



Assessing the effects of sampling frequency on behavioural classification of accelerometer data

J.L. Hounslow^{a,b,*}, L.R. Brewster^{a,c}, K.O. Lear^a, T.L. Guttridge^c, R. Daly^b, N.M. Whitney^d,
A.C. Gleiss^a

^a Centre for Sustainable Aquatic Ecosystems, Harry Butler Institute, Murdoch University, 90 South Street, Murdoch, WA 6150, Australia

^b Save our Seas Foundation- D'Arros Research Centre, Rue Philippe Plantamour 20, CH-9 1201 Geneva, Switzerland

^c Bimini Biological Field Station Foundation, South Bimini, Bahamas.

^d Anderson Cabot Center for Ocean Life, New England Aquarium, Central Wharf, Boston, MA 02110, USA

ARTICLE INFO

Keywords:

Random forest algorithm

Nyquist frequency

Elasmobranch

Bio-logging

Machine learning

ABSTRACT

Understanding the behaviours of free-ranging animals over biologically meaningful time scales (e.g., diel, tidal, lunar, seasonal, annual) gives unique insight into their ecology. Bio-logging tools such as accelerometers allow the remote study of elusive or inaccessible animals by recording high resolution movement data. Machine learning (ML) is becoming a common tool for automatic classification of behaviours from these types of large data sets. These classifiers often perform best using high sampling frequencies; however, these frequencies also limit archival device recording duration through elevated battery and memory use. In this study we assess the effect of sampling frequency on a ML algorithm's ability to correctly classify behaviours from accelerometer data and present a framework for programming bio-logging devices that maintains classifier performance while optimizing data collection duration. Accelerometer data (30 Hz) were collected from juvenile lemon sharks (*Negaprion brevirostris*) during semi-captive trials at Bimini, Bahamas, and were ground-truthed to a discrete catalogue of behaviours through direct observation of sharks during trials. The ground-truthed data were re-sampled to a range of sampling frequencies (30, 15, 10, 5, 3 and 1 Hz) and behaviours (swim, rest, burst, chafe, headshake) were classified using a random forest ML algorithm. We demonstrate that as sampling frequency decreases, classifier performance decreases. Best overall classification was achieved at 30 Hz (F -score > 0.790), although 5 Hz was appropriate for classification of swim and rest (F -score > 0.964). For fine-scale behaviours characterised by faster kinematics (headshake, burst and chafe), classification performance was lower across the entire range of sampling frequencies (0.535–0.846, 1–30 Hz), though did not decrease significantly until sampling frequency was < 5 Hz. We discuss the effects of signal aliasing and recommend that for best classification of fine-scale behaviours, frequencies > 5 Hz are required. However, when seeking to maximise the available device memory and battery capacity and therefore extend deployment duration, 5 Hz is an appropriate sampling frequency for classifying behaviours in similar-sized animals.

1. Introduction

Behavioural studies provide great insight into an animal's ecology; knowledge of what, where and how animals are behaving sheds light on migration patterns, foraging, and reproductive strategies (Hays et al., 2016). As such, understanding an animal's behavioural response to internal and external factors contributes to informing conservation issues (e.g., the consequences of environmental change and reserve or policy planning) (Sutherland, 1998; Cooke, 2008; Abrahms et al., 2016). However, studying the behaviour of cryptic free-ranging animals in their natural environment is often challenging and therefore rarely

attempted (Hays et al., 2016). It is particularly difficult to obtain continuous direct observations of aquatic, highly mobile, migratory, or nocturnal species, especially when the presence of a human observer may inadvertently alter natural behaviour (Gleiss et al., 2009a; Brown et al., 2013; Wilson et al., 2006; Hammond et al., 2016). Recent advances in bio-logging technologies present a number of tools well suited for remotely observing wild animal behaviour, physiology, movements, and their surrounding environments, recorded via an electronic, animal-borne logger (Rutz and Hays, 2009; Wilmers et al., 2015).

Accelerometers are prominent and emerging bio-logging sensors used in behavioural studies. Since particular postures or body

* Corresponding author at: Centre for Sustainable Aquatic Ecosystems, Harry Butler Institute, Murdoch University, 90 South Street, Murdoch, WA 6150, Australia.
E-mail address: jennahounslow@gmail.com (J.L. Hounslow).

movements correspond to distinct behaviours, the measurement of high-resolution acceleration data enables remote study of behaviour in unprecedented detail (Wilson et al., 2006; Shepard et al., 2008b; Sakamoto et al., 2009; Payne et al., 2014). Sampling acceleration continuously in all three axes is considered most favourable for characterising animal behaviours that occur in three dimensions, with higher sampling frequencies assumed to improve determination of behaviours (Nathan et al., 2012; Brown et al., 2013). In general, signal processing theory dictates that accelerometers should be programmed with a sampling frequency at least twice that of the minimum frequency of the most rapid body movement of interest, referred to as the Nyquist criterion (Mallat, 1999; Chen and Bassett Jr, 2005; Halsey et al., 2011; Graf et al., 2015). A wide range of sampling frequencies have been used to relate acceleration data to behaviours in terrestrial and aquatic animals, from as low as 1 Hz (Whitney et al., 2007) up to 100 Hz (Broell et al., 2013; le Roux et al., 2017). When recording at sub-second frequencies (> 1 Hz) in three axes, accelerometers quickly produce millions of data points, making analysis of acceleration data a time-consuming task. This is one of the main challenges when using this technology for the study of animal behaviour (Shepard et al., 2008b; Brown et al., 2013).

Machine learning (ML) enables automatic classification of behaviours from large, complex acceleration data sets (Grünewälder et al., 2012; Nathan et al., 2012; Bidder et al., 2014; Ladds et al., 2017; Leos-Barajas et al., 2017). Supervised learning involves labelling ground-truthed data, from direct observations (in captivity or the wild), to train a classification algorithm and predict behaviour from new data (Gao et al., 2013; Valletta et al., 2017). Recent studies have combined the strengths of multiple supervised ML techniques to improve classification of animal behaviour (Ladds et al., 2017; Brewster et al., 2018). Despite ML overcoming issues associated with manual analysis of acceleration data for behavioural classification, a crucial problem remains: archival accelerometers record data on an internal memory card. Both memory storage and battery capacity of most commercially available devices are quickly exceeded when sampling at frequencies assumed to be high enough to distinguish fine-scale behaviours (Halsey et al., 2008). Acceleration data can therefore only be collected for brief periods (< 1 week), giving a limited representation of how animals respond to environmental conditions. Therefore, the optimal sampling frequency for the classification of behaviours is a compromise between acceleration data-resolution and extending temporal resolution by maximising the available memory and battery capacity. We assume that high sampling frequencies, which come at a cost, are required in order to classify behaviours from acceleration data, however this assumption has not been tested.

The aim of this study was to assess how the choice of sampling frequency influences the classification of behaviour from accelerometer data. To achieve this, we used ground-truthed accelerometer data collected from lemon sharks *Negaprion brevirostris* and presented by Brewster et al. (2018), to develop a supervised ML algorithm for automatic classification of the accelerometer data. We evaluated the effects of sampling frequency on the performance of this algorithm, in order to identify the optimal sampling frequency as a compromise between data resolution and the rate of memory and battery consumption. In addition, we assessed the relative performance of the algorithm for discrete behaviours, their unique acceleration characteristics, and the Nyquist criterion.

2. Materials & methods

2.1. Study species and captive trials

Juvenile lemon sharks (*Negaprion brevirostris*) ($n = 4$) were captured using gillnets off South Bimini, Bahamas, and transported to purpose-built pens for semi-captive trials. Once acclimated to captivity sharks were tagged with a dorsally mounted Cefas G6a + triaxial

accelerometers (Cefas Inc., Lowestoft, UK). A consistent and secure attachment is required, because acceleration amplitude changes with position (Whitney et al., 2012). Tags were attached to individual sharks by passing nylon monofilament through two holes, 1.5 mm in diameter, made in the base of the first dorsal fin with a hypodermic needle, then looping through corresponding pre-drilled holes in the accelerometer package. The monofilament was secured on the reverse side of the fin using stainless steel crimps with two small plastic plates in between. A medical grade porous orthotic foam was placed between the plates and the shark's skin to minimise rubbing and damage to the shark's skin (Brewster et al., 2018). Acceleration data, recorded at 30 Hz, were validated (ground-truthed) by observing tagged sharks whilst in captivity. Each second of the observation period was labelled as one of five behaviours: rest (motionless), swim (steady undulatory motion), burst (fast-start), chafe (rolling motion), and headshake (side to side movement of head). See Brewster et al., 2018 for capture, transport, husbandry and behavioural ethogram details.

2.2. Data analysis

To determine the optimal sampling frequency for classifying the behaviours, the ground-truthed, raw acceleration data collected by Brewster et al. (2018), were re-sampled for each axis, using the 're-sample' function in IGOR Pro v7.06 (WaveMetrics Inc., Lake Oswego, Oregon, USA). Data were 'down-sampled' using decimation by omission, whereby every 'nth' data point of the original 30 Hz data set was systematically deleted according to the new sampling frequency (Broell et al., 2013; Sur et al., 2017). For example, when resampling raw 30 Hz data down to 15 Hz, every other data point was omitted from the original data set using a decimation rate of 2. Resampling by decimation requires the resampled frequencies to be a factor of the raw sampling frequency (30 Hz), precluding certain sampling frequencies from analysis (e.g., 20 Hz and 25 Hz). Alternative resampling functions use a less systematic combination of both interpolation (up-sampling) and decimation, but such methods reflect estimations, rather than real data, and are not representative of data that would be produced from a tag with a lower sampling rate.

Static acceleration, representing body posture in relation to Earth's gravitational field, and dynamic acceleration, representing body movement, were separated in all three acceleration axes (X, Y, and Z). This was executed for each sampling frequency, using a 3-s box smoother (Shepard et al., 2008a). Overall dynamic body acceleration (ODBA) was calculated by summing the absolute values of dynamic acceleration in each axis (Wilson et al., 2006; Shepard et al., 2008a). Continuous wavelet transformation in Ethographer v2.0 was used to derive acceleration signal waveform amplitude and frequency of the dominant cycle (the sway [Z] axis, representing lateral acceleration as a result of tail-beats) (Sakamoto et al., 2009). Descriptive statistics were extracted as per Brewster et al. (2018) and used as predictor variables ($n = 44$; Table 1). These predictor variables included the static and dynamic acceleration from each axis and their derivatives, including ODBA, waveform amplitude and frequency of the dominant cycle for each sampling frequency. To enable time matching of the ground-truthed data to observed behaviours recorded on a per-second basis, these acceleration-derived predictors were calculated for one-second segments. This procedure was repeated for each sampling frequency to create a set of predictor variables for each dataset.

2.3. Machine learning classification algorithm

Random forest (RF) is a supervised ML algorithm that has been used to classify behaviour from acceleration data (Nathan et al., 2012; Wang et al., 2015; Sur et al., 2017; Walton et al., 2018). For this reason, and ease of execution, we selected this method to assess the effect of sampling frequency on classifier performance. RF algorithms are an ensemble classifier whereby multiple unpruned classification or

Table 1

Predictor variables extracted from each of the acceleration axes for one second segments, used to train the RF ML algorithm for each sampling frequency.

Predictor Variable	Label	Definition
1. Static acceleration	Avg_X_Static, avg_Y_Static, avg_Z_Static	1 s means for static acceleration representing body posture in each axis
2. Dynamic acceleration	Avg_X_Dynamic, avg_Y_Dynamic, avg_Z_Dynamic	1 s means for dynamic acceleration representing body movement in each axis
3. Overall Dynamic Body Acceleration	Avg_ODBA	Sum of absolute dynamic body acceleration in each axis
4. Standard deviation	Sdev_X_Static, sdev_Y_Static, sdev_Z_Static, sdev_X_Dynamic, sdev_Y_Dynamic, sdev_Z_Dynamic, sdev_ODBA	Standard deviation of static and dynamic acceleration in each axis and ODBA
5. Minimum	Min_X_Static, min_Y_Static, min_Z_Static, min_X_Dynamic, min_Y_Dynamic, min_Z_Dynamic, min_ODBA	Minimum per 1 second values for static and dynamic acceleration in each axis and ODBA
6. Maximum	Max_X_Static, max_Y_Static, max_Z_Static, max_X_Dynamic, max_Y_Dynamic, max_Z_Dynamic, max_ODBA	Maximum per 1 second values for static and dynamic acceleration in each axis and ODBA
7. Kurtosis	Kurt_X_Static, kurt_Y_Static, kurt_Z_Static, kurt_X_Dynamic, kurt_Y_Dynamic, kurt_Z_Dynamic, kurt_ODBA	Measure of weight of tailedness relative to normal distribution for static and dynamic acceleration in each axis and ODBA
8. Skewness	Skew_X_Static, skew_Y_Static, skew_Z_Static, skew_X_Dynamic, skew_Y_Dynamic, skew_Z_Dynamic, skew_ODBA	Measure of asymmetry about mean for static and dynamic acceleration in each axis and ODBA
9. Cycle	Cycle	Cycle for the dominant frequency obtained through the continuous wavelet transformation generated spectrogram
10. Amplitude	Amp	Amplitude for the dominant frequency obtained through the continuous wavelet transformation generated spectrogram.

regression trees are grown (*ntree* as set by the user). They incorporate three steps: first, for each tree the data are randomly bootstrapped (with replacement) so that 63% of the data are used to train the tree; second, a random subset of predictor variables (*mtry*) are used to split the bootstrapped data at each node and the tree is grown to its full extent; third, the predictions from each tree are aggregated and the observation is assigned to the class with the majority vote. The data remaining from the internal bootstrapping are used by the model for cross-validation to calculate classification errors or Out-Of-Bag (OOB) errors (Breiman, 2001).

RF classification was conducted using the ‘*randomForest*’ package in R version 1.1.383 (Liaw and Wiener, 2002). For each sampling frequency, the predictor variables extracted from the ground-truthed data were randomly split into training and testing sets while maintaining class ratios (Ladds et al., 2017; Sur et al., 2017). Seventy per cent of the data were used for training the RF model, and the classification accuracy was cross-validated and tested on the remaining 30% of the data. Whilst data were randomly allocated to a train or test set, observations were kept in the same set across sampling frequencies to allow for direct comparison of results. A range of *ntree* values (number of trees) were tested (500, 1000 and 1500). Although OOB errors varied, the range was small so the *ntree* value resulting in least computational time was selected for the model (*ntree* = 1000). The number of predictor variables at each tree node, *mtry*, was chosen by using the square root of the total number of predictor variables ($n = 44$) (Hastie et al., 2009).

The behavioural classes for this data set were imbalanced, with two behavioural classes (rest and swim) accounting for > 98% of the ground-truthed data (majority classes), while burst, headshake, and chafe behaviours made up < 2% of the data (minority classes). RF models aim to improve overall classification accuracy by focusing on the predictive power of the majority class, to the detriment of the accuracy for minority classes (Chen et al., 2004; Ganganwar, 2012). Suggestions for dealing with imbalanced data sets include assigning class weights, and random or direct over- or under-sampling methods (Japkowicz, 2000). Some studies have created completely balanced data sets where each behaviour is represented by an equal number of observations (le Roux et al., 2017). However, due to both the nature of the data set and the functionality of the R package used, these options were not available for addressing our class imbalance. In this case, a stratified subset of the training data was incorporated into the RF model. The majority classes were reduced in frequency to be closer to the minority class (burst) by a factor of ten. Despite some loss of training data from the majority classes, this method drastically reduces

computational time (Chen et al., 2004).

2.4. Evaluating classifier performance

The trained RF model was applied to the unseen test data set (30%) to produce predictions and evaluate model performance. Evaluation metrics—accuracy, precision, recall and F_1 (described below)— were calculated from a confusion matrix in the ‘*caret*’ package in R (Kuhn, 2015). In this confusion matrix, rows are actual observed values and columns are predicted values, represented by true positive (TP), false positive (FP), and false negative (FN) values (Breiman, 1999, 2001). TP values occur when the predicted behavioural class has been correctly identified. Conversely, FP values are those which have been incorrectly attributed to a behavioural class. FN predictions are observations which have been incorrectly assigned to a different class. These values were used to calculate the evaluation metrics as follows (Breiman (2002):

Recall: The proportion of predictions of a behaviour class that are correctly classified as that behaviour (completeness).

$$R = TP/(TP + FN) \quad (1)$$

Precision: The proportion of correctly classified behaviours from all predictions (exactness). Precision can be poor if many predictions are incorrectly assigned to a behavioural class.

$$P = TP/(TP + FP) \quad (2)$$

F_1 : The harmonic mean of recall and precision with a value of 0–1. Values near 0 have low classification performance whilst values closest to 1 have the best classification performance.

$$F_1 = 2PR/(P + R) \quad (3)$$

F_M (Macro F_1): The mean of F_1 scores for all classes; used to describe overall classifier performance for each sampling frequency

$$F_M = \Sigma^M F_1/M \quad (4)$$

where M is the number of classes in the classification model.

Accuracy: The overall proportion of behavioural classes predicted correctly.

$$(TP + TN)/(TP + TN + FP + FN) \quad (5)$$

To allow for comparison of classifier performance between each of the individual behaviour classes, recall, precision and an F_1 score were calculated within each RF. Accuracy and the macro F_1 score (F_M), allow comparison of overall classifier performance between sampling frequencies, encompassing all behaviour classes. These methods (Fig. 1)

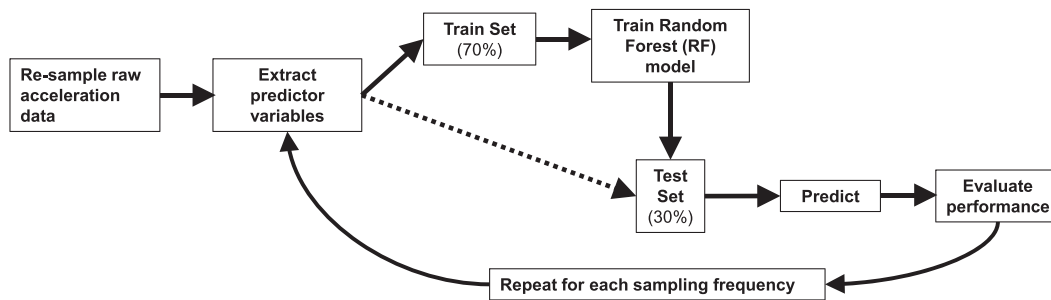


Fig. 1. Machine learning workflow for random forest classification of behaviours observed in semi-captive juvenile lemons sharks (*N. brevirostris*) ($n = 4$). This workflow was repeated for each sampling frequency to allow for direct comparison of classifier performance in predicting behaviour from unseen test data between the six sampling frequencies.

were repeated for each sampling frequency to allow for direct comparison in classification performance as sampling frequency was reduced from 30 Hz to 1 Hz.

3. Results

Accelerometer data were collected for four lemon sharks during captive trials (Brewster et al., 2018). From these data sets, over 35,000 s of data were ground-truthed and labelled as one of five distinct behaviours: swim, rest, burst, chafe, and headshake (Brewster et al., 2018). Swimming and resting behaviours were performed most frequently (97.57 and 1.23%, Table 2), whilst burst, chafe and headshake behaviours were infrequent in comparison (0.13, 0.76 and 0.31% respectively, Table 2).

3.1. Acceleration characteristics and signal aliasing

Upon visual inspection of the acceleration waveform signal output for the five behaviours observed, bouts of constant amplitude and frequency for dynamic sway acceleration were indicative of swimming and resting, with higher positive and negative acceleration (g) values for swimming. Rest behaviour, defined as motionless behaviour, corresponded to acceleration values of near-zero. Acceleration waveform signal amplitude and frequency was greater for chafe, burst and headshake behaviours than for swimming, with headshaking typically displaying both higher frequency and amplitude for longer durations than bursting and chafing (Fig. 2). This visual inspection revealed obvious differences in waveform characteristics for each behaviour; however, the amplitude and frequency of the waveform signal was distorted, with behaviours becoming less visually discernible from each other as sampling frequency was reduced from 30 Hz to 1 Hz (Fig. 2).

3.2. Classifier performance

Regardless of sampling frequency used, overall model accuracy was high (> 96%). F_M was highest at 30 Hz, however little decrease in overall predictive power was observed until sampling frequency was reduced to below 5 Hz (Fig. 3). Classifier performance for the individual

behavioural classes varied (Fig. 4; Table 3).

Swim: The correct classification of events from the swim class was highest for all sampling frequencies. As sampling frequency decreased, precision (exactness) remained high (Fig. A), however, recall (completeness) decreased slightly when sampling frequency was < 5 Hz (Fig. 4B). This was due to swim events being increasingly labelled as other behaviours (Table 3).

Rest: The model classified rest events with high scores in all performance metrics until sampling frequency was < 5 Hz (Fig. 4). Recall remained high across all sampling frequencies, however precision suffered < 3 Hz due to incorrect predictions of rest as swim events (Table 3).

Chafe: Chafe events were classified with high recall across all sampling frequencies, however this was at the expense of decreasing precision (Fig. 4A and B). The model increasingly mislabelled chafe events as swim events as sampling frequency decreased (Table 3).

Burst: Burst behaviour obtained high precision scores when sampling frequency was ≥ 10 Hz (Fig. 4A). Precision scores decreased at < 10 Hz due to the model mistakenly labelling swim events as burst (Table 3). However, due to incorrect predictions of chafe and headshake events from burst behaviours, recall was low for burst events across all sampling frequencies (Table 3, Fig. 4B).

Headshake: Headshakes achieved the lowest classification performance of all behaviour classes, across all sampling frequencies, with the highest F_1 score obtained at 30 Hz (Fig. 4C). Model recall was at the expense of increasingly poor precision (Fig. 4A and B). Decreasing recall was due to events from the headshake class incorrectly predicted as chafe events, whilst low precision was a result of all classes (except rest) being wrongly labelled as headshakes across all sampling frequencies, with the most common FP attributed to the swim class (Table 3).

Classification performance for the minority classes (e.g., headshake, chafe and burst) was lower in all three metrics (precision, recall and F_1) than for the majority classes (swim and rest) (Fig. 4A, B and C). Performance for the minority classes was not significantly improved by reducing the majority classes compared to using the original imbalanced training dataset, however, the reduction in computational time was significant.

Table 2

Ethogram and percent time spent performing observed behaviours for accelerometer equipped semi-captive juvenile lemon sharks (*N. brevirostris*) ($n = 4$) These per-second observations were used to validate the raw accelerometer data.

Shark ID	Behaviour										Total	%
	Burst	%	Chafe	%	Headshake	%	Rest	%	Swim	%		
1	4	0.01	139	0.47	57	0.19	82	0.28	29,406	99.05	29,688	82.75
2	10	0.18	83	1.46	14	0.25	315	5.53	5273	92.59	5695	15.87
3	12	4.32	49	17.63	26	9.35	43	15.47	148	53.24	278	0.77
4	22	10.28	0	0.00	16	7.48	0	0.00	176	82.24	214	0.60
Total	48	0.13	271	0.76	113	0.31	440	1.23	35,003	97.57	35,875	100.00

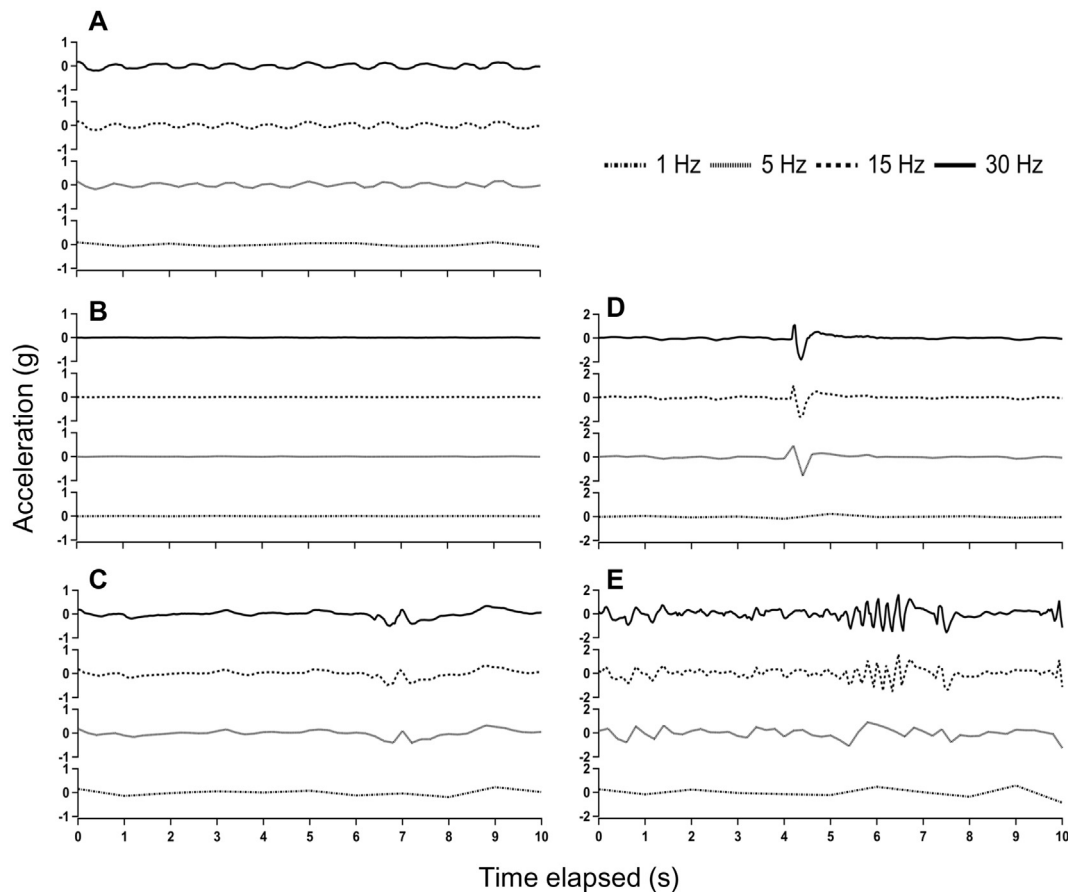


Fig. 2. Example time-series plots for dynamic sway (Z) acceleration for observed behaviours for semi-captive juvenile lemon sharks (*N. brevirostris*) ($n = 4$). Raw acceleration data (30 Hz) were resampled to show the representative change in acceleration waveform signal amplitude and frequency as sampling frequency was reduced for five observed behaviours (A) swim, (B) rest, (C) chafe, (D) burst and (E) headshake. Note different scales for acceleration (g) to aid visualisation of behaviours.

4. Discussion

This study aimed to demonstrate how sampling frequency affects the ability of a machine learning (ML) algorithm to correctly classify behaviours from accelerometer data. Accuracy was high ($> 96\%$) for all sampling frequencies, reflecting the effect of an imbalanced data set due to the nature of RF focusing on the predictive power of the majority class. For this reason, F_M was considered the most appropriate overall evaluation metric to compare sampling frequencies, as this metric combines F_1 score for all classes of behaviour to give a single value describing classifier performance across all behaviour classes.

Our research indicated that sampling frequencies at or above 5 Hz are adequate for the classification of behaviour from similar accelerometer data. Overall, classification of lemon shark behaviours from accelerometer data was best achieved at the highest frequency available, 30 Hz ($F_M > 0.790$), which is comparable to results in sheep (32 Hz; Walton et al., 2018). However, evaluation of sampling frequency across all behaviours revealed that this performance did not decrease substantially if sampling frequencies remained at or above 5 Hz.

4.1. Classifier performance for individual behaviour classes

Although our results suggest that overall classifier performance decreases as sampling frequency decreases, we found that the effect of sampling frequency on this performance varied according to individual behaviour classes. Interpreting classification performance with regard to distinct behaviour classes reveals that classifier performance is

inherently dependent upon the complexity of the behaviour to be classified (McClune et al., 2014). Appropriate sampling frequency is scale dependent, with regard to body size and the frequency of body movement of interest (Nyquist criterion). Whilst overall classification performance decreases as sampling frequency decreases below 5 Hz, the model classified basic swim and rest behaviours with high performance at all sampling frequencies ($F_1 > 0.900$ above 1 Hz). Successful differentiation of active and inactive (resting) behaviours has been achieved with a range of sampling frequencies, including 25 Hz for smaller (< 44.2 cm FL) teleost bonefish (*Albula vulpes*) (Brownscombe et al., 2014). For larger elasmobranchs (> 1 m) considerably lower sampling frequencies have enabled successful identification of swimming and resting behaviours from accelerometer data, including 1 Hz for white tip reef sharks (*Triaenodon obesus*) (Whitney et al., 2007).

In contrast to the classification of swim and rest behaviours, the RF model struggled to classify behaviours characterised by faster kinematics of movement (higher acceleration frequency and amplitude; herein referred to as complex behaviours) across the entire range of sampling frequencies. This trend has been mirrored in previous studies, where complex flight behaviours for golden eagles (*Aquila chrysaetos*) were poorly distinguished using RF classification from accelerometer data collected at 140 Hz (Sur et al., 2017). Additionally, complex behaviours (fast-start and feeding) in great sculpin (*Myoxocephalus polyacanthocephalus*) were classified with low overall accuracy (60%) at sampling frequencies of 20–30 Hz, suggesting that higher sampling frequencies are required for successful classification of these behaviours (Broell et al., 2013). Similarly, higher sampling frequencies of 200 Hz were required for successful identification ($F_1 > 0.7$) of feeding

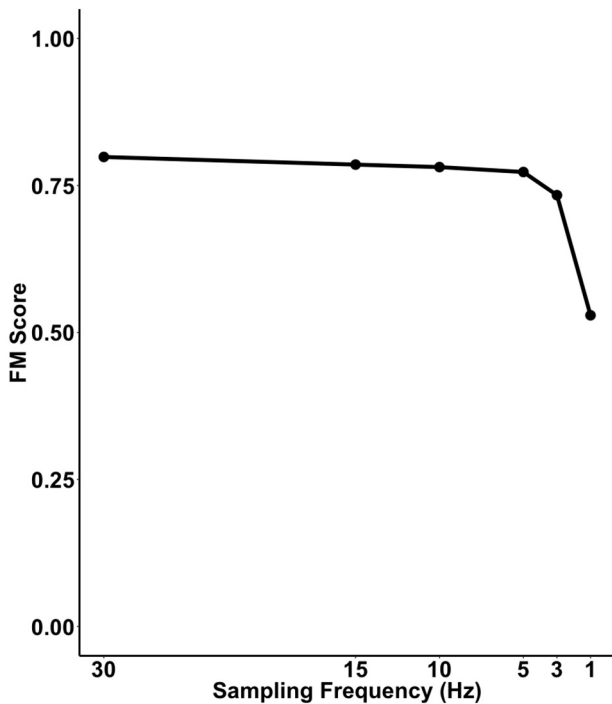


Fig. 3. Overall random forest classification performance for classification of behaviours in juvenile lemon sharks (*N. brevirostris*) ($n = 4$). As sampling frequency decreases from 30 to 1 Hz, overall classification performance (described by macro averaged F₁ Score [F_M]) decreases. Performance decreases more significantly when sampling frequency is reduced to < 5 Hz.

behaviours in red-spotted groupers (*Epinephelus akaara*) (Horie et al., 2017).

There are a number of reasons for the misclassification of complex lemon shark behaviours that involve faster kinematics. Typically, swimming in sharks is represented by regular oscillatory patterns in the sway acceleration waveform signal, with each oscillation representing the individual tail-beats contributing to forward propulsion (Gleiss et al., 2009a; Gleiss et al., 2009b). Any similarities between waveform signal characteristics may contribute to misclassification (McClune et al., 2014; Ladds et al., 2017; le Roux et al., 2017; Walton et al., 2018). For example, although bout-duration varied, complex behaviours in juvenile lemon sharks were similarly characterised by high frequency and amplitude acceleration signals and produce similar per-second predictor variables. This led to the model's incorrect predictions of headshake events from all behaviour classes except rest, a behaviour that does not feature any distinct tail-beat movements and therefore bears little resemblance to active behaviours. Misclassification was also high between behaviours with similar acceleration signal characteristics in cats (*Felis catus*) (Watanabe et al., 2005), sheep (*Ovis aries*) (le Roux et al., 2017; Walton et al., 2018) and seals (*Arctocephalus sp.*) and sea lions (*Neophoca cinerea*) (Ladds et al., 2017).

Additionally, misclassification can be partially attributed to waveform signal aliasing, a distortion effect that causes behaviours to become indistinguishable from each other when sampling frequency is decreased beyond the Nyquist frequency (Beutler, 1966; Mallat, 1999; Broell et al., 2013). Signal aliasing was evident for all observed behaviours as sampling frequency was reduced, though it was most noticeable for burst and headshake behaviours, as these behaviours are characterised by the highest frequency movements. At lower sampling frequencies these behaviours became indiscernible from each other, causing further misclassification.

Whilst not directly related to the effects of sampling frequency,

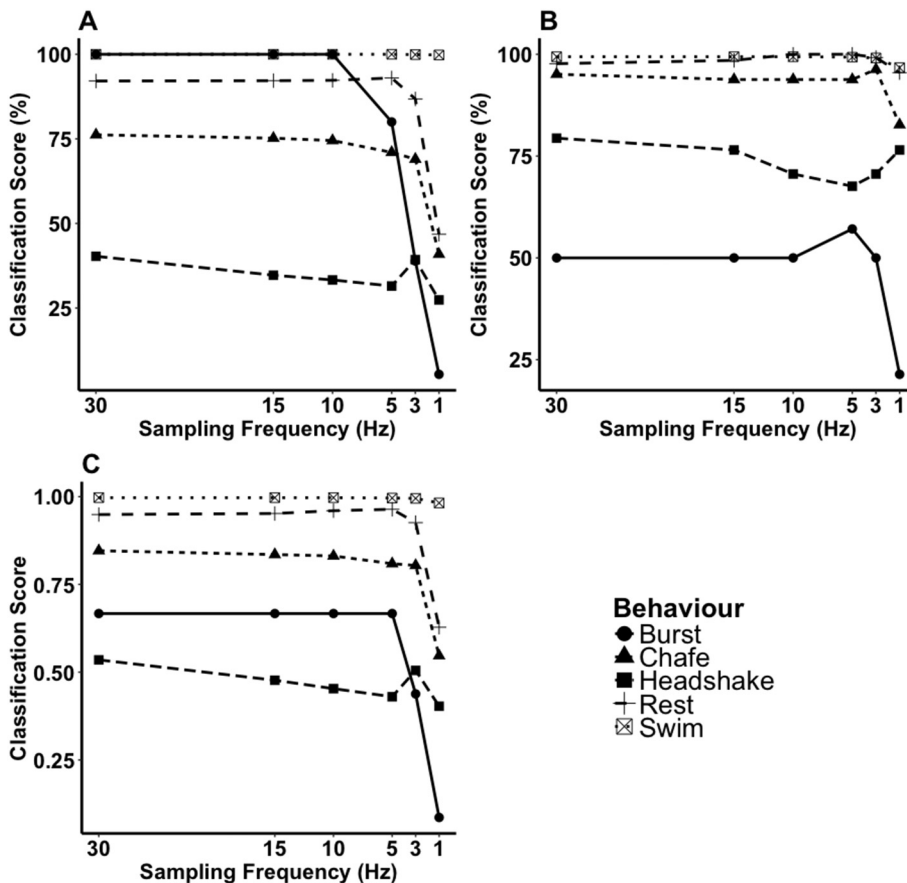


Fig. 4. Random forest performance for classification of individual behaviours observed from semi-captive trials for juvenile lemon sharks (*N. brevirostris*) ($n = 4$) at different sampling frequencies (30, 15, 10, 5, 3 and 1 Hz). For each behaviour class (burst, chafe, headshake, rest, and swim), performance is described by three metrics: (A) Precision, (B) Recall and (C) F₁ Score (as described in text). Performance varied for each behavioural class, with swim (open squares) and rest (crosses) classes achieving higher classification scores in all metrics than chafe (triangles), burst (black circles) and headshake (black squares) behaviour classes.

Table 3

Confusion matrix of random forest model generated predictions from the test set of ground-truthed data for each sampling frequency. Rows indicate actual behaviour observations and columns represent model predicted behaviours. Bold values show observations correctly classified by the model (True Positives; TP).

		Predicted Behaviour					
		Class	Burst	Chafe	Headshake	Rest	Swim
Actual Behaviour	30 Hz	Burst	7	3	4	0	0
		Chafe	0	77	2	0	2
		Headshake	0	7	27	0	0
		Rest	0	0	0	129	3
	15 Hz	Swim	0	14	34	11	10442
		Burst	7	3	4	0	0
		Chafe	0	76	2	0	3
		Headshake	0	8	26	0	0
	10 Hz	Rest	0	0	0	130	2
		Swim	0	14	43	11	10433
		Burst	7	3	3	0	1
		Chafe	0	76	2	0	3
	5 Hz	Headshake	0	9	24	0	1
		Rest	0	0	0	132	0
		Swim	0	14	43	11	10433
		Burst	8	4	1	0	1
	3 Hz	Chafe	0	76	2	0	3
		Headshake	0	10	23	0	1
		Rest	0	0	0	132	0
		Swim	2	17	47	10	10425
	1 Hz	Burst	7	4	1	0	2
		Chafe	1	78	0	0	2
		Headshake	0	5	24	0	5
		Rest	0	0	0	131	1
	1 Hz	Swim	10	26	36	20	10409
		Burst	3	5	3	0	3
		Chafe	0	67	1	0	13
		Headshake	3	2	26	0	3
		Rest	0	0	0	126	6
		Swim	50	90	65	143	10153

misclassification can also be linked to the rarity of burst and headshake events for training and testing the RF model (Shamoune-Baranes et al., 2012; Brown et al., 2013). While little loss of overall predictive power was observed when classifying behaviours in pumas (*Puma concolor*) until sampling frequency was reduced to below 8 Hz, classification of feeding and grooming behaviours was less effective, with small class size recognised as the main reason for misclassification (Wang et al., 2015). Although class size imbalance was managed in this study by selectively down-sampling the majority behaviour classes, there were still a limited number of observations for the rare behaviour classes for training and testing the model. Therefore, the effect of just one mislabelled data point has a larger effect on the evaluation metrics for these behaviour classes than for the classes which contain many thousands of observations (e.g., swim). The small class sizes for these behaviours were a direct result of the logistical difficulties associated with obtaining data for infrequent behaviours from captive lemon sharks (Brewster et al. (2018)).

Misclassification may be also attributed to inter-individual variability within observed behaviours (Walker et al., 2015). High variability was demonstrated for swimming in captive lemon sharks by Gleiss et al. (2009a), where tail-beat frequency ranged from 0.4–1.2 Hz and tail-beat acceleration amplitude ranged from 0.002–0.16 g. Although these kinematic values for each behaviour were not directly assessed in this study, this variation is presumably larger for behaviours such as burst or headshaking that involve more complex kinematics. Whilst this variability was incorporated into the training set for the majority classes (e.g., swim), any extreme variability for the minority classes (burst and headshake) may have a larger effect, causing further misclassification of rare behaviours.

4.2. Recommendations and implications

The ability to remotely classify behaviours and study activity levels

related to foraging behaviour could have a positive impact on future conservation and management strategies for many species. Although misclassification of the fine-scale behaviours related to foraging (i.e., headshaking; Brewster et al., 2018) was disappointing given their ecological importance, our objective here was simply to demonstrate the effects of reducing sampling frequency on classification. There are several approaches that might improve the classification of such behaviours. Poor precision in the headshake class may have masked meaningful results that may have been better detected by using a different classification model, limiting application of the RF model at any sampling frequency to wild data. The ensemble classifier (composed of a weighted combination of multiple base classification models) developed from this data set by Brewster et al. (2018) would improve the classification of these behaviours across all sampling frequencies. For future studies, there are a number of methodological improvements that could be incorporated to improve classification of complex behaviours. More captive trials may increase the number of observations for rare events (Wang et al., 2015). Alternatively, grouping under-represented behaviours together within the model may improve their classification (Jeantet et al., 2018), however this may narrow ecological insight. Classification may also be improved by using a different study species that better proportions behaviours; lemon sharks are ectotherms, therefore feed rarely compared to smaller endotherms.

Additionally, classification performance can be attributed to complex interactions between predictor variables (Liaw and Wiener, 2002), and shark behaviours may be identified by features of acceleration data besides the changes in dynamic acceleration corresponding to tail-beats. For example, mating in nurse sharks was characterised in part by changes in body orientation, represented by static acceleration in the pitch and sway axes (Whitney et al., 2010). Therefore, changes in body orientation, including the rolling body movements displayed by juvenile lemon sharks during chafe events, may be better described by incorporating data collected from additional movement sensors, such as

gyroscopes and magnetometers (Noda et al., 2012; Noda et al., 2013; Kawabata et al., 2014; Noda et al., 2014; Williams et al., 2015; Williams et al., 2017). Increasing the number of predictor variables by including metrics related to the individual shark and environmental factors may also improve both model complexity and classification of fine-scale behaviours (Huynh et al., 2007; Nathan et al., 2012; Brownscombe et al., 2014).

While classification was poor for chafe, burst and headshake behaviours, there was no significant decrease in classifier performance for these behaviours until sampling frequency was < 5 Hz. Therefore, for best determination of both basic and fine-scale fast movement behaviours in animals of similar size and kinematics, accelerometers should be programmed at a sampling frequency of 5 Hz. These findings have major positive implications for the practical aspects of future studies classifying behaviours from accelerometer data. Programming accelerometer devices at the lowest frequency possible could drastically reduce the rate at which the available memory and battery capacity of devices is consumed. For this study, the total available memory capacity of the data logger used was 56 MB, and maximum recording duration at 30 Hz was ≈5 days. Based on our recommendation to programme devices with a sampling frequency of 5 Hz, study durations using this device could last as long as 30 days (depending on digital storage programming e.g., number of bits) without sacrificing classification performance (see Supplementary Material, Fig. S1). With regards to reducing device battery consumption, there can be a trade-off between improving classification performance and extending device battery life. Battery life was halved when sampling frequencies were increased from 16Hz to 32Hz for human activity (Khan et al., 2016), yet the same increase in sampling frequency only achieved a 5% increase in classification accuracy in sheep (Walton et al., 2018). The minimal improvement in classifier performance at higher sampling frequencies is therefore not worth the increased battery consumption, further demonstrating the logistical benefits of programming devices with 5 Hz.

5. Conclusion

This study demonstrates how sampling frequency influences behavioural classification of accelerometer data, enabling future refinements in the programming of accelerometer devices for classification in animals of similar size and kinematics of movement. Sampling frequencies as low as 5 Hz are suitable for classifying behaviours in addition to dramatically reducing demand on archival device memory and battery. The benefits of lengthening study duration include extending insight to ecologically meaningful time scales (e.g., tidal, lunar), reduced study costs and tagging fewer animals.

Funding

J. L. Hounslow's honours work was supported by Save our Seas Foundation [D'Arros Research Centre internal grant] and the School of Veterinary & Life Sciences, Murdoch University, Murdoch, WA. BBFS work was funded by the National Science Foundation [grants number OCE 1156145 and OCE 1156141] to ACG and NMW.

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical approval

All applicable international, national, and/or institutional guidelines for the care and use of animals were followed. All procedures performed in studies involving animals were in accordance with the ethical standards of the institution or practice at which the studies were conducted (University of Miami Institutional Animal Care and Use Committee (IACUC), Protocol Number 12-030).

Acknowledgments

We thank all the staff, volunteers and sponsors of Bimini Biological Field Station Foundation (BBFSF), in particular Dr. S.H. Gruber, who enabled collection of the dataset used for this study and the Department of Fisheries of the Commonwealth of the Bahamas for granting BBFSF a Permit to Conduct Scientific Research (MAF/LIA/22).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jembe.2018.12.003>.

References

- Abrahms, B., Jordan, N., Golabek, K., McNutt, J., Wilson, A., Brashares, J., 2016. Lessons from integrating behaviour and resource selection: activity-specific responses of African wild dogs to roads. *Anim. Conserv.* 19, 247–255. <https://doi.org/10.1111/acv.12235>.
- Beutler, F.J., 1966. Error-free recovery of signals from irregularly spaced samples. *SIAM Rev.* 8, 328–335. <https://doi.org/10.1137/1008065>.
- Bidder, O.R., Campbell, H.A., Gomez-Laich, A., Urge, P., Walker, J., Cai, Y., Gao, L., Quintana, F., Wilson, R.P., 2014. Love thy neighbour: automatic animal behavioural classification of acceleration data using the K-nearest neighbour algorithm. *PLoS ONE* 9, e88609. <https://doi.org/10.1371/journal.pone.0088609>.
- Breiman, L., 1999. *Random Forests*. UC Berkeley.
- Breiman, L., 2001. *Random Forests*. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/a:1010933404324>.
- Breiman, L., 2002. *Manual on setting up, using, and understanding random forests v3*. Vol. 1. Statistics Department University of California Berkeley, CA, USA, pp. 1.
- Brewster, L.R., Dale, J.J., Guttridge, T.L., Gruber, S.H., Hansell, A.C., Elliott, M., Cowx, I.G., Whitney, N.M., Gleiss, A.C., 2018. Development and application of a machine learning algorithm for classification of elasmobranch behaviour from accelerometry data. *Mar. Biol.* 165 (62). <https://doi.org/10.1007/s00227-018-3318-y>.
- Broell, F., Noda, T., Wright, S., Domenici, P., Steffensen, J.F., Auclair, J.-P., Taggart, C.T., 2013. Accelerometer tags: detecting and identifying activities in fish and the effect of sampling frequency. *J. Exp. Biol.* 216, 1255–1264. <https://doi.org/10.1242/jeb.077396>.
- Brown, D.D., Kays, R., Wikelski, M., Wilson, R.P., Klimley, A.P., 2013. Observing the unwatchable through acceleration logging of animal behavior. *Animal Biotelemetry* 1, 20. <https://doi.org/10.1186/2050-3385-1-20>.
- Brownscombe, J.W., Gutowsky, L.F., Danylchuk, A.J., Cooke, S.J., 2014. Foraging behaviour and activity of a marine benthivorous fish estimated using tri-axial accelerometer biologgers. *Mar. Ecol. Prog. Ser.* 505, 241–251. <https://doi.org/10.3354/meps10786>.
- Chen, K.Y., Bassett Jr., D.R., 2005. The technology of accelerometry-based activity monitors: current and future. *Med. Sci. Sports Exerc.* 37, S490–S500. <https://doi.org/10.1249/01.mss.0000185571.49104.82>.
- Chen, C., Liaw, A., Breiman, L., 2004. Using random forest to learn imbalanced data. In: *Technical Report 666*. 110. University of California, Berkeley, pp. 1–12.
- Cooke, S.J., 2008. Biotelemetry and biologging in endangered species research and animal conservation: relevance to regional, national, and IUCN Red list threat assessments. *Endanger. Species Res.* 4, 165–185. <https://doi.org/10.3354/esr00063>.
- Ganganwar, V., 2012. *An overview of classification algorithms for imbalanced datasets*. *Int. J. Emerg. Technol. Adv. Eng.* 2, 42–47.
- Gao, L., Campbell, H.A., Bidder, O.R., Hunter, J., 2013. A Web-based semantic tagging and activity recognition system for species' accelerometry data. *Ecological Informatics* 13, 47–56. <https://doi.org/10.1016/j.ecoinf.2012.09.003>.
- Gleiss, A.C., Gruber, S.H., Wilson, R.P., 2009a. Multi-channel data-logging: towards determination of behaviour and metabolic rate in free-swimming sharks. In: Nielsen, J.L., Fragoso, N., Hobday, A., Lutcaugh, M., Sibert, J. (Eds.), *Tagging and Tracking of Marine Animals with Electronic Devices*. Reviews: Methods and Technologies in Fish Biology and Fisheries. Springer, Dordrecht, pp. 211–228.
- Gleiss, A.C., Norman, B., Liebsch, N., Francis, C., Wilson, R.P., 2009b. A new prospect for tagging large free-swimming sharks with motion-sensitive data-loggers. *Fish. Res.* 97, 11–16. <https://doi.org/10.1016/j.fishres.2008.12.012>.
- Graf, P.M., Wilson, R.P., Qasem, L., Hackländer, K., Rosell, F., 2015. The use of acceleration to code for animal behaviours; a case study in free-ranging Eurasian beavers *Castor fiber*. *PLoS ONE* 10, e0136751. <https://doi.org/10.1371/journal.pone.0136751>.
- Grünewälder, S., Broekhuis, F., MacDonald, D.W., Wilson, A.M., McNutt, J.W., Shawe-Taylor, J., Hailes, S., 2012. Movement activity based classification of animal behaviour with an application to data from cheetah (*Acinonyx jubatus*). *PLoS ONE* 7, e49120. <https://doi.org/10.1371/journal.pone.0049120>.
- Halsey, L., Green, J., Wilson, R., Frappell, P., 2008. Accelerometry to estimate energy expenditure during activity: best practice with data loggers. *Physiol. Biochem. Zool.* 82, 396–404. <https://doi.org/10.1086/589815>.
- Halsey, L.G., Shepard, E.L., Wilson, R.P., 2011. Assessing the development and application of the accelerometry technique for estimating energy expenditure. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* 158, 305–314. <https://doi.org/10.1016/j.cbpa.2010.09.002>.

- Hammond, T.T., Springthorpe, D., Walsh, R.E., Berg-Kirkpatrick, T., 2016. Using accelerometers to remotely and automatically characterize behavior in small animals. *J. Exp. Biol.* 219, 1618–1624. <https://doi.org/10.1242/jeb.136135>.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edn. Springer-Verlag: New York, NY, USA.
- Hays, G.C., Ferreira, L.C., Sequeira, A.M., Meehan, M.G., Duarte, C.M., Bailey, H., Baillieu, F., Bowen, W.D., Caley, M.J., Costa, D.P., 2016. Key questions in marine megafauna movement ecology. *Trends Ecol. Evol.* 31, 463–475.
- Horie, J., Mitamura, H., Ina, Y., Mashino, Y., Noda, T., Moriya, K., Arai, N., Sasakura, T., 2017. Development of a method for classifying and transmitting high-resolution feeding behavior of fish using an acceleration pinger. *Animal Biotelemetry* 5 (12). <https://doi.org/10.1186/s40317-017-0127-x>.
- Huynh, T., Blanke, U., Schiele, B., 2007. Scalable recognition of daily activities with wearable sensors. In: *LoCA*, pp. 50–67.
- Japkowicz, N., 2000. *Learning from Imbalanced Data Sets: A Comparison of Various Strategies*. AAAI workshop on learning from imbalanced data sets, Menlo Park CA, pp. 10–15.
- Jeanet, L., Dell'Amico, F., Forin-Wiart, M., Coutant, M., Bonola, M., Etienne, D., Gresser, J., Regis, S., Lecerf, N., Lefebvre, F., 2018. Combined use of two supervised learning algorithms to model sea turtle behaviours from tri-axial acceleration data. *J. Exp. Biol.* 177378. <https://doi.org/10.1242/jeb.177378>.
- Kawabata, Y., Noda, T., Nakashima, Y., Nanami, A., Sato, T., Takebe, T., Mitamura, H., Arai, N., Yamaguchi, T., Soyano, K., 2014. Use of a gyroscope/accelerometer data logger to identify alternative feeding behaviours in fish. *J. Exp. Biol.* 217, 3204–3208. <https://doi.org/10.1242/jeb.108001>.
- Khan, A., Hammerla, N., Mellor, S., Plötz, T., 2016. Optimising sampling rates for accelerometer-based human activity recognition. *Pattern Recogn. Lett.* 73, 33–40.
- Kuhn, M., 2015. A short introduction to the caret package. <https://cran.r-project.org/web/packages/caret/vignettes/caret.pdf> (Accessed 07 January 2018).
- Ladds, M.A., Thompson, A.P., Kadar, J.-P., Slip, D., Hocking, D., Harcourt, R., 2017. Super machine learning: improving accuracy and reducing variance of behaviour classification from accelerometry. *Animal Biotelemetry* 5 (8). <https://doi.org/10.1186/s40317-017-0123-1>.
- le Roux, S.P., Marias, J., Wolhuter, R., Niesler, T., 2017. Animal-borne behaviour classification for sheep (Dohne Merino) and Rhinoceros (Ceratotherium simum and Diceros bicornis). *Animal Biotelemetry* 5 (25). <https://doi.org/10.1186/s40317-017-0140-0>.
- Leos-Barajas, V., Photopoulou, T., Langrock, R., Patterson, T.A., Watanabe, Y.Y., Murgatroyd, M., Papastamatiou, Y.P., 2017. Analysis of animal accelerometer data using hidden Markov models. *Methods Ecol. Evol.* 8, 161–173. <https://doi.org/10.1111/2041-210X.12657>.
- Liaw, A., Wiener, M., 2002. Classification and Regression by random Forest. *R news*. Vol. 2, pp. 18–22.
- Mallat, S., 1999. *A Wavelet Tour of Signal Processing*. Academic press.
- McClune, D.W., Marks, N.J., Wilson, R.P., Houghton, J.D., Montgomery, I.W., McGowan, N.E., Gormley, E., Scantlebury, M., 2014. Tri-axial accelerometers quantify behaviour in the Eurasian badger (*Meles meles*): towards an automated interpretation of field data. *Animal Biotelemetry* 2, 5. <https://doi.org/10.1186/2050-3385-2-5>.
- Nathan, R., Spiegel, O., Fortmann-Roe, S., Harel, R., Wikelski, M., Getz, W.M., 2012. Using tri-axial acceleration data to identify behavioral modes of free-ranging animals: general concepts and tools illustrated for griffon vultures. *J. Exp. Biol.* 215, 986–996. <https://doi.org/10.1242/jeb.058602>.
- Noda, T., Okuyama, J., Koizumi, T., Arai, N., Kobayashi, M., 2012. Monitoring attitude and dynamic acceleration of free-moving aquatic animals using a gyroscope. *Aquat. Biol.* 16, 265–276. <https://doi.org/10.3354/ab00448>.
- Noda, T., Kawabata, Y., Arai, N., Mitamura, H., Watanabe, S., 2013. Monitoring escape and feeding behaviours of cruiser fish by inertial and magnetic sensors. *PLoS ONE* 8, e79392. <https://doi.org/10.1371/journal.pone.0079392>.
- Noda, T., Kawabata, Y., Arai, N., Mitamura, H., Watanabe, S., 2014. Animal-mounted gyroscope/accelerometer/magnetometer: in situ measurement of the movement performance of fast-start behaviour in fish. *J. Exp. Mar. Biol. Ecol.* 451, 55–68. <https://doi.org/10.1016/j.jembe.2013.10.031>.
- Payne, N.L., Taylor, M.D., Watanabe, Y.Y., Semmens, J.M., 2014. From physiology to physics: are we recognizing the flexibility of biologging tools. *J. Exp. Biol.* 217, 317–322. <https://doi.org/10.1242/jeb.09392>.
- Rutz, C., Hays, G.C., 2009. New frontiers in biologging science. pp. 289–292. <https://doi.org/10.1098/rsbl.2009.0089>.
- Sakamoto, K.Q., Sato, K., Ishizuka, M., Watanuki, Y., Takahashi, A., Daunt, F., Wanless, S., 2009. Can ethograms be automatically generated using body acceleration data from free-ranging birds? *PLoS ONE* 4, e5379. <https://doi.org/10.1371/journal.pone.0005379>.
- Shamoune-Baranes, J., Bom, R., van Loom, E.E., Ens, B.J., Oosterbeek, K., Bouten, W., 2012. From sensor data to animal behaviour: an oystercatcher example. *PLoS ONE* 7. <https://doi.org/10.1371/journal.pone.0037997>.
- Shepard, E.L.C., Wilson, R.P., Halsey, L.G., Quintana, F., Laich, A.G., Gleiss, A.C., Liebsch, N., Myers, A.E., Norman, B., 2008a. Derivation of body motion via appropriate smoothing of acceleration data. *Aquat. Biol.* 4, 235–241. <https://doi.org/10.3354/ab00104>.
- Shepard, E.L.C., Wilson, R.P., Quintana, F., Laich, A.G., Liebsch, N., Albareda, D.A., Halsey, L.G., Gleiss, A.C., Morgan, D.T., Myers, A.E., 2008b. Identification of animal movement patterns using tri-axial accelerometry. *Endanger. Species Res.* 10, 47–60. <https://doi.org/10.3354/esr00084>.
- Sur, M., Suffredini, T., Wessells, S.M., Bloom, P.H., Lanzone, M., Blackshire, S., Sridhar, S., Katzner, T., 2017. Improved supervised classification of accelerometry data to distinguish behaviors of soaring birds. *PLoS ONE* 12, e0174785. <https://doi.org/10.1371/journal.pone.0174785>.
- Sutherland, W.J., 1998. The importance of behavioural studies in conservation biology. *Anim. Behav.* 56, 801–809.
- Valletta, J.J., Torney, C., Kings, M., Thornton, A., Madden, J., 2017. Applications of machine learning in animal behaviour studies. *Anim. Behav.* 124, 203–220. <https://doi.org/10.1016/j.anbehav.2016.12.005>.
- Walker, J.S., Jones, M.W., Laramee, R.S., Holton, M.D., Shepard, E.L., Williams, H.J., Scantlebury, D.M., Nikki, J.M., Magowan, E.A., Maguire, I.E., 2015. Prying into the intimate secrets of animal lives; software beyond hardware for comprehensive annotation in 'Daily Diary' tags. *Movement Ecol.* 3, 29. <https://doi.org/10.3354/esr00064>.
- Walton, E., Casey, C., Mitsch, J., Vázquez-Diosdado, J.A., Yan, J., Dottorini, T., Ellis, K.A., Winterlich, A., Kaler, J., 2018. Evaluation of sampling frequency, window size and sensor position for classification of sheep behaviour. *R. Soc. Open Sci.* 5, 171442. <https://doi.org/10.6084/m9.figshare.c.3985326>.
- Wang, Y., Nickel, B., Rutishauser, M., Bryce, C.M., Williams, T.M., Elkaim, G., Wilmers, C.C., 2015. Movement, resting, and attack behaviors of wild pumas are revealed by tri-axial accelerometer measurements. *Movement Ecol.* 3, 2. <https://doi.org/10.1186/s40462-015-0030-0>.
- Watanabe, S., Izawa, M., Kato, A., Ropert-Coudert, Y., Naito, Y., 2005. A new technique for monitoring the detailed behaviour of terrestrial animals: a case study with the domestic cat. *Appl. Anim. Behav. Sci.* 94, 117–131. <https://doi.org/10.1016/j.applanim.2005.01.010>.
- Whitney, N.M., Papastamatiou, Y.P., Holland, K.N., Lowe, C.G., 2007. Use of an acceleration data logger to measure diel activity patterns in captive whitetip reef sharks, *Triaenodon obesus*. *Aquat. Living Resour.* 20, 299–305. <https://doi.org/10.1051/alr:2008006>.
- Whitney, N.M., Papastamatiou, Y.P., Gleiss, A.C., 2012. Integrative multisensor tagging: emerging techniques to link elasmobranch behavior, physiology, and ecology. In: Carrier, J.C., Musick, J.A., Heithaus, M.C. (Eds.), *Biology of Sharks and Their Relatives*, 2nd Edition. Taylor and Francis Group, LLC, Boca Raton, FL, pp. 265–290.
- Whitney, N.M., Pratt Jr., H.L., Pratt, T.C., Carrier, J.C., 2010. Identifying shark mating behaviour using three-dimensional acceleration loggers. *Endanger. Species Res.* 10, 71–82. <https://doi.org/10.3354/esr00247>.
- Williams, H., Shepard, E., Duriez, O., Lambertucci, S.A., 2015. Can accelerometry be used to distinguish between flight types in soaring birds? *Animal Biotelemetry* 3 (45). <https://doi.org/10.1186/s40317-015-0077-0>.
- Williams, H.J., Holton, M.D., Shepard, E.L., Largey, N., Norman, B., Ryan, P.G., Duriez, O., Scantlebury, M., Quintana, F., Magowan, E.A., 2017. Identification of animal movement patterns using tri-axial magnetometry. *Movement Ecol.* 5 (6). <https://doi.org/10.1186/s40462-017-0097-x>.
- Wilmers, C.C., Nickel, B., Bryce, C.M., Smith, J.A., Wheat, R.E., Yovovich, V., 2015. The golden age of bio-logging: how animal-borne sensors are advancing the frontiers of ecology. *Ecology* 96, 1741–1753. <https://doi.org/10.1890/14-1401.1>.
- Wilson, R.P., White, C.R., Quintana, F., Halsey, L.G., Liebsch, N., Martin, G.R., Butler, P.J., 2006. Moving towards acceleration for estimates of activity-specific metabolic rate in free-living animals: the case of the cormorant. *J. Anim. Ecol.* 75, 1081–1090. <https://doi.org/10.1111/j.1365-2656.2006.01127.x>.