



A simple tool for linking photo-identification with multimedia data to track mammal behaviour

Alexandre M. S. Machado^{1,2} · Mauricio Cantor^{1,3,4,5,6}

Received: 5 February 2021 / Accepted: 6 October 2021 / Published online: 27 December 2021
© The Author(s) 2021, corrected publication 2022

Abstract

Identifying individual animals is critical to describe demographic and behavioural patterns, and to investigate the ecological and evolutionary underpinnings of these patterns. The traditional non-invasive method of individual identification in mammals—comparison of photographed natural marks—has been improved by coupling other sampling methods, such as recording overhead video, audio and other multimedia data. However, aligning, linking and syncing these multimedia data streams are persistent challenges. Here, we provide computational tools to streamline the integration of multiple techniques to identify individual free-ranging mammals when tracking their behaviour in the wild. We developed an open-source R package for organizing multimedia data and for simplifying their processing a posteriori—“*MAMMals*: Managing Animal MultiMedia: Align, Link, Sync”. The package contains functions to (i) align and link the individual data from photographs to videos, audio recordings and other text data sources (e.g. GPS locations) from which metadata can be accessed; and (ii) synchronize and extract the useful multimedia (e.g. videos with audios) containing photo-identified individuals. To illustrate how these tools can facilitate linking photo-identification and video behavioural sampling in situ, we simultaneously collected photos and videos of bottlenose dolphins using off-the-shelf cameras and drones, then merged these data to track the foraging behaviour of individuals and groups. We hope our simple tools encourage future work that extend and generalize the links between multiple sampling platforms of free-ranging mammals, thereby improving the raw material needed for generating new insights in mammalian population and behavioural ecology.

Keywords Behavioural sampling · Hydrophone · Individual · Photo-identification · Unmanned aerial vehicle

Introduction

Natural populations change in size and composition, propelling the dynamics of ecological communities, species interactions, and energy flow through the ecosystem (Odum and Barrett 1971). At the heart of these changes, are individual animals being born, growing, behaving, and dying. Individual-based data provide the raw material to investigate the mechanics and dynamics of these natural populations, their ecological and behavioural interactions and evolution

Handling editors: Leszek Karczmarski and Scott Y.S. Chui.

This article is a contribution to the special issue on “Individual Identification and Photographic Techniques in Mammalian Ecological and Behavioural Research – Part 1: Methods and Concepts” — Editors: Leszek Karczmarski, Stephen C.Y. Chan, Daniel I. Rubenstein, Scott Y.S. Chui and Elissa Z. Cameron

✉ Mauricio Cantor
mauricio.cantor@oregonstate.edu

¹ Departamento de Ecologia e Zoologia, Universidade Federal de Santa Catarina, Florianópolis, Brazil

² Department of Collective Behaviour, Max Planck Institute of Animal Behavior, Radolfzell, Germany

³ Department for the Ecology of Animal Societies, Max Planck Institute of Animal Behavior, Konstanz, Germany

⁴ Department of Fisheries, Wildlife and Conservation Sciences, Marine Mammal Institute, Oregon State University, Newport, USA

⁵ School of Plant, Animal and Earth Sciences, University of Witwatersrand, Johannesburg, South Africa

⁶ Centro de Estudos do Mar, Universidade Federal do Paraná, Pontal do Paraná, Brazil

(Coulson 2020), which is particularly necessary in longitudinal studies (Clutton-Brock and Sheldon 2010). Therefore, a deep understanding of these patterns and processes in animal ecology requires identifying and tracking individual animals over time and space (Coulson 2020).

The available invasive and non-invasive methods for sampling individual animals present trade-offs on the accuracy, content and quality of the data they provide. Invasive methods require capturing animals to mark (e.g. with collars, tattoos, tags, freeze branding; Silvy et al. 2005) or fit tracking devices (RFID, GPS, acoustic, satellite tags: e.g. Krause et al. 2013) but provide detailed information about the individuals (e.g. identity, location, behaviour, health). Actively capturing and marking animals, however, can be unfeasible, expensive or disrupt natural behaviour or physiology (Walker et al. 2012). By contrast, non-invasive identification methods, such as photographic, acoustic and video recording (Karczmarski et al. 2022a, b), rely on systematic comparison of natural marks or behaviours (e.g. Karanth and Nichols 1998; Muller et al. 2018; Longden et al. 2020) to track individuals from a distance (e.g. Clapham et al. 2020; Ferreira et al. 2020). Although efficient in providing individual identities, non-invasive methods generally provide fewer information on other biological variables (but see Toms et al. 2020), which has motivated the simultaneous use of other multimedia sampling platforms, such as video (e.g. Raoult et al. 2018; Francisco et al. 2020; Landeo-Yauri et al. 2020) and audio recordings (Cheng et al. 2012; Erbe et al. 2020). Novel technologies for identifying and tracking individuals using such multimedia data are becoming increasingly more precise in the lab or captivity (e.g. Mersch et al. 2013; Dell et al. 2014; Pérez-Escudero et al. 2014; Alarcón-Nieto et al. 2018; Graving et al. 2019; Marks et al. 2021), but doing so in situ remains more challenging (e.g. Ferreira et al. 2020; Guo et al. 2020). In the field, where animals are not spatially constrained, recording data from multiple sampling platforms simultaneously, or syncing large volumes of data to then link with that of individual identification a posteriori, can be troublesome.

In wild mammal research, cetacean studies exemplify the continuous development of non-invasive individual identification methods based on multimedia data. Photo-identification has been the go-to technique to recognize individual whales and dolphins in the last five decades (e.g. Würsig and Würsig 1977; Katona and Whitehead 1981; Hammond et al. 1990). Since whales and dolphins can range over large areas and spend long times underwater, photo-identification has been increasingly coupled to other multimedia sampling to detect the presence of individuals and/or describe their behavioural patterns. For instance, while cameras and acoustic sampling provide invaluable underwater perspectives, the growing market of unmanned aerial vehicles (drones) has popularized the recording of behaviour, movement and

health of cetaceans from an overhead view (e.g. Torres et al. 2018; Gray et al. 2019; Hartman et al. 2020). With few exceptions, however, these sampling techniques do not provide individual identities—but see, e.g., identification from overhead images (e.g. Payne et al. 1983; Durban et al. 2015) or acoustic signals (e.g. Janik and Sayigh 2013). Combining traditional photo-identification sampling with hydrophones, underwater and drone cameras can resolve this limitation, but it inevitably creates another one: individual behavioural tracking from multiple platforms generates a large and multi-dimensional dataset that rapidly become unfeasible to handle manually. These technological advances have therefore produced a need for corresponding advances in computational tools to organize and process multiple data streams (e.g. Schneider et al. 2019).

Here, we introduce a free and open computational tool for aligning, linking and syncing photo-identification data with other multimedia data of free-ranging vertebrates. The R package *MAMMals*—*Managing Animal MultiMedia: Align, Link, Sync*—contains functions to synchronize different multimedia data streams a posteriori and so facilitate their post-processing to measure relevant biological and behavioural data. Using *MAMMals*, one can (i) extract, organize and line up the metadata of photographs, videos, audios, drone flight logs and any other timestamped text data; (ii) select, trim and export clips or stills of the footage or audio recording containing individual photo-identification; and (iii) wrangle, convert and plot data from cameras, drones, hydrophones, microphones and other timestamped data sources. In what follows, we describe the workflow for pre-processing individual photo-identification and link it to other multimedia data (Fig. 1). Next, we illustrate the utility of these tools by applying them to process and analyse empirical data on the foraging behaviour of coastal bottlenose dolphins. We conclude by discussing the caveats of our approach and how future work can alleviate them.

Workflow overview: coupling photo-identification with other multimedia data

The *MAMMals* R package targets the challenge of coupling large volumes of observational and multimedia data to traditional techniques of identifying individuals, extending therefore the possibilities for studies that use methods of focal-animal and focal-group sampling (Altmann 1974). The minimum requirements are image files with assigned individual identification and at least one other multimedia data source. The workflow follows four steps (Fig. 1): (i) extracting the metadata of photographs and any other multimedia files available; (ii) aligning the metadata of these files to select the useful multimedia containing photo-identified

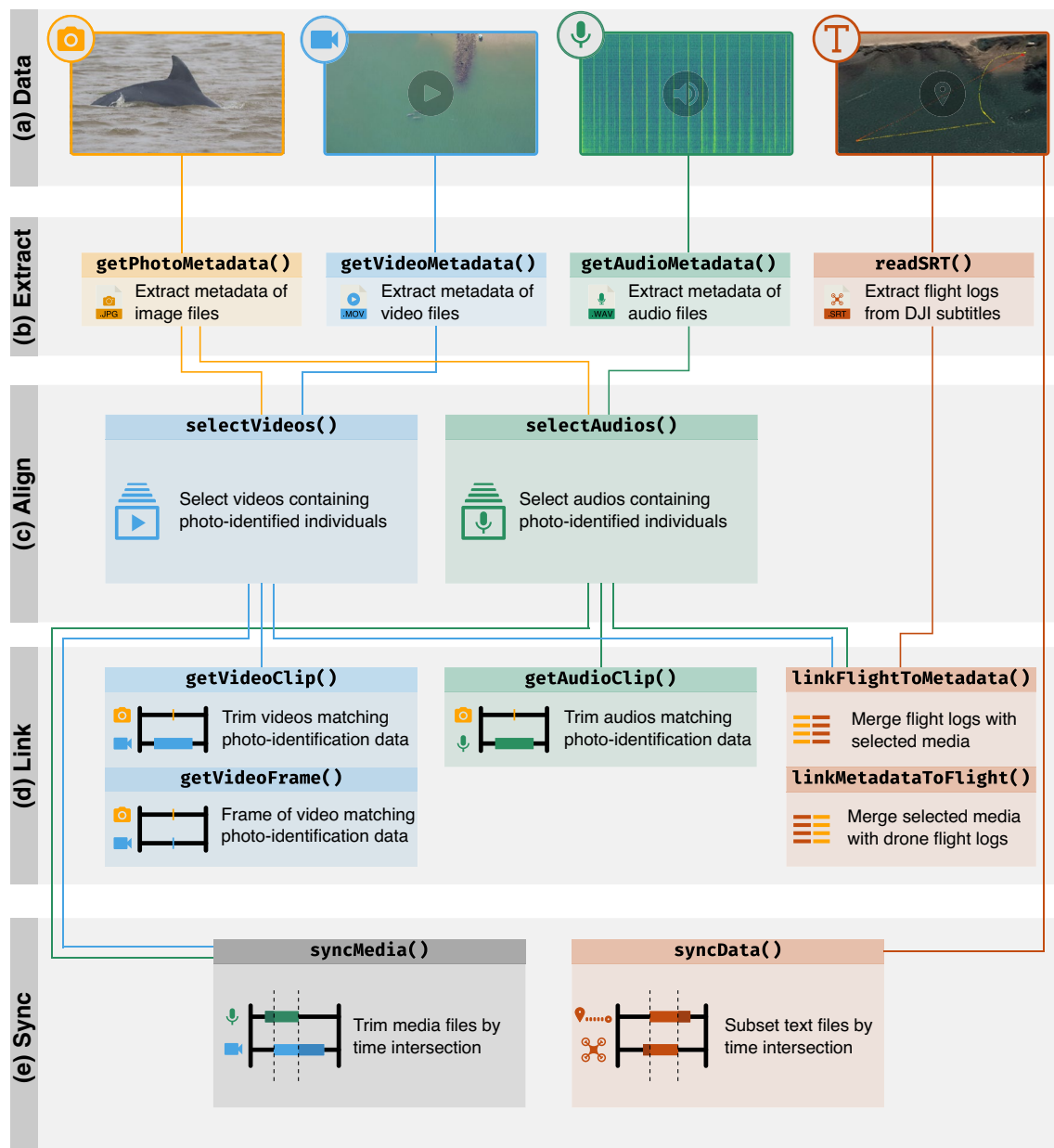


Fig. 1 The *MAMMals* workflow to align, link and sync multimedia and timestamped text data. **a** The inputs are files commonly produced in individual identification and behavioural sampling methods, such as images (.jpg, .tiff, .png), videos (.mov, .mp4), audios (.wav, .mp3) and/or text files (.csv, .txt, .srt). After aligning, linking and syncing the inputs, the outputs can be text files with metadata and/or synced image, audio and/or video files. The minimum requirements for the *MAMMals* workflow are the photo-identification data (i.e. the image files associated to individual identification text data), and at least one more multimedia data source, such as videos, audios or text files. **b** The first step is to extract the metadata of all multimedia files (and flight logs, if available), or from captions in .srt files of commercially available drones). One can also export the metadata for posterior processing, such as attributing individual IDs to each photo processed by the *getPhotoMetadata* function, or assign individual identifica-

tion from pre-processed data directly to the function *getPhotoMetadata*. **c** The second step is to align the metadata of photographs (or timestamped field notes) with that of the other media to automatically select the video or audio files containing individual photo-identification data. **d** The third step is to link the selected media by clipping the videos and audios around the information of interest (e.g. photo-identified individuals) to facilitate the post-processing of videos (*getVideoClip*), audios (*getAudioClip*) or stills from the video (*getVideoFrame*). If sampling includes drone videos, selected media can be linked to information from the flight, such as latitude, longitude and altitude. **e** The final step is to sync media and/or text by subsetting only the time intersect between data coming from different sampling platforms. The synced multimedia and text data can be exported as a single merged file or multiple separate files

individuals; (iii) linking these selected files by clipping the multimedia containing photo-identified individuals; (iv) and syncing media and text data around their intersection time. We detail each step of the *MAMMals* workflow in the next sessions, and provide instructions and examples of the input and output files in a documentation in an online tutorial (<https://mammals-rpackage.netlify.app/index.html>). The *MAMMals* R package can be installed from the online repository (instructions at <https://bitbucket.org/maucantor/mammals/>). It depends on the installation of the R environment (R Core Team 2021) and key R packages such as *lubridate* (Grolemund and Wickham 2011) to manage date-time formats (full list of dependencies, see the package repository), as well as external software ExifTool (<https://exiftool.org>) to extract the metadata of media files, and FFmpeg (<https://ffmpeg.org>) to clip video and audio files.

To align, link and sync multiple data sources, the *MAMMals* workflow relies on timestamped files: essentially, the recording times of the multiple sampling equipment are extracted from the metadata of the media files and lined up. Therefore, the most important recommendations for field sampling are to synchronise the clocks of all collection platforms, and to keep the original metadata of the media files unaltered. For accurate results, we recommend the clocks of cameras, drones, audio recorders and auxiliary equipment (such as cell phones used to pilot the drone or tablet apps to record observation data) to be adjusted to the maximum precision possible via information—either from the GPS satellite, or manually—and to be always double-checked and fine-tuned before each sampling occasion to account for clock drift. For example, one can photograph and film a reference clock prior to sampling or use audio or visual signals during sampling (e.g. flash in our case study detailed below) to offset time differences across images and videos.

When photographing animals for individual identification using natural marks, we recommend following the protocols for collecting, processing and organizing such data, which have been extensively detailed elsewhere (e.g. Speed et al. 2007; Urián et al. 2015). We highlight that using DSLR cameras equipped with GPS and digital compass can be useful when teasing apart photo-identified individuals in the field, especially when tracking them with overhead videos. For instance, when tracking multiple individuals or groups distributed in space, one can assign the photographs taken to each group recorded in the overhead footage by interpreting the GPS coordinates and shooting angle extracted from the photograph metadata. After the photographic data sampling, we recommend first processing the photo-identification data and organizing it in a plain text data frame, in which the first column contains the photograph file name and extension (e.g. ‘6Q1A8164.JPG’), and the second contains the individual (alphanumeric) identification code (e.g. ‘ID1248’).

When recording audio, we recommend using recorders that can produce timestamped files. Otherwise, one can manually check the end time of recordings after sampling and rename files accordingly with date and time. When recording videos from small drones (e.g. DJI Phantom, DJI Mavic Pro, DJI Inspire, Splash Drone) while simultaneously collecting photo-identification or audio recordings, we recommend keeping a constant flight height and point the camera straight down (i.e., drone and camera pitch = -90° , roll = 0°) to ensure the centre of the frame matches the coordinates recorded by the drone GPS and to reduce the distortion from any measures taken from the drone footage. If measuring animals from the drone footage using photogrammetry, there will be additional requirements. In addition to the camera tilt, the aircraft altitude data are the main issues for precise and unbiased photogrammetric measurements. Off-the-shelf drones record the altitude relative to the aircraft’s take-off position (“home point”). Hence, if the aircraft takes off from the deck of a ship or a higher ground, the zero in the aircraft’s barometer does not match the sea level. To mitigate this, an object of known length can be used to calibrate a scale (details in Burnett et al. 2019). Another solution is to couple a LiDAR sensor to the drone (e.g. Dawson et al. 2017) to precisely measure the distance from the aircraft to the sea level. Correcting lens distortion and camera calibration also reduce errors in measurement estimates (see Dawson et al. 2017).

Step 1: extracting metadata of multimedia files

After conducting photo-identification as per standard protocols, the first step in the *MAMMals* workflow is to extract the metadata of all multimedia files (Fig. 1b) and organize them into a text database, such as an R data frame. We suggest allocating each media type in separate subfolders within the root folder of the project, then using the following functions to read and organize the metadata into a data frame where the number of rows equals the number of files, and each column corresponds to the available metadata. To extract the metadata of the photographs, access the subfolder with the image files with the function *getPhotoMetadata*, which handles many common image extensions (e.g. .jpg, .tiff, .png) and accesses the available metadata of each photograph—at least the date and time, but also the camera GPS coordinates and shooting angle, if available. The *getPhotoMetadata* function also assigns the individual ID to the full metadata of the photographs, by matching the file names with that of the simple two-column data frame containing the photo file name and the individual identification code. While we recommend having the individual identification ready prior linking and syncing with the other multimedia files, we highlight that, alternatively, one can also perform the photo-identification afterwards. In this case, the *getPhotoMetadata* function can be used to export the metadata of photographs to common text files

(e.g. .csv or .txt) and to then assign individual ID to the database using any text editor or spreadsheet software (e.g. Microsoft Excel, Apple Numbers). Bear in mind, however, that issues with the date and time formats and precision are common when using spreadsheet software; thus, we suggest using plain text editors to avoid lack of precision when aligning, linking or syncing the photo-identification to the multimedia data.

For the audio subfolder, use *getAudioMetadata* to extract metadata of audio files (at least duration, initial and final time). If the audio files do not contain date and time in the metadata, initial and final time of recordings can be extracted from the filename automatically generated with date-time stamps, as exported by commonly used autonomous recorders (e.g. Whytock and Christie 2017; Hill et al. 2019). To extract the metadata of the videos (at least duration, initial and final time), access the video subfolder with the *getVideoMetadata*. If videos were recorded with drones, additional metadata can be available (e.g. altitude, GPS coordinates) and will be extracted and organized into a text database as well. Most commercially available drones save detailed logs of every flight. Information on aircraft sensors, motors, battery, remote controller and media are logged on-board and on remote applications, often using proprietary file structures. Hobbyists (e.g. DatCon, TXTlogtoCSVtool), companies (e.g. <https://airdata.com>) and forensics (e.g. Clark et al. 2017) have been developing tools to decode flight logs into readable .csv files. Alternatively, the *MAMMals* R package can extract the basic flight log data recorded by DJI drones. These drones can produce timestamped subtitles (1 Hz data) logging the aircraft latitude, longitude and height (calculated from the aircraft barometer), the home point latitude and longitude, and camera settings. However, subtitles do not contain auxiliary information on the aircraft and camera roll, pitch and angle; and the accuracy of latitude and longitude is limited to 10 m. But conveniently, subtitles are natively exported from DJI drones as text files (.srt) along video files, and the *MAMMals readSRT* function can read all .srt files in a folder and return an R data frame with the formatted metadata of the DJI drone flight logs.

Step 2: aligning multimedia files

After extracting the metadata of the multimedia files, large volumes of multimedia data can be aligned with the *MAMMals* functions that subset media files containing photo-identification data (Fig. 1c). Use the *selectVideos* or *selectAudios* functions to get the video and audio files of interest, respectively, by aligning their metadata with the metadata of the photographs of individuals (previously generated by the functions *getVideoMetadata*, *getAudioMetadata*, *getPhotoMetadata*, respectively). The *select* set of functions calculates the time of the photograph in the video or audio files for all photographs taken during the sampling event,

and they return an R data frame with data matching the time in the video or audio files. Then, one can export an R data frame containing only the photo-identified individuals, or other events of interest, into a .csv or .txt. We highlight that while these functions are based on photograph metadata, they also work with other text data in which events are correctly timestamped (Fig. 1c), such as observed behavioural events recorded in the field notes, and GPS positions from loggers fitted to the animals.

Step 3: linking photographs with multimedia data

After aligning the metadata of the media files, the photo-identification data can be linked with video or audio files by trimming these media files (Fig. 1d) based on the information generated by the *selectVideos* and *selectAudios* functions. If the aim is to get a still from the video for every photo-identified individual, the *getVideoFrame* function can export a frame of the video in the moment each photo was taken. If the aim is to perform further video or audio analyses, one can export short clips around the time of each photo-identification for both video (*getVideoClip*) and audio files (*getAudioClip*). If sampling with drones, one can automatically link data from flight logs to every event exported by the *selectVideos* or *selectAudios* functions. The *linkFlightToMetadata* function returns an R data frame in which the number of lines is equal the number of photo-identification photographs, and the columns contain all available metadata. The *linkMetadataToFlight* function merges the media data with the flight data, returning an R data frame with all the flight logs, or a list with a data frame for each flight log data.

Step 4: syncing multimedia data

Finally, the multiple media data sources can be synchronized based on the intersection of their recording time (Fig. 1e). Using the function *syncMedia*, video and audio files that were sampled concurrently and selected by the *selectVideos* and *selectAudios* functions can be trimmed to match the time intersection, and merged into a single file or exported as separate media files. Other auxiliary text data (e.g. GPS trackers, heart rate loggers, flight logs) recorded simultaneously in the field can be synchronized based on the intersection of their sampling time and merged into a single text database using the function *syncData*, as long as the input clocks are precisely synced.

Auxiliary functions for post-processing multimedia data

The *MAMMals* R package was designed to streamline the pre-processing of photo-identification and multimedia data; thus its workflow does not include the post-processing of the

biological data of interest. After linking the photographs with the useful parts of the videos and audios, manual or semiautomatic extraction of the target data is required. This may include video playback to quantify behavioural states and events (e.g. Torres et al. 2018), morphometry or health variables (e.g. Christiansen et al. 2020); automatic detection of species (e.g. Gray et al. 2019); or photographic comparison needed to identify individual animals (e.g. Urián et al. 2015). To efficiently measure and extract such biological data from photos, videos, and audio data, we point the reader to the growing number of computational tools available elsewhere (e.g. Abràmoff et al. 2004; Friard and Gamba 2016, Beery et al. 2020; Schneider et al. 2018; Torres and Bierlich 2020; Bird and Bierlich 2020). We exemplify one case of post-processing behavioural data in the next section, but here we highlight that the *MAMMals* R package also contains some functions and utilities to assist with the post-processing of the linked multimedia data or auxiliary data. For instance, one can use *MAMMals* to wrangle and convert information from the drone flight log data, such as gimbal and camera angles, GPS coordinates, digital compass and barometer sensors. We conceptually divide these functions into data tools and visualization (Table 1), which are, respectively, identified by the prefixes *do* and *view*. For instance, *doCorrectAngle* can be used to correct drone yaw ranging from 0 to 180 or -180 to 0 to 0 to 360, and the function *viewFlightPath* can be used to visualize a 2D drone flight path with photos as points, using data from the *linkMetadataToFlight*.

An illustrative case study

To illustrate the utility of the *MAMMals* R package, we used individual and behavioural data collected from a coastal bottlenose dolphin population in Laguna, southern Brazil,

where some individual dolphins forage near the coast with net-casting fishers (Simões-Lopes et al. 1998). To explore the dolphins' foraging behaviour, we combined standard photo-identification with overhead video, recorded using a commercially available drone (DJI Mavic Pro) with a built-in high-resolution camera mounted on a gimbal. We hovered the drone over the study area above 60 m to minimize potential disturbance to the dolphins (Fettermann et al. 2019), and follow all safety flight guidelines (Fiori et al. 2017; Raoult et al. 2020). The drone camera covered an area of *ca.* 7500 m², including the coast where the fishers wait for dolphins and *ca.* 60 m of the lagoon canal. Simultaneously, two photographers registered the dolphins' dorsal fins for posterior individual identification based on nicks, notches, scars and skin lesions, following photo-identification protocols (Hammond et al. 1990). One photographer positioned ashore used a DSLR Canon 60D camera equipped with a 100–400 mm lenses to photograph all dolphins in the video footage area, while the second photographer stood on a 1.5 m platform 3 m behind the fishers and used a DSLR Canon 7D MkII with built-in GPS and digital compass and a 70–300 mm lenses to identify the individual dolphins that approach the fishers to interact. This photographer was always captured in the drone footage and used a flash (Yongnuo) pointing up, so the timing of the photographs taken could be verified in the video to double-check if the clocks of the camera and drone were properly synced.

To illustrate two types of behavioural data that can be measured from the merged video and photo-identification dataset, we tracked (i) the foraging behaviour of individual dolphins in terms of distance and heading angles relative to the coast over time (Fig. 2a); and (ii) the foraging behaviour of a group of dolphins in terms of spatial cohesion and diving synchrony (Fig. 2b). In both cases, we used the *MAMMals* R package to automatically select examples of drone videos containing photo-identified dolphins from a total of

Table 1 Auxiliary functions provided in the *MAMMals* R package to assist data wrangling, conversion and visualization

Function	Description
<i>doConvertAngle</i>	Convert angle in degrees ranging from 0 to 180 for the right side and -180 to 0 for the left side, and return angle in degrees ranging from 0 to 360
<i>doCorrectCameraYaw</i>	Correct camera yaw relative to drone yaw in degrees relative to geographical North
<i>doFilterDroneHeight</i>	Filter drone data to data in height equal or above a value in meters
<i>doFilterGimbalPitch</i>	Filter drone data to when the camera is pointing straight down, i.e. $\text{pitch} = -90^\circ$
<i>doCalcDistanceX</i> and <i>doCalcDistanceY</i>	Calculate horizontal (<i>doCalcDistanceX</i>) or vertical (<i>doCalcDistanceY</i>) distance of an object marked in any image tool (e.g. ImageJ). This distance will be used to transform the marked object into a GPS position using the origin point as reference
<i>doNewLatitude</i>	Calculate latitude of an object given the distance and angle to a reference point using photogrammetry data
<i>doNewLongitude</i>	Calculate longitude of an object given the distance and angle to a reference point using photogrammetry data
<i>doAngleToDec</i>	Convert the degree-min-sec format (e.g. $28^\circ 29' 44.77''$ S) into degree decimals
<i>doCalcRadiusEarth</i>	Calculate radius of Earth at any given latitude and altitude
<i>viewFlightPath</i>	Plot drone flight path and relate with photo-identification data, if available

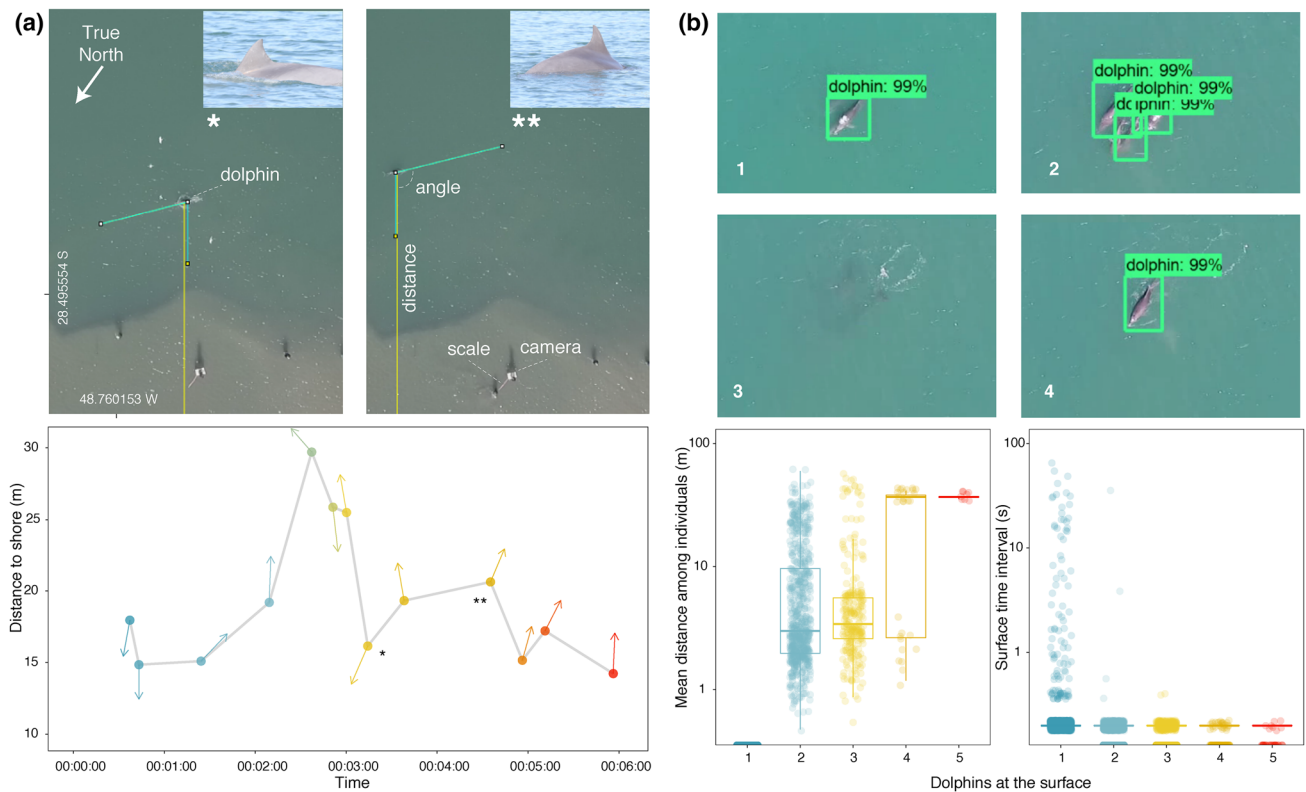


Fig. 2 Examples of individual- and group-level behaviour of photo-identified mammals extracted from overhead videos. **a** Tracking the foraging behaviour of individual coastal dolphins, in terms of distance and angle to the shore. The *MAMMals* package was used to automatically select and clip a video containing a solitary photo-identified dolphin (inset photo-identification). The video was then post-processed, when dolphins' distances (yellow lines in the picture; y-axis in the plot) and angles (cyan lines in the picture, with the middle point centred on the dolphin; arrows in the plot, whose colours indicate temporal sequence) relative to shore were measured each time they surfaced to breathe. Distances measured in pixels were converted to meters based on a 1-m scale placed behind the photographer; angles measured in degree relative to the shore, were converted to radians, considering the True North as a reference. **b** Group cohesion and dive synchrony of photo-identified bottlenose dolphins, in terms of relative distance to each member and timing of surfacing.

56.6 h of footage and 3614 photographs of 21 identified individual dolphins. First, we used the functions *getPhotoMetadata* and *getVideoMetadata* to extract and organize the metadata of photographs and videos, extracted the drone flight logs and used some of the auxiliary functions to correct angles of drone footage (*doConvertAngle*, *doCorrectCameraYaw*) and filter off flights that were too low (*doFilterDroneHeight*) or in which the camera was not pointed straight down (*doFilterGimbalPitch*).

To describe (i) the individual-level foraging, we then used the function *selectVideos* to identify drone videos taken when there were 1 or 2 dolphins at the interaction site, and the function *getVideoClip* to crop 6-min video clips

The *MAMMals R* package was used to select the photographs with the dolphins' dorsal fins for posterior identification of the 5 group members. The group of 5 dolphins were then tracked over time with a custom computer vision model trained to detect dolphins in drone videos. Cohesion was estimated as the average Euclidean distances among the centroids of all dolphins detected (i.e. the green rectangles with detection scores) every 0.2 s and converted to meters using a known 1-m scale captured in the video (not shown here). Synchrony was estimated as the time difference between detections. Pictures 1–4 illustrate a case of diving sequence of a subgroup of 5 dolphins, in which 1 individual is detected first, followed by three that surfaced simultaneously, and then by the fifth individual after a 2-s lag. Box plots present the distribution of mean distances and breath intervals (y-axes) across different number of simultaneous detections (circles) of dolphins at the surface (x-axes) during a ~20-min drone video

around the photographs taken. Next, we manually processed these clips with the open-source software *imageJ* (Abràmoff et al. 2004); each time the photo-identified dolphin surfaced to breathe, we used the 'straight line' tool to measure the distance of the dolphin from shore, and the 'angle' tool to measure the angle between the dolphin's heading and the shore. In videos with more than one dolphin at the site, we distinguished photo-identified dolphins recorded in the video at the same time but in different places using the angle (available in the metadata of the photographs) between the dolphin and the camera equipped with built-in compass used for photo-identification. Finally, we converted the distances measured in pixels to meters based on a 1-m scale captured

in the drone video, and converted the angles measured in degrees relative to the shore to radians, considering the True North as a reference. In Fig. 2a, we present an example of these data on the distances and angles of a photo-identified individual dolphin foraging close to shore.

To describe the group-level foraging, we used the functions *selectVideos* and *getVideoClip* to select the photographs of all dolphins foraging in groups and trim the complete 20-min drone video into a shorter video around the time that the photos were taken. We first photo-identified individuals manually, and then measured group cohesion and dive synchrony, in terms of relative distance to each member of the group and timing of surfacing. To do that, we have used a convolutional neural network object detection classifier (He et al. 2016) to automatically detect and count dolphins in the drone footage. We have re-trained a TensorFlow pre-trained classifier with Faster-RCNN model architecture (Ren et al. 2015) using 838 drone video frames in which dolphins were manually labelled using LabelImg (Tzutalin 2015), and 200 other such images for testing the model. We then applied this supervised learning computer-vision model to detect and count the number of dolphins at every 0.2 s of the drone video, i.e. every 5 frames of a 25 FPS video (for a similar approach, see Guo et al. 2020). We highlight that although we have used machine learning to post-processes the video clips, this procedure could also be done manually. For instance, one can extract short .avi clips with a framerate of 1 fps using the *getVideoClip* function, and then import the clip to *imageJ* to measure the inter-individual distances and surface timing. To estimate the group cohesion, we calculated the relative time between each dolphin detection, considering greater cohesion when individuals are closer together; to estimate diving synchrony, we calculated the lag between dolphin surfacing times, considering greater synchrony when their breath intervals were shorter. We measured the group cohesion as the average Euclidean distance, in pixels, between the centroids of all dolphins detected in each frame, and converted these distances into meters based on a known 1-m scale recorded in the drone video. We measured the diving synchrony as the time lag between detections, considered the group to be in synchronous diving when more than one dolphin was detected in the same video frame. In Fig. 2b, we present these data on group cohesion and dive synchrony as the distribution of mean distances and breath intervals among different number of dolphins at the surface.

Caveats

The tools herein presented assist the organization of simultaneous sampling methods, but caveats exist. First, the level of detail of the outputs—be them the merged databases or the cropped and synced media—may depend on the accessibility of the study system.

We have illustrated how the *MAMMals* tools work when recording and tracking coastal dolphins, but these tools could be used to process multimedia of mammals individually identifiable from photographs taken from the ground or sea level (e.g. sperm and humpback whale caudal fins, or blue whale pigmentation; Hammond et al. 1990) and from overhead (e.g. head of right whales, or other identifiable body parts of marine and terrestrial mammals; Landeo-Yauri et al. 2020; Maeda et al. 2021). However, in our example, we had the advantage of keeping the photographer in the overhead video frame at all times for recording the position of the GPS-equipped camera as a reference point, and for double-checking the synchrony between the video and photograph data streams. This setup is rather unusual for studies of free-ranging mammals, and require the sampling design to be adapted to fit the reality of other study systems. For example, boat-based focal follows of cetaceans could aim to keep the boat close to the group most of the time to allow the photographer to be in the overhead video frame, or overhead behavioural sampling of terrestrial mammals can be focused on a relatively small open area.

The second limitation of our tools is that the precision of the link between the photo-identification and the other multimedia can be dependent on group size and group cohesion. In our example, we tracked solitary and small groups of animals that can be easily photo-identified, but mismatches in individual identification can occur when collecting data from multiple individuals at the same time, such as in large and tight groups. Our drone videos can contain multiple individuals, leading to the possibility that an individual photographed at a given time could be linked to multiple individuals that appear in the drone video at that time. We have resolved this by keeping the photographer in the overhead video frame and relying on the angle of the built-in digital compass of the camera to tease apart individuals in the overhead footage. However, these decisions become increasingly more difficult to make as the group size, and the rate of pictures taken, increase, and/or the groups become tighter and closer to the photographer. In such situations, our tools could still help defining the timestamps of sampling events to extract group-level (but not individual) data or identify subgroups of animals.

Closing remarks

Our tools to streamline the use of multimedia data with traditional individual identification methods are steps toward the integration of multiplatform behavioural sampling on free-ranging mammals. We acknowledge there is room for improvement and, to encourage further development of these

tools collectively, we provide all the code of the *MAMMals* package in an open repository (<https://bitbucket.org/maucantor/mammals/>). We hope to inspire further collective work in the scientific community to generalize the process of linking multiple sampling platforms to refine the collection and processing of data of individual animals. More importantly, we hope these computer tools improve the raw material needed to promote new insights on the population dynamics, ecological interactions and behaviour of free-ranging animals.

Appendix

The *MAMMals* R Package can be downloaded and installed from its open repository <https://bitbucket.org/maucantor/mammals/>. Please visit the *MAMMals* R Package website at <https://mammals-rpackage.netlify.app> for list of all functions and a step-by-step guide on (i) extracting meta-data of multimedia files; (ii) aligning multimedia files; (iii) linking photographs with multimedia data; and (iv) syncing multimedia data.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s42991-021-00189-0>.

Acknowledgements We are grateful to FG Daura-Jorge and DR Farine for the insightful discussions, field logistics and support during the development of these tools. We also thank all researchers involved in data sampling for their hard work (FG Daura-Jorge, C Bezamat, B Romeu, DR Farine, JVS Valle-Pereira, PV Castilho, BS Silva, N da Silva, LF da Rosa, CF Alves), and FG Daura-Jorge, DR Farine, JVS Valle-Pereira and ED Strauss for thoughtful comments on the manuscript.

Author contributions Both authors contributed to the study conception and design, to the development of the computational tools and the R package, to data sampling and analyses. The manuscript was written, read and approved by both authors.

Funding Open Access funding enabled and organized by Projekt DEAL. This study is part of the Long-term Ecological Research Program ‘Sistema Estuarino de Laguna e Adjacências’ (SELA PELD) funded by the Conselho Nacional de Pesquisa e Desenvolvimento Tecnológico (CNPq Brazil #445301/2020-1), and by the Brazil-Germany PROBRAL Research Grant (#23038.002643/2018-01) from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and the Deutscher Akademischer Austauschdienst (DAAD). Fieldwork was further supported by the National Geographic Society (Discovery Grant WW210R-17), The Society for Marine Mammalogy, and The Max Planck Society (granted to MC) and The Animal Behaviour Society (granted to AMSM). AMSM received doctoral scholarship by CAPES Brazil (Finance Code 001) and an exchange scholarship from CAPES-DAAD PROBRAL Grant. MC received post-doctoral fellowship from CNPq (#153797/2016-9), CAPES (#88881.170254/2018-01), and the Department for the Ecology of Animal Societies at the Max Planck Institute of Animal Behavior.

Declarations

Conflict of interest There are no competing interests with this study.

Data availability We provide all the code and data to reproduce the analyses in the open-access repository <https://bitbucket.org/maucantor/mammals/src/master/>.

Code availability All the code is available in the open-source R package (*MAMMals*) at <https://bitbucket.org/maucantor/mammals/src/master/>. We also provide a step-by-step tutorial at <https://mammals-rpackage.netlify.app/index.html>.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abràmoff MD, Magalhães PJ, Ram SJ (2004) Image processing with imageJ. *Biophotonics Int* 11:36–41
- Alarcón-Nieto G, Graving JM, Klarevas-Irby JA, Maldonado-Chaparro AA, Mueller I, Farine DR (2018) An automated barcode tracking system for behavioural studies in birds. *Methods Ecol Evol* 9:1536–1547. <https://doi.org/10.1111/2041-210X.13005>
- Altmann J (1974) Observational study of behavior: sampling methods. *Behaviour* 49:227–266
- Beery S, Wu G, Rathod V, Votel R, Huang J (2020) Context R-CNN: long term temporal context for per-camera object detection. *IEEE/CVF Conf Comput vis Pattern Recognit*. <https://doi.org/10.1109/CVPR42600.2020.01309>
- Bird CN, Bierlich KC (2020) CollatriX: a GUI to collate Morpho-MetriX outputs. *J Open Source Soft* 5:2328. <https://doi.org/10.21105/joss.02328>
- Burnett JD, Lemos L, Barlow D, Wing MG, Chandler T, Torres LG (2019) Estimating morphometric attributes of baleen whales with photogrammetry from small UASs: a case study with blue and gray whales. *Mar Mamm Sci* 35:108–139. <https://doi.org/10.1111/mms.12527>
- Cheng J, Xie B, Lin C, Ji L (2012) A comparative study in birds: call-type-independent species and individual recognition using four machine-learning methods and two acoustic features. *Bioacoustics* 21:157–171. <https://doi.org/10.1080/09524622.2012.669664>
- Christiansen F, Dawson S, Durban J, Fearnbach H, Miller C, Bejder L, Uhart M, Sironi M, Corkeron P, Rayment W, Leunissen E, Haria E, Ward R, Warick H, Kerr I, Lynn M, Pettis H, Moore M (2020) Population comparison of right whale body condition reveals poor state of the North Atlantic right whale. *Mar Ecol Prog Ser* 640:1–16. <https://doi.org/10.3354/meps13299>
- Clapham M, Miller E, Nguyen M, Darimont CT (2020) Automated facial recognition for wildlife that lack unique markings: a deep learning approach for brown bears. *Ecol Evol* 10:12883–12892. <https://doi.org/10.1002/ece3.6840>

- Clark DR, Meffert C, Baggili I, Breitinger F (2017) DROP (DRone open source parser) your drone: forensic analysis of the DJI Phantom III. *Digit Investig* 22:S3–S14. <https://doi.org/10.1016/j.diin.2017.06.013>
- Clutton-Brock TH, Sheldon BC (2010) Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology. *Trends Ecol Evol* 25:562–573. <https://doi.org/10.1016/j.tree.2010.08.002>
- Coulson T (2020) Ecology and evolution is hindered by the lack of individual-based data. In: Dobson A, Tilman D, Holt RD (eds) *Unsolved problems in ecology*. Princeton University Press, Princeton
- Dawson SM, Bowman MH, Leunissen E, Sirguy P (2017) Inexpensive aerial photogrammetry for studies of whales and large marine animals. *Front Mar Sci* 4:1–7. <https://doi.org/10.3389/fmars.2017.00366>
- Dell AI, Bender JA, Branson K, Couzin ID, de Polavieja GG, Noldus LPJJ, Pérez-Escudero A, Perona P, Straw AD, Wikelski M, Brose U (2014) Automated image-based tracking and its application in ecology. *Trends Ecol Evol* 29:417–428. <https://doi.org/10.1016/j.tree.2014.05.004>
- Durban JW, Fearnbach H, Barrett-Lennard LG, Perryman WL, Leroi DJ (2015) Photogrammetry of killer whales using a small hexacopter launched at sea. *J Unmanned Veh Syst* 3:131–135. <https://doi.org/10.1139/juvs-2015-0020>
- Erbe C, Salgado-Kent C, de Winter S, Marley S, Ward R (2020) Matching signature whistles with photo-identification of Indo-Pacific bottlenose dolphins (*Tursiops aduncus*) in the Fremantle Inner Harbour, Western Australia. *Acoust Aust* 48:23–38. <https://doi.org/10.1007/s40857-020-00178-2>
- Ferreira AC, Silva LR, Renna F, Brandl HB, Renoult JP, Farine DR, Doutrelant CRC (2020) Deep learning-based methods for individual recognition in small birds. *Methods Ecol Evol* 11:1072–1085. <https://doi.org/10.1111/2041-210X.13436>
- Fettermann T, Fiori L, Bader M, Doshi A, Breen D, Stockin KA, Bollard B (2019) Behaviour reactions of bottlenose dolphins (*Tursiops truncatus*) to multirotor Unmanned Aerial Vehicles (UAVs). *Sci Rep* 9:8558. <https://doi.org/10.1038/s41598-019-44976-9>
- Fiori L, Doshi A, Martinez E, Orams MB, Bollard-Breen B (2017) The use of unmanned aerial systems in marine mammal research. *Remote Sens* 9:543. <https://doi.org/10.3390/rs9060543>
- Francisco FA, Nührenberg P, Jordan A (2020) High-resolution, non-invasive animal tracking and reconstruction of local environment in aquatic ecosystems. *Mov Ecol* 8:1–12. <https://doi.org/10.1186/s40462-020-00214-w>
- Friard O, Gamba M (2016) BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods Ecol Evol* 7:1325–1330. <https://doi.org/10.1111/2041-210X.12584>
- Graving JM, Chae D, Naik H, Li L, Koger B, Costelloe BR, Couzin ID (2019) DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. *Elife* 8:620245. <https://doi.org/10.7554/eLife.47994>
- Gray PC, Bierlich KC, Mantell SA, Friedlaender AS, Goldbogen JA, Johnston DW (2019) Drones and convolutional neural networks facilitate automated and accurate cetacean species identification and photogrammetry. *Methods Ecol Evol* 10:1490–1500. <https://doi.org/10.1111/2041-210X.13246>
- Grolemund G, Wickham H (2011) Dates and times made easy with lubridate. *J Stat Softw* 40:1–25. <https://doi.org/10.18637/jss.v040.i03>
- Guo S, Xu P, Miao Q, Shao G, Chapman CA, Chen X, He G, Fang D, Zhang H, Sun Y, Shi Z, Li B (2020) Automatic identification of individual primates with deep learning techniques. *iScience* 23:101412. <https://doi.org/10.1016/j.isci.2020.101412>
- Hammond PS, Mizroch SA, Donovan GP (1990) Individual recognition of cetaceans: use of photo-identification and other techniques to estimate population parameters. *Rep Int Whaling Commis (Specieal Issue 12)*
- Hartman K, van der Harst P, Vilela R (2020) Continuous focal group follows operated by a drone enable analysis of the relation between sociality and position in a group of male Risso's dolphins (*Grampus griseus*). *Front Mar Sci* 7:1–13. <https://doi.org/10.3389/fmars.2020.00283>
- He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. *Proc IEEE Conf Comput vis Pattern Recognit*. <https://doi.org/10.1109/CVPR.2016.90>
- Hill AP, Prince P, Snaddon JL, Doncaster CP, Rogers A (2019) AudioMoth: A low-cost acoustic device for monitoring biodiversity and the environment. *HardwareX* 6:e00073. <https://doi.org/10.1016/j.ohx.2019.e00073>
- Janik VM, Sayigh LS (2013) Communication in bottlenose dolphins: 50 years of signature whistle research. *J Comp Physiol A* 199:479–489. <https://doi.org/10.1007/s00359-013-0817-7>
- Karanth KU, Nichols JD (1998) Estimation of tiger densities in India using photographic captures and recaptures. *Ecology* 79:2852–2862. [https://doi.org/10.1890/0012-9658\(1998\)079\[2852:EOT-DII\]2.0.CO;2](https://doi.org/10.1890/0012-9658(1998)079[2852:EOT-DII]2.0.CO;2)
- Karczmarski L, Chan SCY, Rubenstein DI, Chui SYS, Cameron EZ (2022a) Individual identification and photographic techniques in mammalian ecological and behavioural research – Part 1: Methods and concepts. *Mamm Biol (Special Issue)* 102(3). <https://link.springer.com/journal/42991/volumes-and-issues/102-3>
- Karczmarski L, Chan SCY, Chui SYS, Cameron EZ (2022b) Individual identification and photographic techniques in mammalian ecological and behavioural research – Part 2: Field studies and applications. *Mamm Biol (Special Issue)* 102(4). <https://link.springer.com/journal/42991/volumes-and-issues/102-4>
- Katona SK, Whitehead HP (1981) Identifying humpback whales using their natural markings. *Polar Rec* 20:439–444. <https://doi.org/10.1017/S003224740000365X>
- Krause J, Krause S, Arlinghaus R, Psorakis I, Roberts S, Rutz C (2013) Reality mining of animal social systems. *Trends Ecol Evol* 28:541–551. <https://doi.org/10.1016/j.tree.2013.06.002>
- Landeo-Yauri SS, Ramos EA, Castelblanco-Martínez DN, Niño-Torres CA, Searle L (2020) Using small drones to photo-identify Antillean manatees: a novel method for monitoring an endangered marine mammal in the Caribbean Sea. *Endanger Species Res* 41:79–90. <https://doi.org/10.3354/esr01007>
- Longden EG, Elwen SH, McGovern B, James BS, Embling CB, Gridley T (2020) Mark–recapture of individually distinctive calls—a case study with signature whistles of bottlenose dolphins (*Tursiops truncatus*). *J Mammal*. <https://doi.org/10.1093/jmammal/gyaa081>
- Maeda T, Ochi S, Ringhofer M, Sosa S, Sueur C, Hirata S, Yamamoto S (2021) Aerial drone observations identified a multilevel society in feral horses. *Sci Rep* 11:71. <https://doi.org/10.1038/s41598-020-79790-1>
- Marks M, Qiuhan J, Sturman O, von Ziegler L, Kollmorgen S, von der Behrens W, Mante V, Yanik BJMF (2021) Deep-learning based identification, pose estimation and end-to-end behavior classification for interacting primates and mice in complex environments. *bioRxiv*. <https://doi.org/10.1101/2020.10.26.355115>
- Mersch DP, Crespi A, Keller L (2013) Tracking individuals shows spatial fidelity is a key regulator of ant social organization. *Science* 340:1090–1093. <https://doi.org/10.1126/science.1234316>
- Muller Z, Cantor M, Cuthill IC, Harris S (2018) Giraffe social preferences are context dependent. *Anim Behav* 146:37–49. <https://doi.org/10.1016/j.anbehav.2018.10.006>
- Odum EP, Barrett GW (1971) *Fundamentals of ecology*. Saunders, Philadelphia

- Payne R, Brazier O, Dorsey EM, Perkins J, Rowntree V, Titus A (1983) External features in southern right whales (*Eubalaena australis*) and their use in identifying individuals. Communication and behavior of Whales. Westview Press, Boulder, pp 371–445
- Pérez-Escudero A, Vicente-Page J, Hinz RC, Arganda S, De Polavieja GG (2014) IdTracker: tracking individuals in a group by automatic identification of unmarked animals. Nat Methods 11:743–748. <https://doi.org/10.1038/nmeth.2994>
- R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Raoult V, Tosetto L, Williamson JE (2018) Drone-based high-resolution tracking of aquatic vertebrates. Drones 2:1–14. <https://doi.org/10.3390/drones2040037>
- Raoult V, Colefax AP, Allan BM, Cagnazzi D, Castelblanco-Martínez N, Ierodiaconou D, Johnston DW, Landeo-Yauri S, Lyons M, Pirotta V, Schofield G, Butcher PA (2020) Operational protocols for the use of drones in marine animal research. Drones 4:64. <https://doi.org/10.3390/drones4040064>
- Ren S, He K, Girshick R, Sun J (2015) Faster R-CNN: towards real-time object detection with region proposal networks. Adv Neural Info Process Syst 28:91–99
- Schneider S, Taylor GW, Kremer S (2018) Deep learning object detection methods for ecological camera trap data. In: 2018 15th Conference on computer and robot vision (CRV), pp. 321–328
- Schneider S, Taylor GW, Linquist S, Kremer SC (2019) Past, present and future approaches using computer vision for animal re-identification from camera trap data. Methods Ecol Evol 10:461–470. <https://doi.org/10.1111/2041-210X.13133>
- Silvy NJ, Lopez RR, Peterson MJ (2005) Wildlife marking techniques. In: techniques for Wildlife Investigations and Management. The Wildlife Society, Bethesda, MD
- Simões-Lopes PC, Fabián ME, Menegheti JO (1998) Dolphin interactions with the mullet artisanal fishing on southern Brazil: a qualitative and quantitative approach. Rev Bras Zool 15:709–726. <https://doi.org/10.1590/S0101-81751998000300016>
- Speed CW, Meekan MG, Bradshaw CJA (2007) Spot the match: wildlife photo-identification using information theory. Front Zool 4:1–11. <https://doi.org/10.1186/1742-9994-4-2>
- Toms CN, Stone T, Och-Adams T (2020) Visual-only assessments of skin lesions on free-ranging common bottlenose dolphins (*Tursiops truncatus*): reliability and utility of quantitative tools. Mar Mammal Sci 36:744–773. <https://doi.org/10.1111/mms.12670>
- Torres WI, Bierlich KC (2020) MorphoMetriX: a photogrammetric measurement GUI for morphometric analysis of megafauna. J Open Source Soft 5:1825. <https://doi.org/10.21105/joss.01825>
- Torres LG, Nieukirk SL, Lemos L, Chandler TE (2018) Drone up! quantifying whale behavior from a new perspective improves observational capacity. Front Mar Sci 5:1–14. <https://doi.org/10.3389/fmars.2018.00319>
- Tzatalin (2015) LabelImg. <https://github.com/tzatalin/labelImg>
- Urián K, Gorgone A, Read A, Balmer B, Wells RS, Berggren P, Durban J, Eguchi T, Rayment W, Hammond PS (2015) Recommendations for photo-identification methods used in capture-recapture models with cetaceans. Mar Mammal Sci 31:298–321. <https://doi.org/10.1111/mms.12141>
- Walker KA, Trites AW, Haulena M, Weary DM (2012) A review of the effects of different marking and tagging techniques on marine mammals. Wildl Res 39:15–30. <https://doi.org/10.1071/WR10177>
- Whytock RC, Christie J (2017) Solo: an open source, customizable and inexpensive audio recorder for bioacoustic research. Methods Ecol Evol 8:308–312. <https://doi.org/10.1111/2041-210X.12678>
- Würsig B, Würsig M (1977) The photographic determination of group size, composition, and stability of coastal porpoises (*Tursiops truncatus*). Science 198:755–756. <https://doi.org/10.1126/science.198.4318.755>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.