resolution would allow fine-tuning existing models. Then, we expect to achieve better results and to reliably obtain dolphin detections from the calibration flights and obtain counts of dolphins within the schools.

7. CONCLUSIONS

In summary, as regards preparation for a main survey, the following conclusions about equipment and methodology can be drawn from the trial survey project:

- The Jorge Carranza can be used as a survey vessel for the next ETP survey upon which our team of experienced observers, in combination with the ship’s command, were able to implement the NMFS survey protocol.
- Flying bridge equipment worked well, although a few fixes and alterations are needed for the main survey.
- No significant differences in detection probabilities between observation platforms of the Jorge Carranza and previous ETP survey vessels could be identified.
- The Jorge Carranza with its custom-made drone platform can be used for conducting drone operations; drones could be launched and landed in Beaufort sea states up to and including 5.
- We were not able to show that the zigzag flights could be successfully implemented with the Seahawk drone, which was not the drone of our choice but was provided to us for the project. In particular, for implementing zigzag flights during closing-mode effort a drone with much longer endurance is needed.
- Even with additional highly skilled pilots and crew on board the survey vessel to man multiple launches and landings, the Seahawk drone does not have the endurance necessary to make it a viable option for the zigzag flights during the main survey, and its use during a main survey for zigzag flights would be a major safety concern.

- A more flexible solution compared to uploading fixed waypoints pre-flight is needed for directing the drone along zigzag survey legs, including the option to adjust the flight plan and waypoints easily in-flight.
- Detections of cetaceans via the drone could be made both via real-time observation and via video analysis and geo-referencing the detections with drones is generally possible; hence, using drones to collect MRDS data is possible during the main survey, as long as much better cameras and video data collection and transmission systems are used than during the trial survey.
- To achieve the required sample size for assessing trackline detection probability during the main survey, cameras with much better resolution will be needed for the zigzag flights to cover a larger area and provide better ground resolution. Higher resolution video would also allow the drone to operate at higher altitudes while maintaining the same ground resolution. Increased altitude would also increase the area covered by the drone and thus increasing the sample size of detected schools for the trackline detection probability assessment.
- The video data collection (for which the main source was on-screen recording of transmitted video, contrary to our protocol) and transmission systems used during the trial survey will not be adequate for a main survey, and these must include continuous on board recording of high-resolution video imagery and ancillary data.
- Recording of camera angles in three dimensions (in addition to pitch and roll of the drone) on board the drone is required for obtaining accurate assessments of the swath width and area covered by the drone during the zigzag flights as well as estimated lengths of objects of potential interest for the image analyses.
- Flying bridge survey protocol should incorporate logging more resights while the drone is conducting zigzag flights in order to improve matching of duplicate detections. This could be done by the main observers, in the case they resight the schools during their regular scanning, or by an additional observer on bigeyes.
- A more rigorous protocol for note taking by the drone observers is needed, in combination with suitable software for log-keeping for both zigzag and calibration flights to ensure essential information along with its sources (e.g. flying bridge, bridge, drone observers, pilots) is logged.
- There is a need for an alternative to drone-Wincruz software, e.g. simple note-taking software that would automatically record the time and which source the information comes from, as well as allow tracking ship and drone GPS and the sightings made by either platform simultaneously.
- Calibration flights can be completed successfully with the Seahawk drone; however, a higher resolution camera is needed to identify *all* individuals to species (which is not possible by zooming in with the drone camera) and to ensure animals swimming in close proximity to each other can be distinguished.
- Recording multiple sweeps across a given calibration school with slightly varying camera angles is important to alleviate potential glare issues
- For video analysis models, using image data or motion data alone does not yield as good a result as using both in one combined model. Hence, for a given resolution, video is preferred over only recording still frames. Video also adds the advantage that it allows tracking individuals through the frame, an essential feature used for manually counting the calibration schools during this project. However, should the resolution of the still frames improve sufficiently for the next phase of the project, the image-only models applied to single frames might be sufficient to detect dolphins and obtain counts reliably.
- Using a polarising filter on the drone camera helps alleviate some of the glare issues in the videos.

8. RECOMMENDATIONS

We recommend that, prior to a main survey, a different drone-camera system with longer endurance and greater video resolution than the Seahawk should be tested in a short sea-trial on a vessel from which drones can be launched under similar conditions to those aboard the Jorge Carranza. We recommend that before such a trial, any potential drone provider should provide an assessment of their services, taking into account the following:

- A. For the zigzag flights, they should aim to provide a 100% coverage of a corridor with half-width $w = 5.5$ km while maintaining a ground resolution of 5 cm/pixel in the video and while maintaining station ahead of the ship at 5 nm or further.
- B. For the calibration flights, drones should be flown that can be easily manoeuvred to and hovered above the calibration school as well as be flown in slow sweeps across the school with cameras taking video footage of 2 cm/pixel ground resolution that is both recorded on-board the drone and transmitted back to the ship.
- C. Drones should be able to fly at the height and cruising speed required for A. for a minimum of 4hrs in order to implement zigzag flights during closing-mode effort; endurance of 1hr are sufficient for drones flying the calibration flights.
- D. Drone providers should demonstrate their ability to fly the drones in a zigzag pattern (or parallel lines) while maintaining station at 5nm ahead of the ship, and to adjust their flight plan mid-flight to accommodate flexible flight paths required when the flying bridge operates in closing-mode effort and turns on sightings.
- E. For all flights, the video footage must be saved to the disk on the drone with minimal compression.
- F. Drone providers should suggest solutions for dealing with glare issues on the ocean surface, e.g. use of polarised filters.
- G. Information about the drone and camera should be continuously recorded, including GPS, altitude, pitch and tilt of the drone, camera angle in three dimensions and focal length.

The duration of such a sea trial should be long enough to collect data suitable for improving video analysis models. This requires that schools of dolphins should be captured with the video recorded during the zigzag flights flown using the parameters required to implement A. (i.e. the drone altitude and speed as well as the video resolution). This will require the capability for at-sea review of the video imagery after the drone has returned to the ship, and the necessary software and hardware to conduct the post-flight video review should be made available by the potential drone provider. The necessity of conducting a further sea trial and post-trial image analyses will need to be factored into the timing of the main survey.

9. ACKNOWLEDGEMENTS

The R/V Dr Jorge Carranza Fraser was provided by the Mexican government (INAPESCA). We thank the crew for accommodating us and executing our protocol. Funding was provided by the Pacific Alliance for Sustainable Tuna (PAST). We extend special thanks to a number of people who generously provided us with guidance during our preparation for the trial survey: Lisa Ballance, Tim Gerrodette, Annette Henry, Wayne Perryman and Jay Barlow (all NMFS). We further thank Robert Holland (SWFSC) for his help with the customised WinCruz software. We thank Ramón Isaac Rojas González (INAPESCA) for serving as the logistical coordinator, Alvin Suárez (PAST) for dealing with countless logistical matters, José Carranza Beltrán (Pesca Azteca) and Manuel Vazquez Escudero (Pinsa) for providing equipment.

10. REFERENCES

- Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 175-185.
- Barlow, J. (2015). Inferring trackline detection probabilities, $g(0)$, for cetaceans from apparent densities in different survey conditions. *Marine Mammal Sciences*. 31(3), 923-943.
- Bartlett, M. S., Littlewort, G., Lainscsek, C., Fasel, I., & Movellen, J. (2004). Machine learning methods for fully automatic recognition of facial expressions and facial actions. *IEEE International conference on Systems, Man and Cybernetics*. IEEE.
- Batista, G. E., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD explorations newsletter*, 20-29.
- Borchers, D. (2012). A non-technical overview of spatially explicit capture-recapture models . *Journal of Ornithology*.152, (2), 435-444.
- Britannica, E. (2020, July 28). *Triton*. Retrieved from Britannica: <https://britannica.com/topic/Triton-Greek-mythology>
- Buckland, S., Rexstad, E., TA, M., & Oedekoven, C. (2015). *Distance Sampling: Methods and Applications*. Methods in Statistical Ecology, Springer.
- Burt, M., Borchers, D., Jenkins, K., & Marques, T. (2014). Using mark–recapture distance sampling methods on line transect surveys. *Methods in Ecology and Evolution* 5(11) , 1180-1191.
- Cyc. (2008, February 16). *File:Svm separating hyperplanes.png*. Retrieved from Wikimedia: [https://en.wikipedia.org/wiki/Support_vector_machine#/media/File:Svm_separating_hyperplanes_\(SVG\).svg](https://en.wikipedia.org/wiki/Support_vector_machine#/media/File:Svm_separating_hyperplanes_(SVG).svg)
- Deng, J., Socher, R., Li, J., Li, K., & Fei-fei, L. (2009). ImageNet: a large-scale hierarchial image database. *IEEE Conference on computer vision and pattern recognition*. IEEE.
- Geron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media.
- Gerrodette, T., & Forcada, J. (2005). Non-recovery of two spotted and spinner dolphin populations in the eastern tropical Pacific Ocean. *Marine Ecology Progress Series* 291, 1-21.
- Gerrodette, T., Perryman, W., & Oedekoven, C. (2019). Accuracy and precision of dolphin group size estimates. *Marine Mammal Science* 35(1) , 22–39.
- Gerrodette, T., Watters, G., Perryman, W., & Balance, L. (2008). *Estimates of 2006 dolphin abundance in the eastern tropical Pacific with revised estimates from 1986-2003*. La Jolla, CA: NOAA Technical Memorandum NOAA-TM-NMFS-SWFSC-422.
- Gujunoori, S., & Oruganti, M. (2017). Tracking and Size Estimation of Objects in Motion using Optical flow and K-means Clustering. *2nd International Conference On Emerging Computation and Information Technologies (ICECIT)*, (pp. 1-6).
- Harley, A. W. (2015). An Interactive Node-Link Visualization of Convolutional Neural Networks. *ISVC*, 867-877.

- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE.
- He, T., Zhang, Z., Zhang, H., Zhang, Z., & Xie, J. (2019). Bag of tricks for image classification with convolutional neural networks. *IEEE conference on computer vision and pattern recognition*.
- Huang, G., Liu, Z., Van Der Maaten, L., & Weinberg, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE.
- Hubel, D. H., & Torsten, N. W. (2004). *Brain and visual perception: the story of a 25-year collaboration*. Oxford University Press.
- Hubel, D., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, pp. 215-243.
- Hulstaert, L. (2018, August 18). *A Beginner's Guide to Object Detection*. Retrieved from DataCamp: <https://www.datacamp.com/community/tutorials/object-detection-guide>
- Johnson, K., Punt, A., & Lennert-Cody, C. (2018). *Report of the Inter-American Tropical Tuna Commission Workshop on Methods for Monitoring the Status of Eastern Tropical Pacific Ocean Dolphin Populations, 18–20 October 2016*. La Jolla, California: SWFSC.
- Kinzey, D., Olson, P., & Gerrodette, T. (2000). *Marine mammal data collection procedures on research ship line-transect surveys by the Southwest Fisheries Science Center*. La Jolla, California: Southwest Fisheries Science Center, Administrative Report LJ-00-08, 32 p. .
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, (pp. 2278-2324).
- Lin, C. (2019). *Introduction to Motion Estimation with Optical Flow*. Retrieved from nanonets: <https://nanonets.com/blog/optical-flow/>
- Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. *Proceedings DARPA Image Understanding Workshop*, 121-130.
- Maire, F., Mejias, L., & Hodgson, A. (2014). A convolutional neural network for automatic analysis of aerial imagery. *International Conference on Digital Image Computing: Techniques and Applications*. IEEE.
- Marie, F., Mejias, L., Hodgson, A., & Duclos, G. (2013). Detection of dugongs from unmanned aerial vehicles. *IEEE/RSJ international conference on intelligent robots and systems*. IEEE.
- Marques, T., Thomas, L., Fancy, S., & Buckland, S. (2007). Improving estimates of bird density using multiple-covariate distance sampling. *The Auk*, 124(4), 1229-1243.
- McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, 205-205.
- McInnes, L., Saul, N., & Horn, S. (2020, July 29). *How HDBSCAN Works*. Retrieved from The hdbscan Clustering Library: https://hdbscan.readthedocs.io/en/latest/how_hdbscan_works.html
- Mori, S., Nishida, H., & Yamada, H. (1999). *Optical character recognition*. John Wiley & Sons, Inc.

- Moshen, H., El-dashan, E. S., El-horbaty, E., & Salem, A. (2018). Classification using deep learning and neural networks for brain tumors. *Future Computing and Informatics Journal*, 68-71.
- Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M., Packer, C., & Clune, J. (2018). Automatically identifying, counting and, describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences*. Proceedings of the National Academy of Sciences.
- Oedekoven, C., Buckland, S., Marshall, L., & Lennert-Cody, C. (2018). *Design of a survey for eastern tropical Pacific dolphin stocks*. San Diego, California (USA): Agreement on the international dolphin conservation program – IATTC 37TH MEETING OF THE PARTIES, DOCUMENT MOP-37-02. https://www.iattc.org/Meetings/Meetings2018/IATTC-93/Docs/English/MOP-37-02_Design%20of%20a%20survey%20for%20eastern%20tropical%20Pacific%20dolphin%20stocks.pdf
- Rankin, S., Oedekoven, C., & Archer, F. (2020). Mark recapture distance sampling: using acoustics to estimate the fraction of dolphins missed by observers during shipboard line-transect surveys. *Environmental and Ecological Statistics* 27, 233–251.
- Rasband, W. S. (2008). ImageJ. Bethesda, Maryland, USA: U.S National Institutes of Health. Retrieved from <https://imagej.nih.gov/ij>
- Reilly, S. (1990). Seasonal changes in distribution and habitat differences among dolphins in the eastern tropical Pacific. *Marine Ecology Progress Series* 66, 1-11.
- Schwartz, L., Gerrodette, T., & Archer, F. (2010). Comparison of closing and passing mode from a line-transect survey of delphinids in the eastern Tropical Pacific Ocean. *J. Cetacean Res. Manag.* 11(3), 253–265.
- Shaikh, S. H., Saeed, K., & Chaki, N. (2014). Moving object detection using background subtraction. In *Moving Object Detection Using Background Subtraction* (pp. 15-23). SpringerBriefs in Computer Science. Springer, Cham.
- Sharma, V. (2018, October 15). *Deep Learning – Introduction to Convolutional Neural Networks*. Retrieved from Vinod Sharma's Blog: <https://vinodsblog.com/2018/10/15/everything-you-need-to-know-about-convolutional-neural-networks/>
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint*.
- Zhang, J., Xie, Y., Wu, Q., & Xia, Y. (2019). Medical image classification using synergic deep learning. *Medical image analysis*, 10-19.

11. APPENDICES

Appendix 1: Drone operations protocol

Appendix 2: Seahawk general technical specifications

Appendix 3: Flight paths of drones

DESIGN OF A SURVEY FOR EASTERN TROPICAL PACIFIC DOLPHIN STOCKS

Protocol for drone operations

Cornelia S. Oedekoven¹ and Stephen T. Buckland¹

¹ Centre for Research into Ecological and Environmental Modelling, University of St Andrews, The Observatory, Buchanan Gardens, KY16 9LZ, UK

Goals for using the drone

Estimate absolute abundance of the priority stocks

One of the critical assumptions for estimating absolute abundance using the conventional distance sampling methods that were used previously during eastern tropical Pacific line transect surveys for cetaceans (e.g. as in Gerrodette *et al.* 2008) is that all schools on the transect line are detected. However, questions have been raised concerning whether probability of detection of schools on the transect line – often referred to as $g(0)$ – is close to one in all sea states up to Beaufort 5 (Barlow 2015). Therefore, to meet Objective 2 (Estimate absolute abundance of the priority stocks, Oedekoven *et al.* 2018), methods are needed to estimate this probability, and how it varies by sea state. Given that some schools that were initially close to the line may be evading detection, these methods should accommodate responsive movement and behavioral responses.

We propose to use the drone to collect data that will allow us to estimate $g(0)$ (the probability that schools on the trackline are detected). The preferred method for addressing the $g(0)$ issue for the ETP survey is mark-recapture distance sampling (MRDS, e.g. Borchers 2012). In comparison to conventional distance sampling where, e.g., line-transect data are collected from a single platform, MRDS methods require double-observer platform data. Here, detections made from one platform, say platform 2, represent trials for the other platform, say platform 1. In this context, trial outcomes refer to whether or not platform 1 detects a group of dolphins initially detected by platform 2. It is crucial that the two observation platforms are such that platform 2 does not influence the observers on platform 1 and that sightings are matched across platforms correctly.

For this survey, a drone would survey the area in front of the ship during all daylight hours while flying bridge observers are on-effort, likely beyond the maximum sighting range of the observers, and serve as platform 2. A drone observer monitors the video footage from the drone transmitted back to the ship in real-time. The sightings made by the drone observer represent the trials for the flying bridge observers on platform 1. An observer monitoring images in real-time identifies duplicate detections – those detected by both drone and observers. **It is crucial, that sightings are matched and identified as duplicates correctly.** The drone will also record high-resolution video and still images for later analysis. Recordings of the video footage might allow detecting additional schools during post-cruise image analysis.

School size calibration

During previous ETP surveys, calibration of school size estimates for ETP observers was done by comparing estimates to counts from aerial photographs taken from manned helicopter or fixed-wing aircraft (e.g. Gerrodette *et al.* 2018). For this survey, we propose to replace the helicopter or fixed-wing aircraft with a drone to collect equivalent still photographs or video of the dolphin groups. During the trial, we need to assess the practicality of using such a drone as the aerial platform for collecting suitable high-resolution imagery. Video and still camera equipment aboard the drone will record the

high-resolution imagery, which will allow the drone to operate at a height where disturbance of the dolphins is highly unlikely. This is not possible with observers on an aircraft.

With these goals in mind, the video footage that is transmitted back to the ship in real-time needs to be good enough to detect dolphin schools, including those that consist of only a few animals, while providing good coverage of the trackline (see below). The imagery (video and/or stills) that is recorded on the drone for calibration purposes needs to be good enough to determine species identification of individual dolphins while covering the entire school with only a few passes over the school.

Drone operations

The drones will be operated from the vessel whenever the primary observer teams on the flying bridge are on-effort, i.e. during all daylight hours when the Beaufort sea state is below 6. On-effort generally entails that the ship moves along the transect lines at a constant speed of 10 nm. However, when the primary team detects a school of cetaceans, the ship closes on the school for obtaining school size estimates and species id which generally requires course and speed changes of the ship (see Diagram 1 below for an example of closing mode without drone operations). As $g(0)$ is likely to be lower in poor conditions, the drone would ideally need to operate up to Beaufort 5, and it would need to be able to stay aloft for extended periods, to ensure that it is searching for most of the time that the observers are on effort. Weather and drone technology permitting, the drone will be launched at sunrise just before flying bridge observers begin daily effort and retrieved at sunset at the conclusion of effort. This will provide maximum time use of the drone and minimal interruptions for the survey effort of the primary observer team which requires constant speed of 10 knots (at least 8 knots). It is at the discretion of the drone pilots to request changes in speed of the ship for launching and landing of the drone.

Should the drone not allow for such long uninterrupted airtime (~12 hours), two or three drones operating in shifts need to be operated in rotation. For example, if the drone only allows for one hour of flight uninterrupted flight time before recharging, a second drone needs to be launched in time to replace the former before it needs to return to the ship to allow uninterrupted monitoring of the areas in front of the ship. The drone flying in the zig-zag pattern across the transect line is always identified as the main drone. In this example the first drone is the main drone until it is replaced by the second drone; then, the second drone becomes the main drone until it is replaced again by the former, and so on until the end of operations for the day. It is always the main drone that will be recording and transmitting the video and GPS locations back to the ship. If two drones are scheduled to operate in rotation, it is important to have sufficient batteries and chargers on board to allow for this rotation schedule.

The main drone will operate several km ahead of the vessels, flying in either a zig-zag pattern or in parallel lines back and forth across the trackline out to a defined distance, e.g. ~3nm (the exact distance depends on the flight speed of the drone and the targeted coverage probability on the trackline and should be determined during the trials) either side of the line, with the angle of the zig-zag or the distance between the parallel lines determined to allow the drone to maintain station ahead of the ship (see Diagram 2 below for an example). It will fly at an altitude that is unlikely to generate a response from the dolphins. It will be part of the trials to determine the best flight pattern and the altitude at which the drone will be flying.

High-resolution video will be recorded on the drone to allow examining the footage for species identification, school composition and obtaining (at least approximate) true school size counts used for observer calibration at the analysis stage. This imagery will also be transmitted to the ship in real-time and monitored by drone observers on the vessel. If a school is detected by the drone observer monitoring the drone footage, they will alert the cruise leader (but not the flying bridge observers),

and if the flying bridge observers subsequently detect the same school, both the drone and the ship will close on the school to secure better data on school size and species present, and to record any movement of the school towards or away from the line following initial detection by the drone. If it is considered feasible for the drone to monitor a school without alerting shipboard observers, it will do so for schools detected from the air, until either the shipboard observers detect the school or it passes abeam; however, to avoid cueing observers, it is likely that the drone will need to remain some distance ahead of the vessel unless the shipboard observers detect the school, as any change in flight pattern when a school is detected by the drone might cue observers to the presence of a school.

Drone coverage probability on the trackline

We did a preliminary evaluation of what the coverage probability of the drone (the proportion of the transect line that would be captured by the drone footage) would be based on a zigzag flight pattern of the drone intersecting the transect line at a constant distance ahead of the ship. This constant distance depends on the behavior of the dolphin schools and will be determined during the trials. **The aim is to capture the dolphin schools with the drone footage before they reacted to the presence of the ship.**

Here we used an endurance speed of the Flexrotor drone of 85km/hr and the ship survey speed of 18.5km/hr (10knots). If the drone is to survey in a zigzag pattern out to 4km to either side of the transect line, the drone is expected to cross the transect line at approximately 1.78km intervals. If the drone flies at 300m above sea level where the strip width in wide-angle mode is 368m (Table 9), the resulting coverage probability of the transect line is 0.21 (0.368km/1.78km). If the drone flies at 500m, the resulting coverage probability of the transect line is 0.34 (0.613km/1.78km). It is important, however, that the drone footage will provide enough ground resolution to detect dolphin schools.

TABLE 1. Drone flight height, resulting survey strip width covered by the video footage as well as ground resolution (cm per pixel) using 720 x 1280 video (GR_{1280px}) and 1080 x 1920 video (GR_{1920px}) shown for the maximum and minimum angle of the Trillium camera Orion HD50 (<http://w3.trilliumeng.com>).

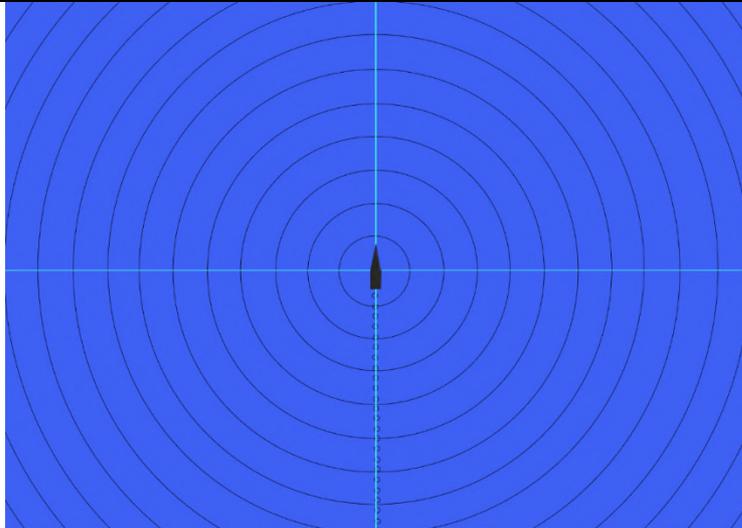
Height (m)	Wide angle: 63°			Narrow angle: 2.2°		
	Strip width (m)	GR _{1280px} (cm)	GR _{1920px} (cm)	Strip width (m)	GR _{1280px} (cm)	GR _{1920px} (cm)
100	123	9.6	6.4	4	0.3	0.2
200	245	19.2	12.8	8	0.6	0.4
300	368	28.7	19.2	12	0.9	0.6
400	490	38.3	25.5	15	1.2	0.8
500	613	47.9	31.9	19	1.5	1.0
600	735	57.5	38.3	23	1.8	1.2
700	858	67	44.7	27	2.1	1.4
800	980	76.6	51.1	31	2.4	1.6
900	1103	86.2	57.5	35	2.7	1.8
1000	1226	95.8	63.8	38	3	2.0

Capturing footage for school size calibration will bring further challenges in that these images require both a ground resolution good enough for species or stock identification as well as a wide enough strip to cover the school with as few as possible passes. We acknowledge that using the drone for our purposes will be a challenge, in particular with regards to school size calibration. However, it is the purpose of the trial to determine the feasibility of these operations.

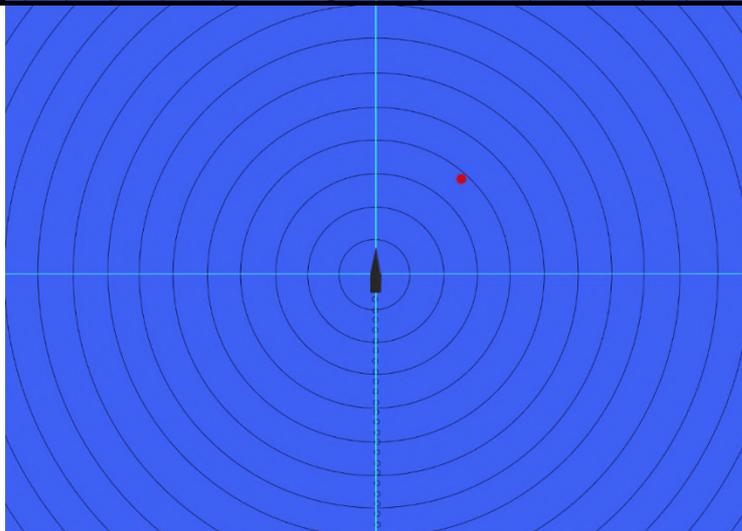
General closing mode effort

In this section we describe the NMFS closing mode procedure (Kinzey et al. 2000) by illustrated example.

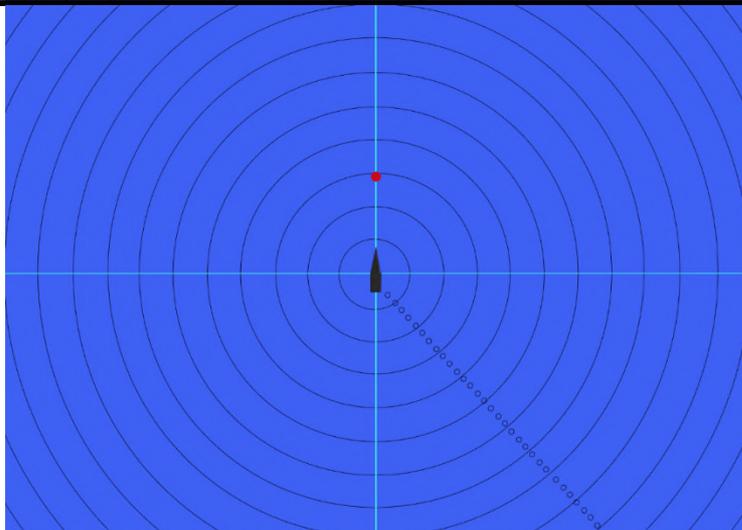
Diagram 1 showing an example of closing mode effort without considering the drone. Black polygon in the centre is the ship at the current position where the direction of current travel is always pointing directly up in the image (as in program WinCruz, the line transect data collection software) and small black circles are recent ship positions, black concentric circles indicate increasing distance to the ship in 1nm increments, red dot indicates sighting position in relation to the ship.



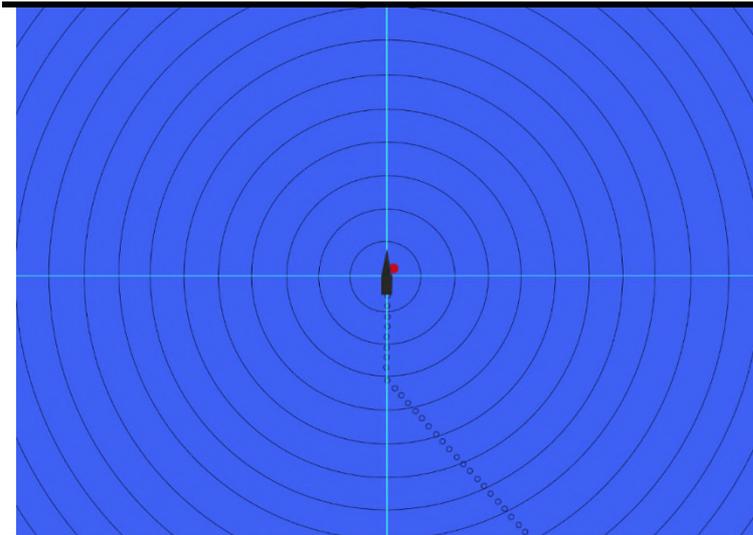
1. Ship transits along the transect line, observers on effort.



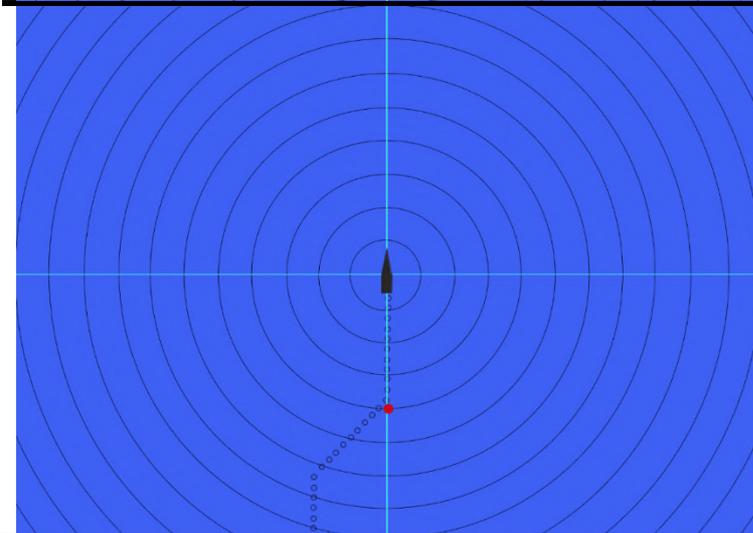
2. Observers on the flying bridge spot a school (red dot). Note that all sightings are closed on that fall within a 3nm perpendicular distance from the transect line (or rather from the projected transect line given that it may change due to closing mode effort – see 5. in this diagram for details).



3. Ship turns to approach the school. Observers are logged off-effort. All on-watch observers focus on the school.



4. Ship closes on school for school size estimates and species id. This may require more than one pass through the school.



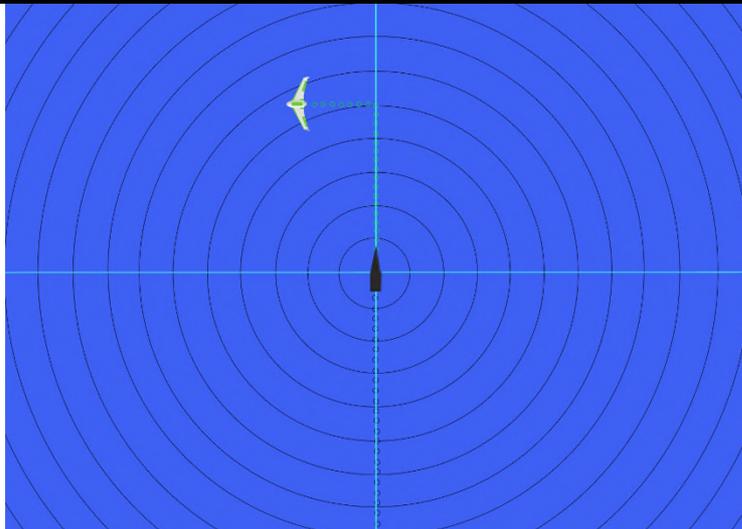
5. After finishing with the sighting, the ship continues along a course parallel to the transect, without returning to the transect line first, and observers resume search effort. Note that if the sighting or a sequence of sightings has taken the ship >10nm away from the transect line, the ship resumes searching on a 20° course back to the original transect.

Flight path of the drone during closing mode effort

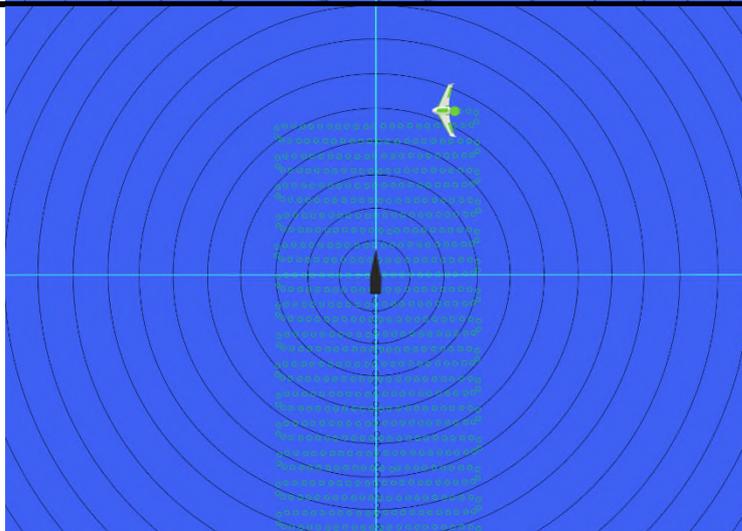
Here, we illustrate a possible flight path of the drone in relation to the ship while the ship is in closing mode effort. Note that the feasibility and suitability needs to be tested in the field. For illustration purposes we used the following parameters for the drone flight in this example, but these will likely need to be updated:

1. The drone flies in parallel lines out to 3nm on either side of the trackline (3nm perpendicular distance of a detected school from the transect is the maximum distance for the ship to close on; however, **the width of the zig-zag will depend on the speed of the drone and the coverage probability on the trackline we aim for, Table 1**).
2. The drone flies 5nm ahead of the ship. This distance will depend on the behavior of the dolphins. While it will be easier to keep this distance relatively short for identifying duplicate sightings between the drone and flying bridge, **we need to capture the dolphins with the drone before they react to the ship**. We anticipate that this distance could be much further than 5nm and the drones should be able to fly 20 – 30 nm ahead of the ship. The distance that the drone needs to fly ahead of the ship will be tested during the trial in November.

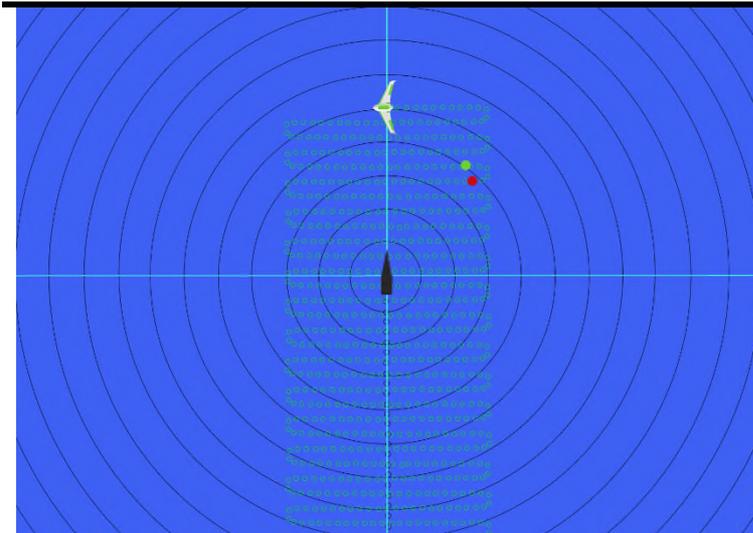
Diagram 2 showing an example of the drone flight path during closing mode effort. Black polygon in the centre is the ship at the current position where the direction of current travel is always pointing directly up and small black circles are recent positions, black concentric circles indicate increasing distance to the ship in 1nm increments. Green-white symbol is the drone and small green circles indicate the recent positions. Red and green dots indicate the ship and drone sighting positions, respectively. Note that in this example, no tracking of the drone sighting occurs. This might require that the drone continues passing over the school until the flying bridge detects the school or passes the beam.



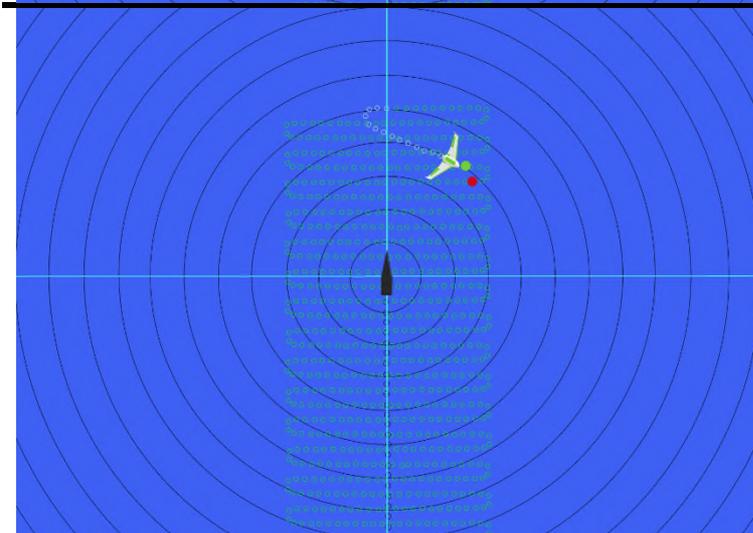
1. After take-off, the drone flies to 5nm distance from the ship and begins search mode in zig-zag lines crossing the transect line. Ship transits along the transect line, observers on effort.



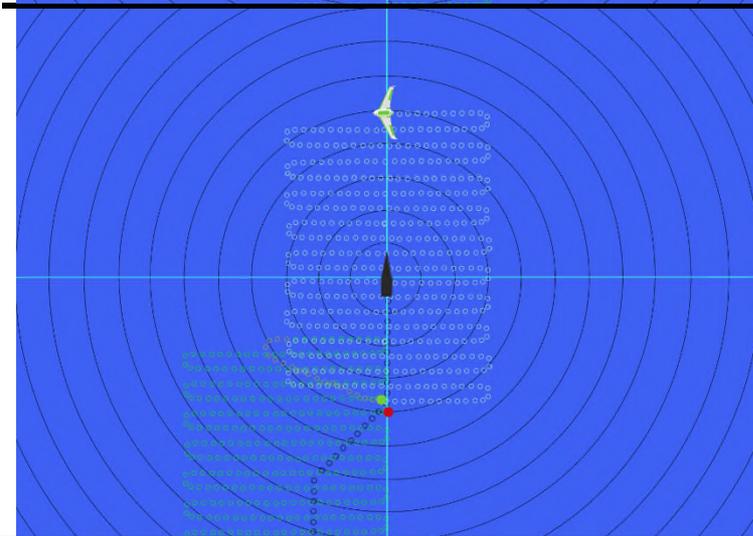
2. When a school is detected from the drone, a sighting (green dot) is logged in the software used by the drone observers. No information about this sighting is shared with the flying bridge observers. It may be decided to make multiple passes over the school to obtain imagery of the whole school. This is particularly important for calibration schools.



3. If it is decided that the drone does not track its sighting, the drone continues the search ahead of the ship until a sighting is made by the flying bridge observers. The drone observers then need to evaluate whether the drone and flying bridge sightings are the same school using all information available.



4. When the ship starts closing on the school that the flying bridge observers detected, the drone flies towards the same school.



5. While the ship is with the school, the drone restarts search effort across a projected transect line that the ship will resume on after finishing with the school (see Diagram 1 above). When the ship finishes with the school, it will continue search effort along the same projected transect line.

References

Barlow J (2015) Inferring trackline detection probabilities, $g(0)$, for cetaceans from apparent densities in different survey conditions. *Marine Mammal Sciences*. 31(3):923-943.

- Borchers D (2012) A non-technical overview of spatially explicit capture-recapture models. *Journal of Ornithology*.152, (2), 435-444.
- Gerrodette T, G Watters, W Perryman & L Balance (2008) Estimates of 2006 dolphin abundance in the eastern tropical Pacific with revised estimates from 1986-2003. NOAA Technical Memorandum NOAA-TM-NMFS-SWFSC-422.
- Gerrodette T, WL Perryman & CS Oedekoven (2018) Accuracy and precision of dolphin group size estimates. *Marine Mammal Science* 35(1): 22-39.
- Kinzey D, P Olson & T Gerrodette (2000) Marine mammal data collection procedures on research ship line-transect surveys by the Southwest Fisheries Science Center. Southwest Fisheries Science Center, Administrative Report LJ-00-08, 32 p.
- Oedekoven CS, ST Buckland, L Marshall & CE Lennert-Cody (2018). Agreement on the international dolphin conservation program. 37TH meeting of the parties. San Diego, California (USA). DOCUMENT MOP-37-02. Dated 17 August 2018. 46 pages available at http://www.iattc.org/Meetings/Meetings2018/AIDCP-37/PDFs/Docs/_English/MOP-37-02_Design%20of%20a%20survey%20for%20eastern%20tropical%20Pacific%20dolphin%20stocks.pdf

GEO DRONES

SeaHawk S

General Specs

The Carbon-Based SeaHawk

Professional unmanned aerial vehicle (UAV); exclusive of Gtt NetCorp for Mexico and Latin America (LATAM); part of the **UAVER** family of aerial robots.

DO TASKS NOT POSSIBLE BEFORE. SAFELY. FASTER. ECONOMICALLY.

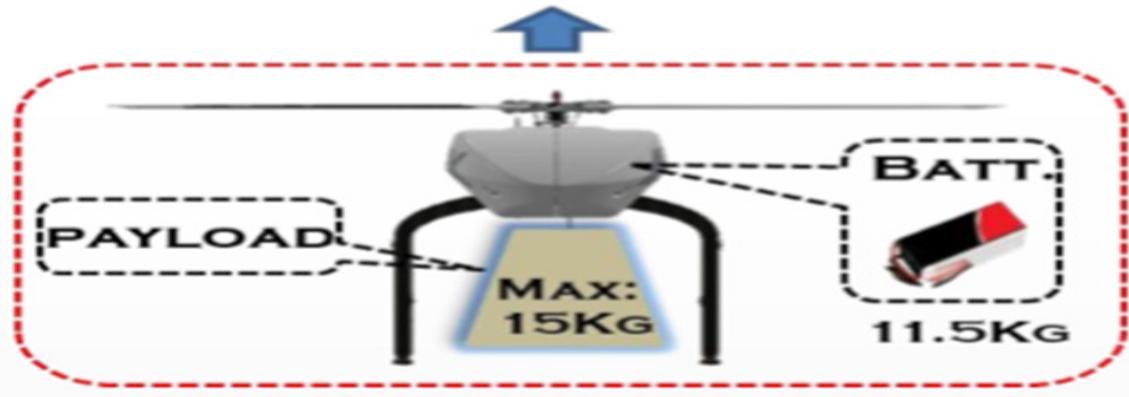
SeaHawk S, professional multi-function unmanned aerial robot (helicopter), specially designed for the long-distance real-time **monitoring** reconnaissance and surveillance application with the following specifications:

1. Made of carbon fiber, and metal
2. Large payload space
3. High mobility
4. Low noise
5. Low vibration
6. Long endurance
7. HD video camera 1080p, 10x zoom with real time video transmission
8. HD frame camera 20 mega pixels for imagery post processing and analytics
9. Onboard radio/telecom
10. Reception stations on ship or land based to receive video with onsite TV monitor and computer. Optional GeoDrones machine learning analytics.

Max. Take-off Weight	30 kg
Dry Weight	9 kg
Battery Weight	11.5 kg
Max. Payload Weight	15 kg
Wind Resistance	Beaufort 8
Max. Cruise Speed	80 km/hr
Max. Service Ceiling	3000 m
Endurance	> 60 min
Data Transmission Distance	> 50 km



MAX. TAKE-OFF WEIGHT : 30KG



ETP trial survey: Project report

Cornelia S. Oedekoven¹, Stephen T. Buckland¹, Laura Marshall¹, Cleridy E. Lennert-Cody²,

Lewis McMillan³, Sakina Fatima³, David Harris-Birtill³, Claudia Faustino¹

¹Centre for Research into Ecological and Environmental Modelling / University of St Andrews

²Inter-American Tropical Tuna Commission, 8901 La Jolla Shores Drive, La Jolla CA 92037-1509, USA

³Computer Science / University of St Andrews

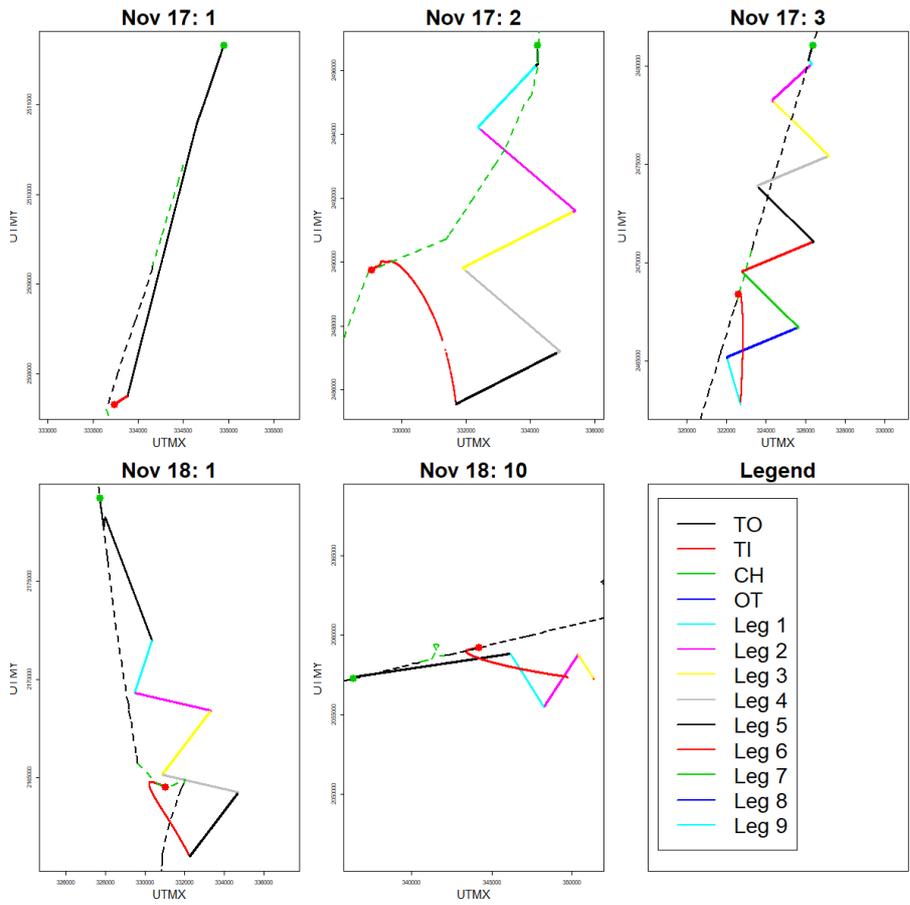
Appendix 3: Flight paths of drones

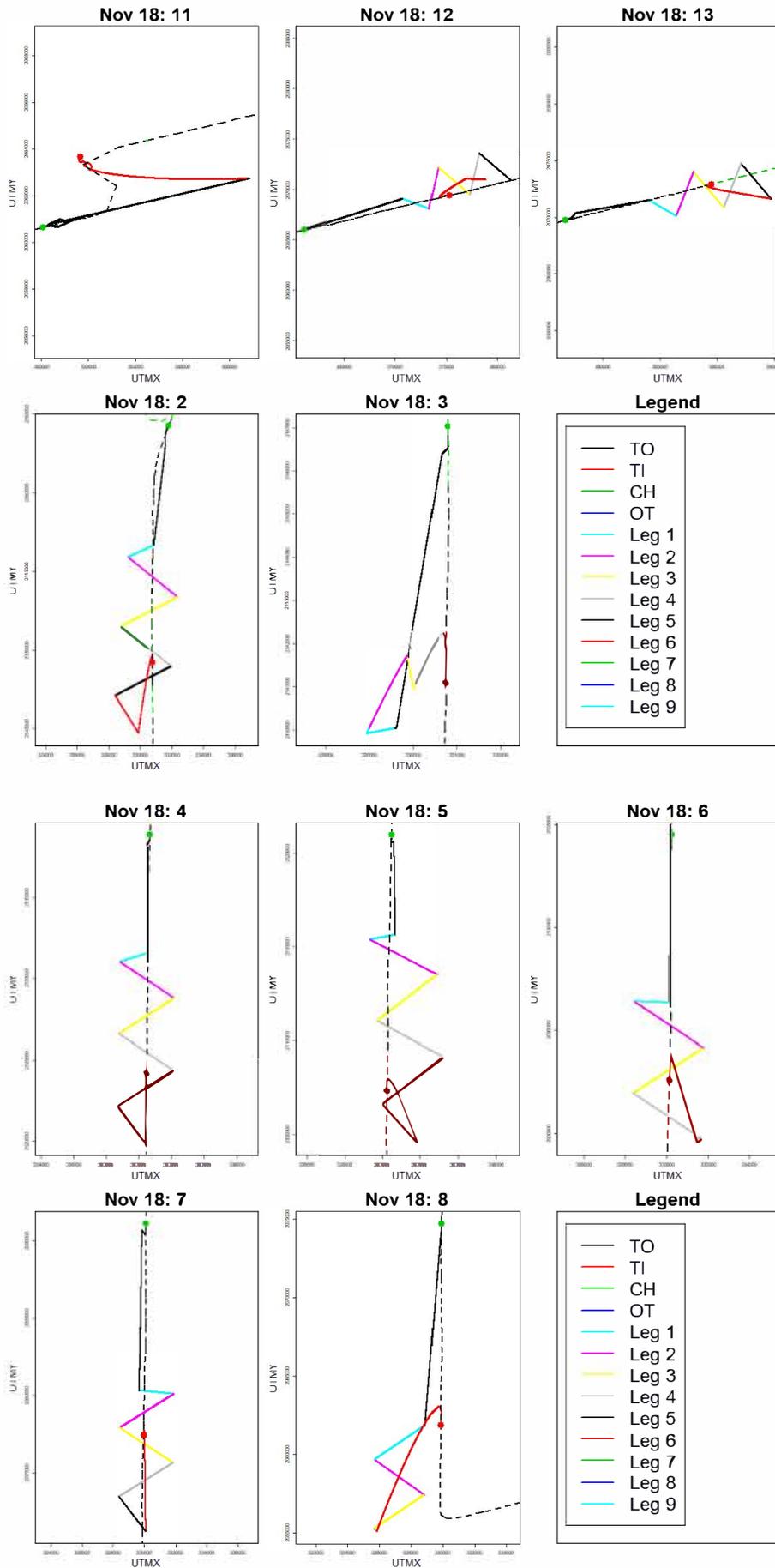
This appendix contains maps of the paths of all drone flights conducted during the 14 day trial survey. The legends in the Figures provide the key to the colour-coded effort type of the drone flight sections. Table 1 provides the type of effort the abbreviations from the legends refer to. Dashed lines represent the track completed by the ship where black and red lines refer to on-effort in closing and passing mode, respectively and green refers to off-effort. Green and red dots indicate the launch and landing locations of the drone, respectively.

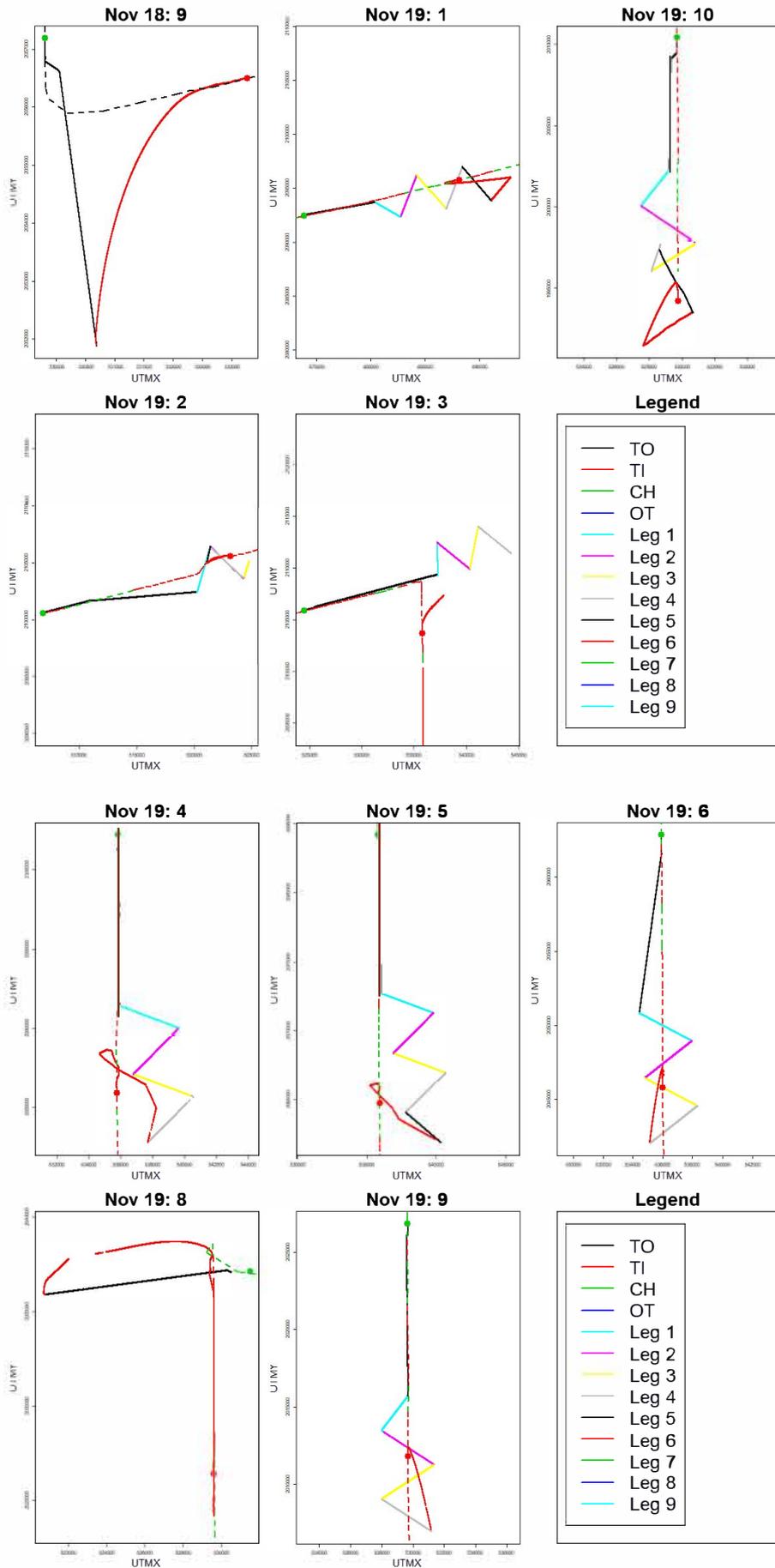
Table 1. Effort types during drone flights and their abbreviations used in the Figures below.

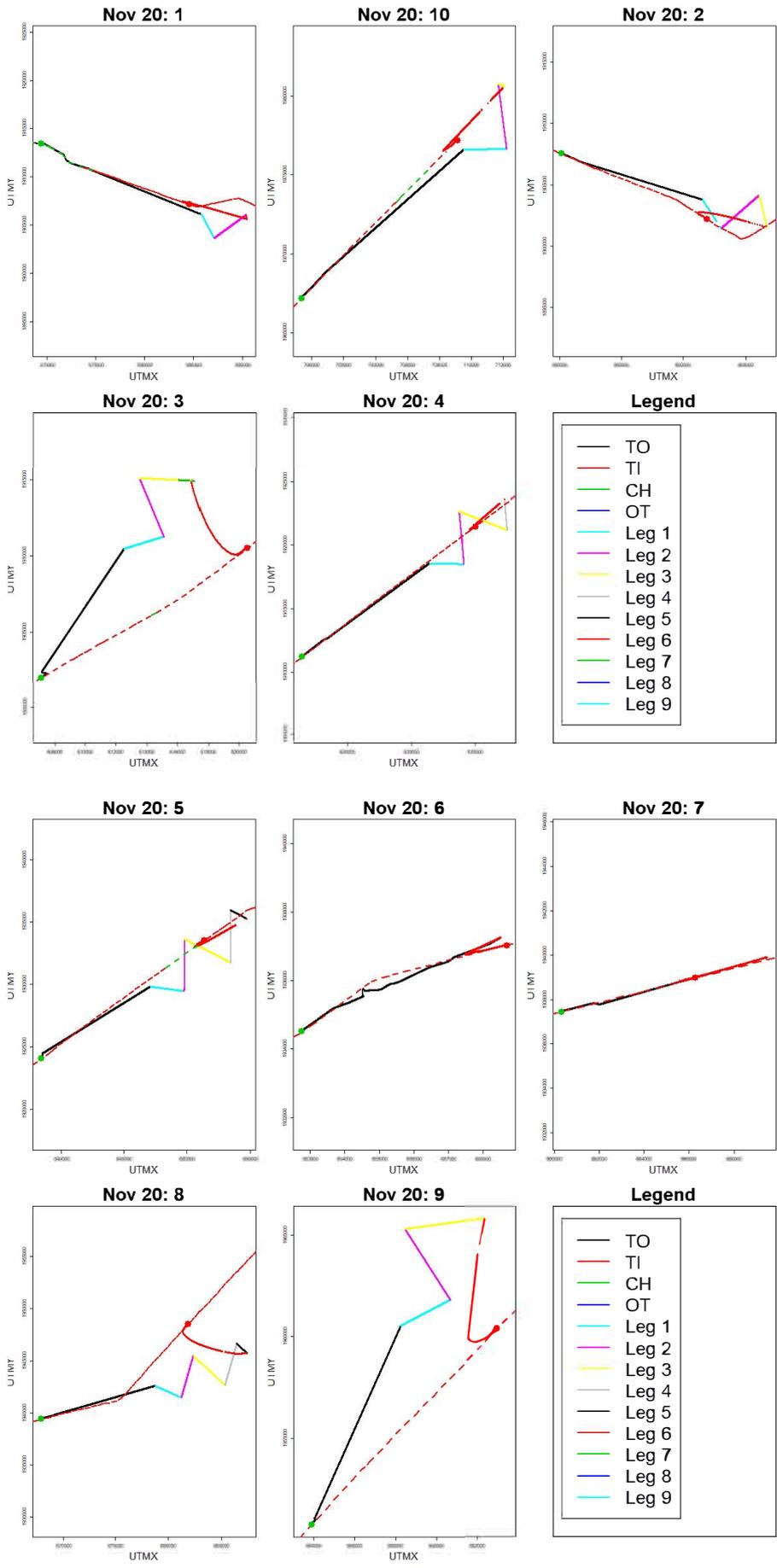
Zigzag effort types	
TO	Transit outbound
Leg	On effort
TI	Transit inbound
CH	Checking
Calibration effort types	
TOC	Transit outbound
WS	With school
TIC	Transit inbound
Other	
OT	Other

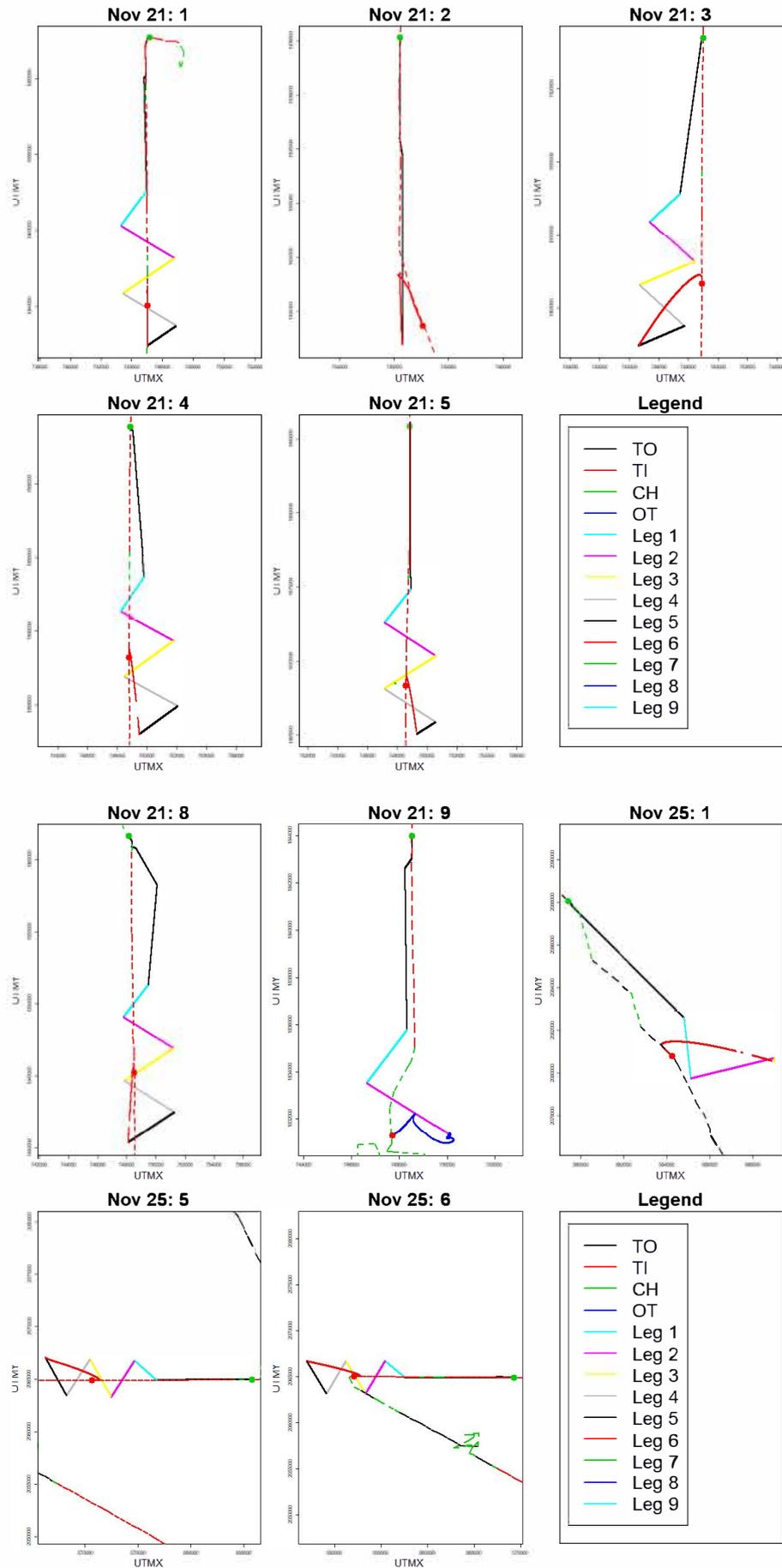
Pure zigzag flights

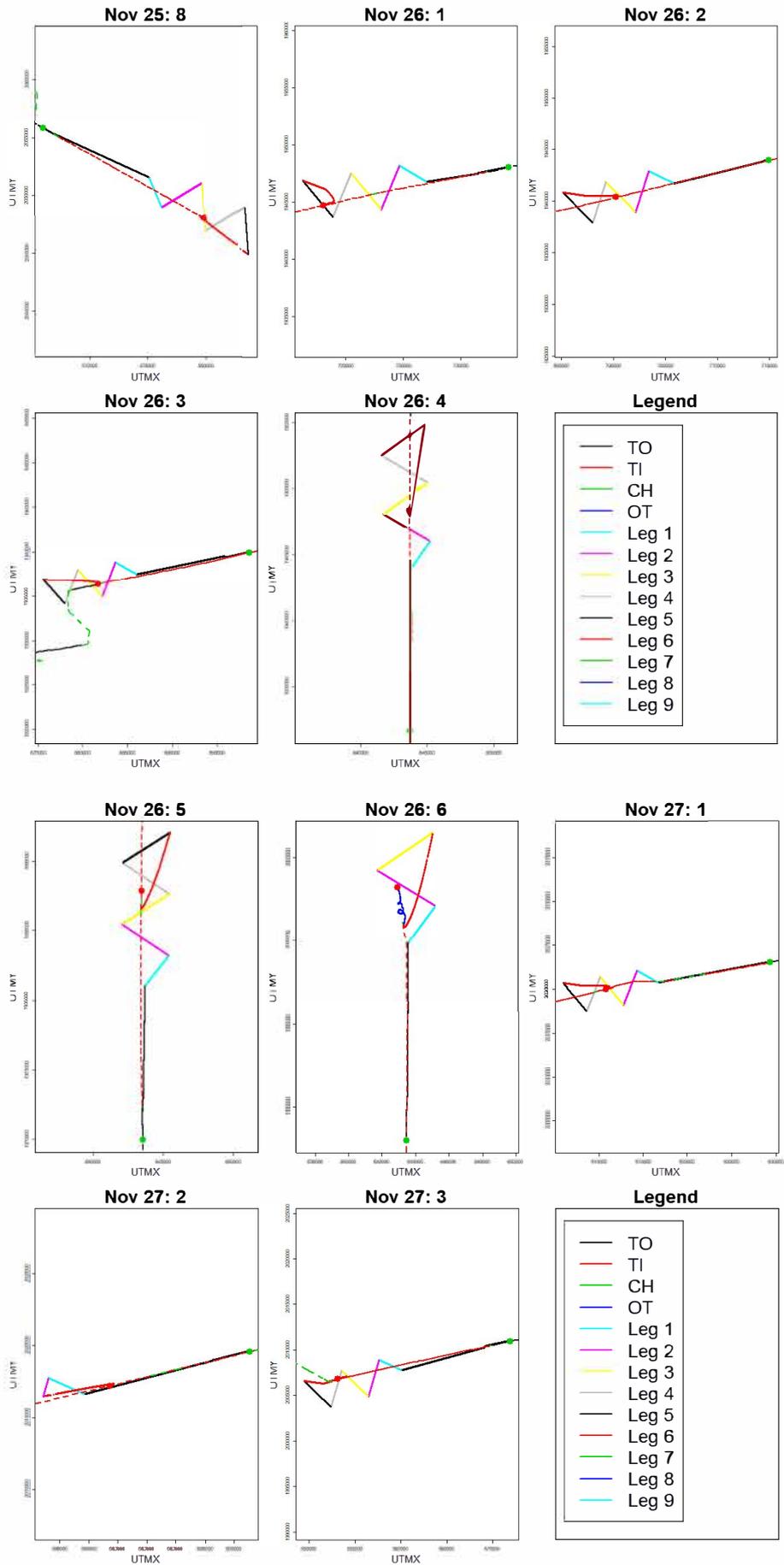


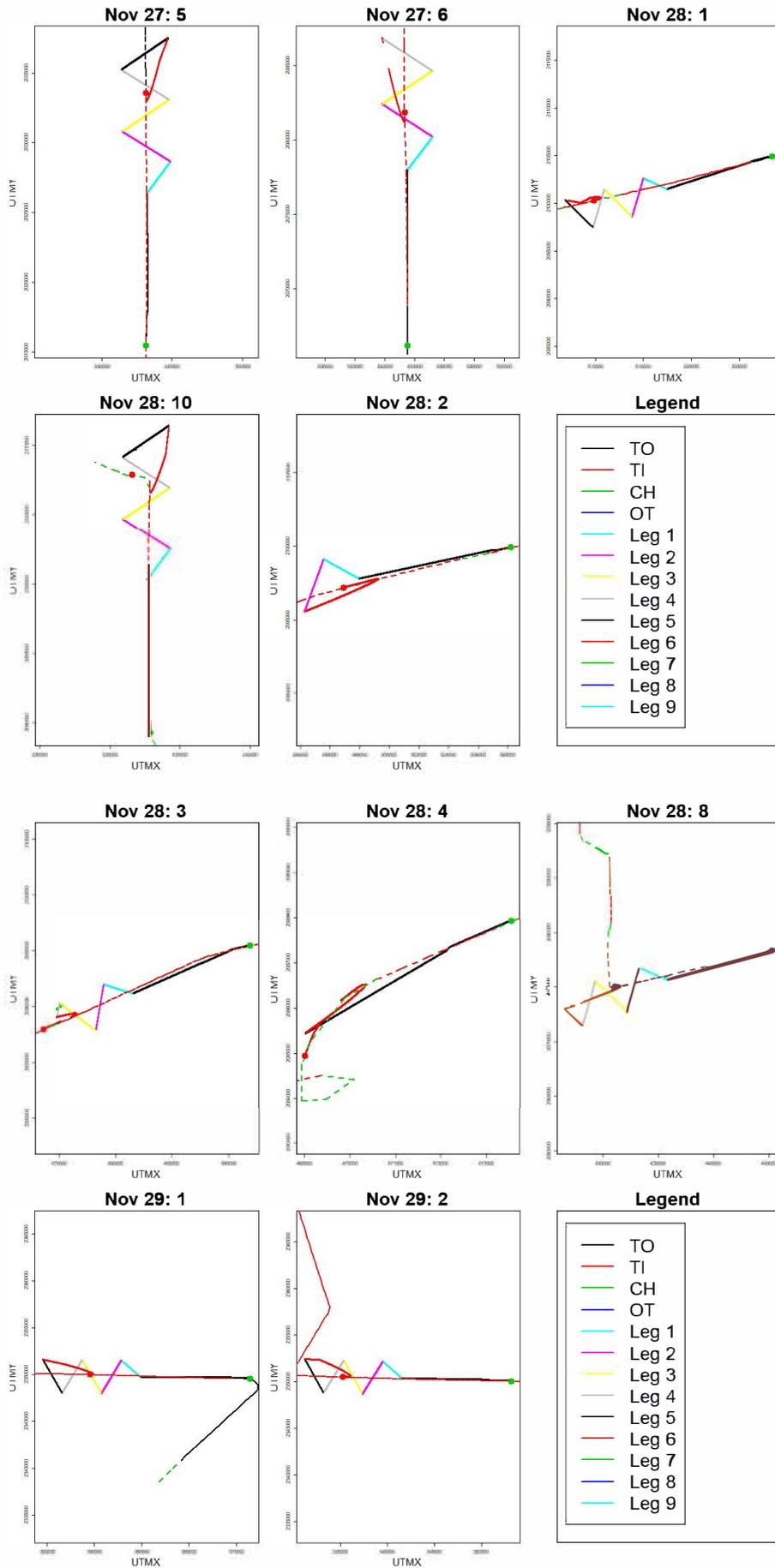


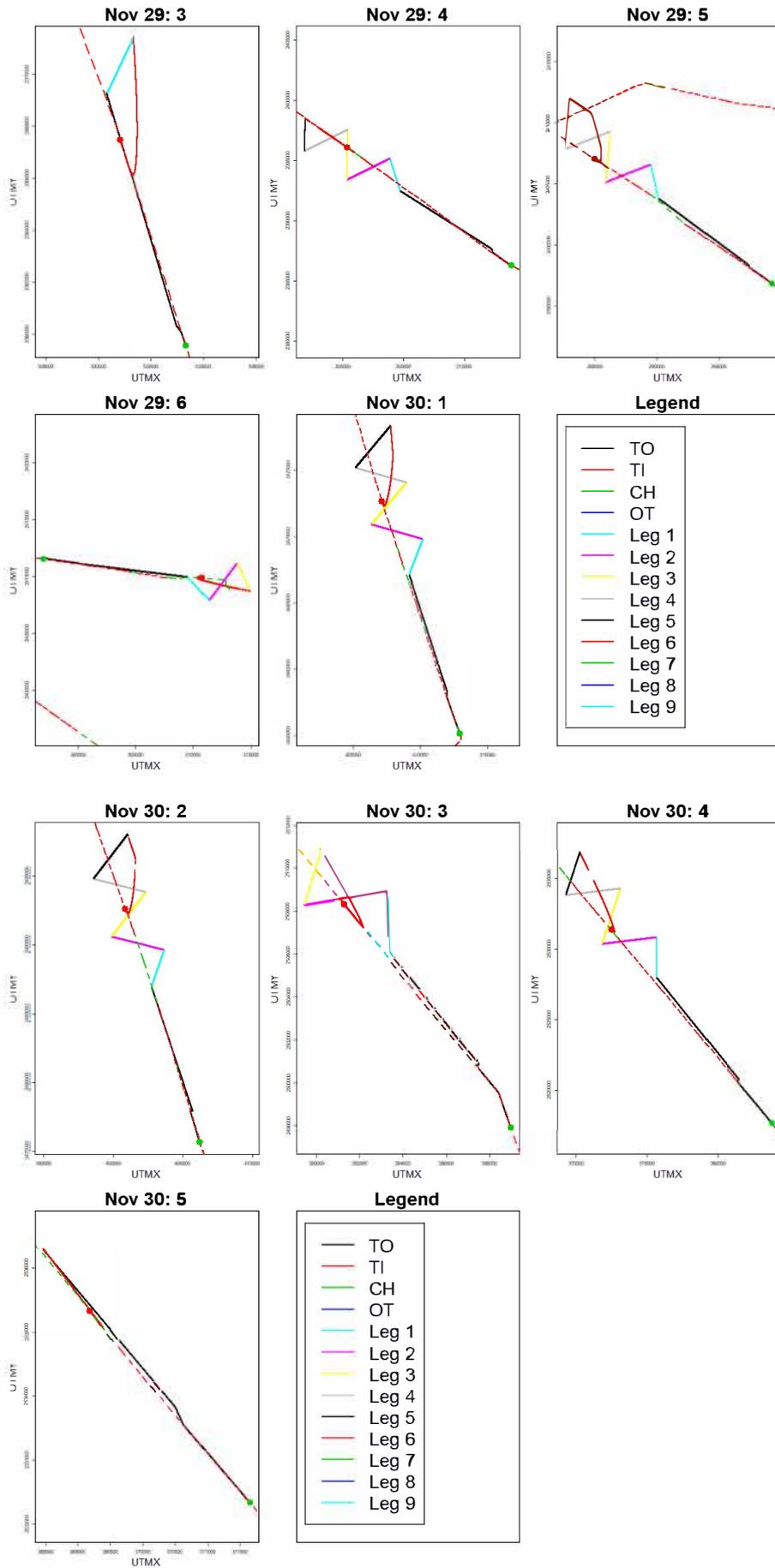




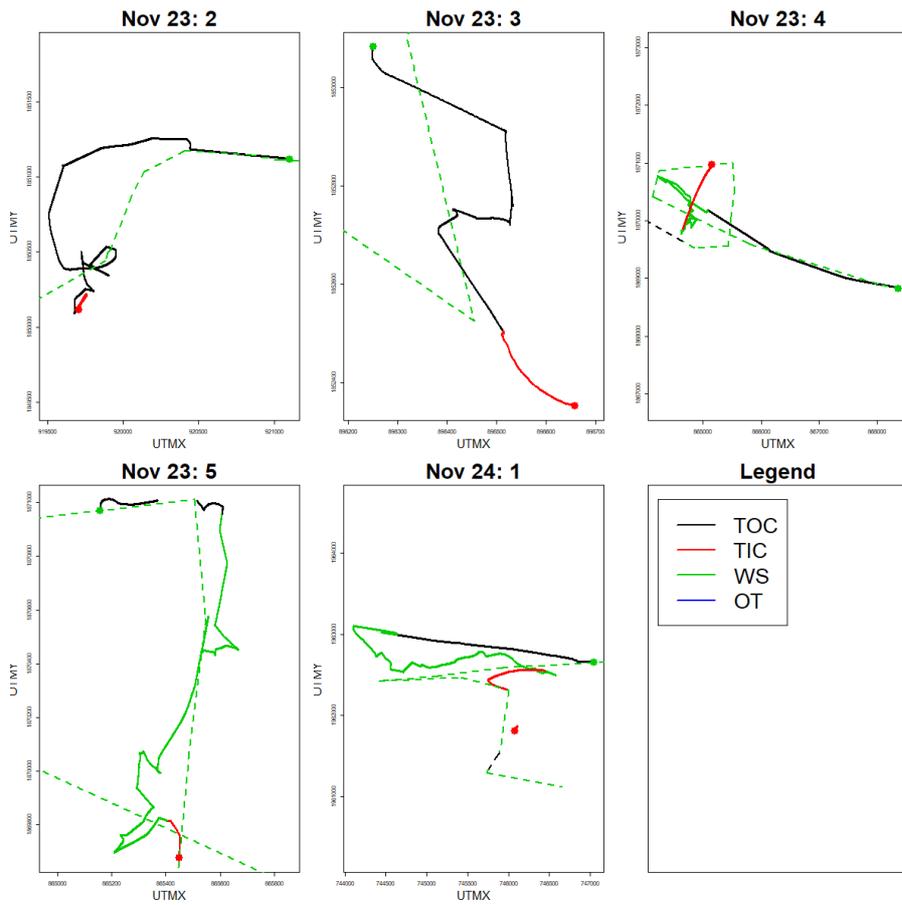


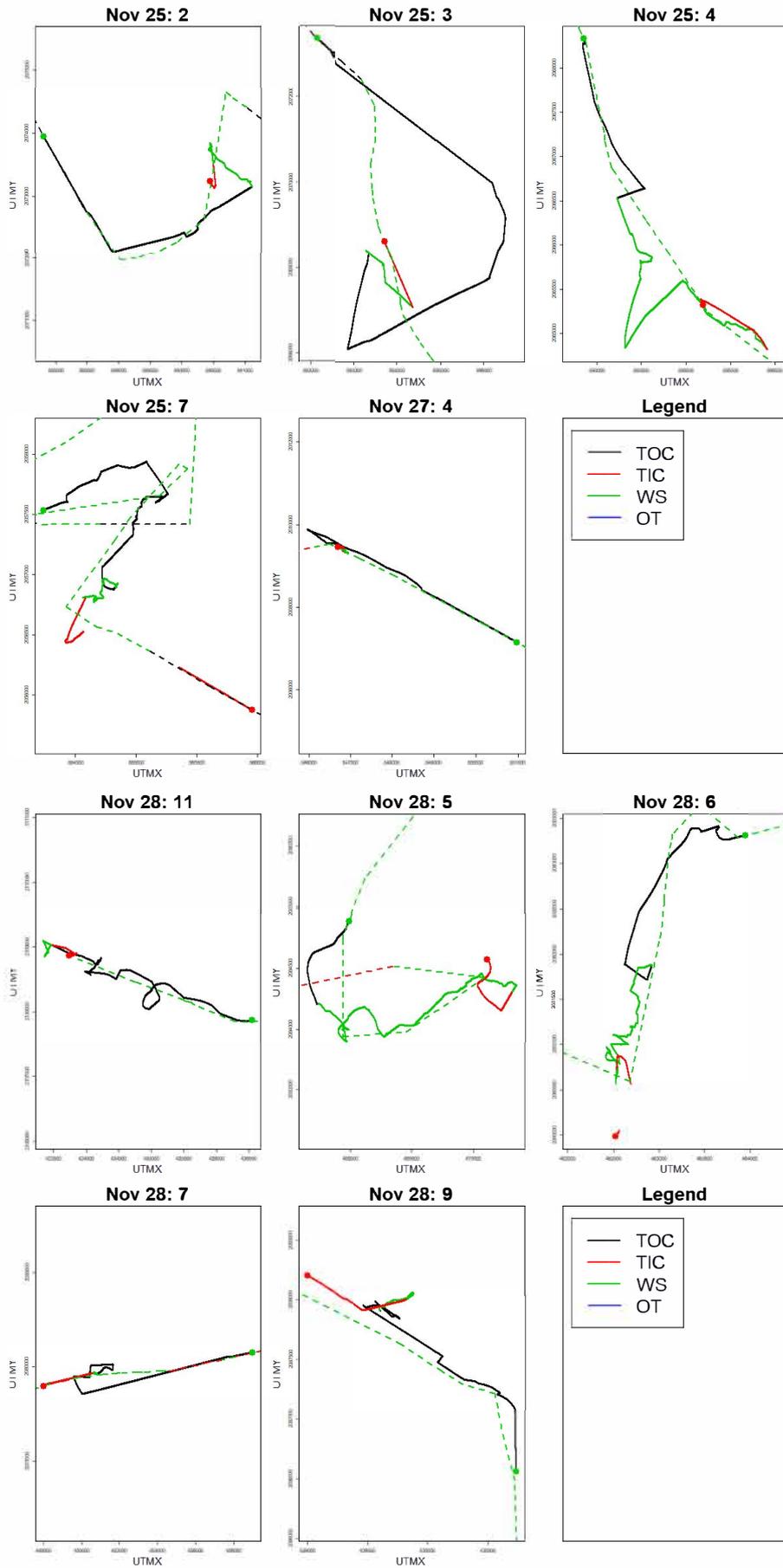






Pure Calibration flights





Mixed zigzag-calibration flights and other

