Establishing best practice for the classification of shark behaviour from bio-logging data

Submitted by

Jenna L. Hounslow, BSc.

This Thesis is presented for the degree of
Bachelor of Science Honours in Marine Science
School of Veterinary and Life Sciences, Murdoch University, 2018

Declaration

I declare this thesis is my own account of my research and contains as its main content work which has not previously been submitted for a degree at any tertiary education institution.

Jenna Hounslow

Abstract

Understanding the behaviours of free-ranging animals over biologically meaningful time scales (e.g. diel, tidal, lunar, seasonal, annual) gives important insights into their ecology. Bio-logging tools allow the remote study of elusive or inaccessible animals by recording high resolution multi-channel movement data, however archival device recording duration is limited to relatively short temporal-scales by memory and battery capacity. Machine learning (ML) is becoming common for automatic classification of behaviours from large data sets. This thesis develops a framework for the programming of bio-loggers for the classification of shark behaviour through the optimisation of sampling frequency (Chapter 2) and the choice of movement sensor (Chapter 3).

The effects of sampling frequency on behavioural classification were assessed using data published in a previous study collected from accelerometer equipped juvenile lemon sharks (*Negaprion brevirostris*) during captive trials in Bimini, Bahamas. The impacts of different combinations of movement sensors (accelerometer, magnetometer and gyroscope) were assessed using data collected from sub adult sicklefin lemon sharks (*Negaprion acutidens*). Sharks were equipped with multi-sensor devices recording acceleration, angular rotation and angular velocity during captive trials at St Joseph Atoll, Seychelles. Catalogues of discrete classes of behaviours (ethograms) were developed by observing sharks during captive trials.

Behaviours (swim, rest, burst, chafe, headshake) were classified using a random forest ML algorithm with predictor variables extracted from the ground-truthed data. A range of sampling frequencies (30, 15, 10, 5, 3 and 1 Hz) and combinations of movement sensors were tested. For each dataset, a confusion matrix was determined from model predictions for calculation and comparison of evaluation metrics. Classifier performance was best described by the class or macro *F*- score, a measure of model performance, one indicating perfect classification and zero indicating no classification.

As sampling frequency decreased, classifier performance decreased. Best overall classification was achieved at 30 Hz (*F*- score >0.790), although 5 Hz was appropriate for classification of swim and rest (>0.964). Behaviours characterised by complex movements (headshake, burst, chafe) were best classified at 30 Hz (0.535- 0.846). Classification of behaviours was best with a tri-sensor combination (0.597), although incorporating an additional sensor (magnetometer or gyroscope) resulted in little increase in classifier performance compared to using an accelerometer alone (0.590 compared to 0.535 respectively).

These results demonstrate the ideal sampling frequencies and movement sensors for best-practice programming of bio-logging devices for classifying shark behaviour over extended durations. This thesis will inform future studies incorporating behaviour classification, enabling improved classifier performance and extending recording duration of bio-logging devices.

Table of Contents

Abstract…… iii

Table of Contents v

Acknowledgments vi

Chapter 1. Introduction 1

1.1 General introduction 1

1.2 Bio-logging technology 2

1.3 Studying animal movement and behaviour using accelerometers 2

1.4 Using machine learning to classify animal behaviours 8

1.5 Classifying shark behaviour from accelerometers using machine learning 11

1.6 Challenges in behavioural classification 14

1.7 Current study and research aims 17

Chapter 2. Assessing the effect of sampling frequency on the classification of shark behaviour from accelerometers 19

2.1 Introduction 19

2.2 Method 20

2.3 Results 30

2.4 Discussion 41

2.5 Conclusion 50

Chapter 3. Evaluating the impact of movement sensors for the classification of shark behaviour 51

3.1 Introduction 51

3.2 Method 53

3.3 Results 63

3.4 Discussion 73

3.5 Conclusion 80

Chapter 4. General Discussion 81

References… 84

Supplementary Material 91

Acknowledgments

Most importantly thank you to my inspirational supervisor, Adrian Gleiss. I consider myself privileged to have had the opportunity to study under your continually supportive guidance and for being considered to undertake such an exciting and novel project. I am so grateful and wholeheartedly look forward to the prospect of future collaboration and research.

Thank you to all of my colleagues and friends at the lab. Lauran Brewster, for your help with data and support when I could not see the forest for the trees. Karissa Lear, for your early support and hints with IGOR and your advice for improving my writing. Thank you, Oliver Jewell, for proof reading. Heartfelt thank you to Evan Byrnes. Your kind, calm and practical nature was a massive contribution to the success of my field work. Honourable mentions: sashimi, bingo and zinc. To everyone else at the lab, thank you for contributing to the positive, motivated and enthusiastic fishy- group I have been lucky to work with on a daily basis.

Thank you to Save Our Seas Foundation (SOSF) for their generous funding by awarding a small grant for this project, and for their dedication to research, education and conservation of sharks. Thank you also to Murdoch University for providing additional funding.

Thank you to both my co-supervisor, Ryan Daly and to Clare Daly at SOSF D’Arros Research Centre, for providing access to the study site, logistical support in the field and for extending your warm hospitality. Thank you both for sharing your impassioned knowledge of D’Arros and St Joseph Atoll, and all of its bright and beautiful creatures. D’Arros and St Joes really took my breath away.

Lastly, I thank my friends and family for their continued, unwavering support of my academic pursuits. I hope that some of you have been enlightened, amused and inspired by my shark stories. Dad, from one fishy thesis to another, thank you for passing on to me your vivid fascination with nature.

Chapter 1. Introduction

1.1 General introduction

Animal behaviour is the response of an animal to different internal and external factors (Cooke *et al.* 2004). Studying behaviour is an essential aspect of animal 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 animals' behaviour is therefore an essential component for informing conservation measures. For example, behavioural studies provide great insight into understanding an animal's diet choice, home range, social dynamics and breeding patterns, providing a significant contribution towards solving conservation issues such as the consequences of environmental change, reserve or policy planning and captive breeding programmes (Sutherland 1998, Cooke 2008, Abrahms *et al.* 2016). However, the study of the behaviour of free-ranging animals in their natural environment is often challenging and therefore under-utilised in conservation (Hays *et al.* 2016). This is partially due to the difficulty in directly observing aquatic, migratory or nocturnal species, particularly when the presence of a human observer may inadvertently alter natural animal behaviour (Gleiss *et al.* 2009a, Brown *et al.* 2013, Wilson *et al.* 2015, Hammond *et al.* 2016). However, recent advances in bio-logging technologies present a tool well suited for remotely observing free-ranging animals' behaviours, physiology, movements, and their surrounding environments (Wilmers *et al.* 2015).

1.2 Bio-logging technology

Bio-logging is the remote measurement of environmental, physiological and behavioural data, which can be recorded via an electronic animal-borne tag (logger) (Cooke *et al.* 2004, Shepard *et al.* 2008b, Ropert-Coudert *et al.* 2012, Sherub *et al.* 2017). The term bio-logging was devised in 2003 and has since become more prevalent for the remote study of animals (Ropert-Coudert *et al.* 2012). Bio-logging technology affords us the opportunity to follow animals without direct observation, enabling us to monitor individuals over greater temporal and coarser geographical scales than direct observation alone (Sundström *et al.* 2001). High resolution, fine scale data can be stored on (or logged to) an on-board memory and retrieved for download (Domenici and Blake 1997, Sundström *et al.* 2001, Brown *et al.* 2013). One of the most recent, and significant advances in bio-logging technology is the accelerometer (Rutz and Hays 2009), the application of which has surged in recent years due to increased commercial availability and affordability (Cooke *et al.* 2004, Whitney *et al.* 2012, Brown *et al.* 2013).

1.3 Studying animal movement and behaviour using accelerometers

Accelerometers measure body acceleration, or the acceleration of an animal's mass due to the movement of its body (Wilson *et al.* 2006, Shepard *et al.* 2008a). Since particular movements correspond to distinct behaviours, accelerometers enable the remote study of animal behaviour in unprecedented detail. Acceleration is generally recorded from three orthogonal axes (Figure 1.1); dorso-ventral (heave), anterior-posterior (surge) and lateral (sway) (Brown *et al.* 2013), although in some cases single or bi-axial measures are reported (Yoda *et al.* 1999, Yoda *et al.* 2001, Ropert-Coudert *et al.* 2004, Watanabe *et al.* 2005). These measurements of raw acceleration have both static and dynamic components. Static acceleration represents an animals posture in relation to Earth's gravitational field, whilst dynamic acceleration corresponds directly to an animal's body movement (Wilson *et al.* 2006, Shepard *et al.* 2008a).

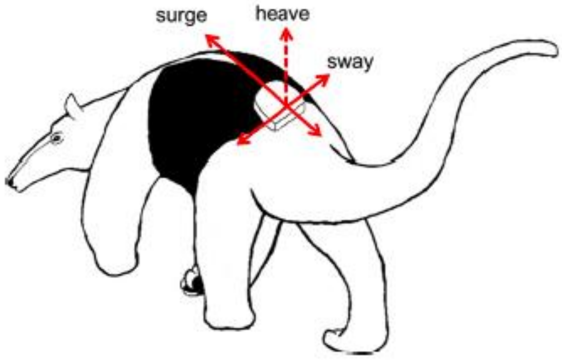

Figure 1.1 Schematic showing the orientation of three accelerometer axes on a terrestrial animal. Surge represents the anterior-posterior axis, heave corresponds to the dorso- ventral axis and sway represents the lateral axis. Reproduced from Brown *et al.* (2013).

Static acceleration can be used to distinguish postural behaviours based on the orientation of an animal's body, such as pitch and roll (Sato *et al.* 2003, Shepard *et al.* 2008b). Whilst early studies identified behaviours based on static acceleration, not all behaviours are characterised by distinct postures. Dynamic acceleration characterises body motion largely resulting from movement of the limbs. Consequently, dynamic acceleration is most representative and hence often more useful for distinguishing an animals locomotory behaviours (Whitney *et al.* 2012). Shepard *et al.* (2008b) demonstrated how different types of repetitive limb movements correspond to accelerometer output, or waveform signals. Importantly different locomotor modes were represented in different axes (Figure 1.2). For example, flying in birds is represented by dynamic acceleration in the dorso-ventral heave axis, due to the vertical up-down wing movements.

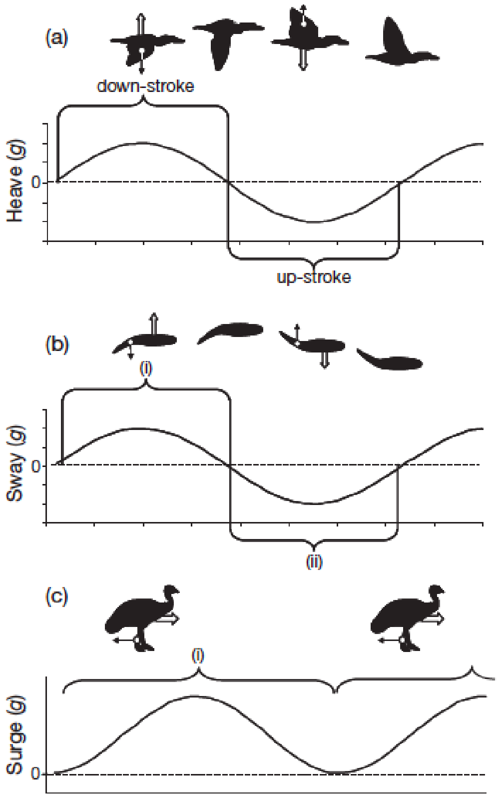

Figure 1.2 Schematic diagram of different types of locomotion (flying, swimming and walking) represented by the waveform signal for dynamic acceleration in the heave, sway and surge axes. Reproduced from Shepard *et al.* (2008b)

Classification of behaviour is a result of combining accelerometer data with direct observations of the tagged animal, thus providing a direct link between acceleration signals and behaviour. One of the earliest studies to classify behaviour in animals using accelerometers was by Yoda *et al.* (2001), where data were used to deduce whether Adelie penguins (*Pygoscelis adeliae*) were standing, walking, porpoising or tobogganing. Accelerometers have since been used to present behavioural time budgets for a variety of terrestrial and aquatic animals; including seals (*Leptonychotes weddellii*), gannets **(***Morus capensis),* domestic cats (*Felis catus*) and imperial cormorants (*Phalacrocorax atriceps*; (Sato *et al.* 2003, Ropert-Coudert *et al.* 2004, Watanabe *et al.* 2005, Laich *et al.* 2008).

In order to classify behaviour in wild animals using accelerometers, it is essential to understand how different types of behaviour translate into accelerometer data output (waveform signals) for a particular species. Active locomotory (e.g. walking, flying, swimming) and inactive non-locomotory (e.g. sitting, resting, standing) behaviours correspond to unique patterns in acceleration waveforms for each species (Figure 1.2; Shepard *et al.* 2008b). Behaviours can then be deduced from the frequency and amplitude of the acceleration signal waveform patterns (Sakamoto *et al.* 2009, Brown *et al.* 2013). Some groups of behaviours are clearly distinguishable in accelerometer data. For example, active and inactive behaviours are easy to visually separate, due to the relative low amplitude and frequency of inactive waveform signals in comparison to higher values for active moving behaviours (Figure 1.3). However, for more detailed classification distinguishing between different active behaviours (e.g. travelling or escaping) it is necessary to first validate accelerometer data using direct observational studies, so that acceleration waveform signals can be matched with some degree of certainty to observed behaviours for classification (Brown *et al.* 2013).

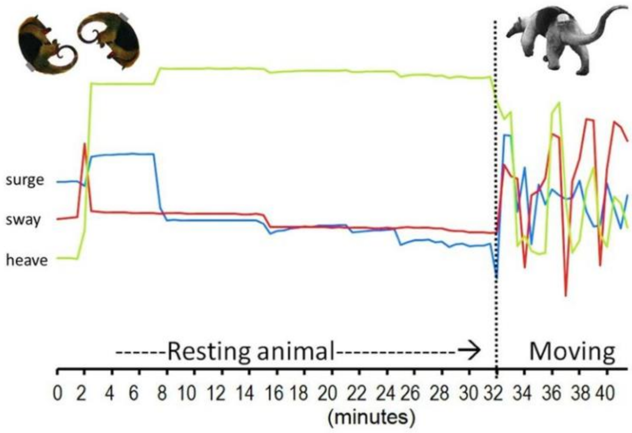

Figure 1.3 Accelerometer signal waveform patterns resulting from different forms of locomotion are related to active (moving) and inactive (resting) behaviours in a Tamandua anteater (*Tamandua mexicana*). Reproduced from Brown *et al.* (2013).

A fundamental part of the behavioural classification process is observation, despite accelerometers being commonly used to classify behaviour which cannot be observed (Brown *et al.* 2013). It is therefore essential that acceleration data are validated with observations of known behaviours. This process of validation is called ground-truthing and can be achieved via a period of direct observation or the use of video footage of the test animals whilst they are equipped with accelerometers (Watanabe *et al.* 2005, Laich *et al.* 2008). Animal behaviour is visually observed simultaneously to recording acceleration, so that accelerometer data can be calibrated with associated observed behaviours in pre-determined reporting intervals, for example every second (Robert *et al.* 2009, Campbell *et al.* 2013). It is difficult however, to ensure that all naturally occurring behaviours have been witnessed whilst the animal is in captivity for data validation, even when direct observation is employed (Jule *et al.* 2009).

Following observations, known behaviours are assigned to patterns in the accelerometer waveform signals by calculating a range of statistics calculated from each of the three axes. These statistics have been used to produce rudimentary keys to identify and differentiate between behaviours (Figure 1.4; Watanabe *et al.* 2005, Laich *et al.* 2008). However, there is no standard protocol for data analysis, with methods varying from comparing simple statistics such as mean, frequency and variance (Laich *et al.* 2008) to more complex analysis methods such as Fast Fourier Transformation (Watanabe *et al.* 2005).

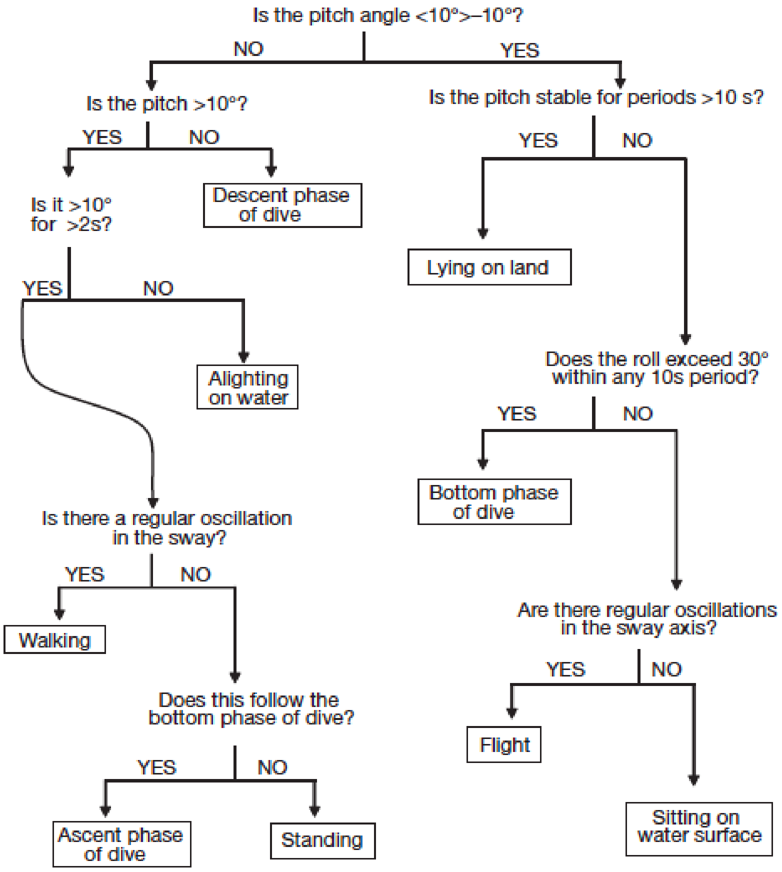

Figure 1.4 Example of a behavioural identification key where classification of behaviour is achieved by manually evaluating different features of the accelerometer data. Reproduced from (Laich *et al.* 2008).

1.4 Using machine learning to classify animal behaviours

Perhaps the most poignant challenge in the use of accelerometer techniques is that data collection processes are overtaking the ability to analyse the data itself (Valletta *et al.* 2017). Accelerometers can produce millions of rows of data, as a result, manual data analysis is a time-consuming task presenting one of the main challenges in using this technology (Shepard *et al.* 2008b). Consequently, there is a need for automatic classification of behaviour. Animal behaviour has been classified from machine learning (ML) algorithms in the past (Nathan *et al.* 2012, Gao *et al.* 2013, Hussey *et al.* 2015). ML is a hypothesis free method that learns from patterns in data to make predictions, unlike statistical modelling that makes an assumption about the data to later be rejected or accepted (Valletta *et al.* 2017). The introduction of ML allows for automatic classification of behaviour directly from large, complex data sets, helping to overcome the practical challenges in manual data analysis (Bidder *et al.* 2014). ML comprises either supervised or unsupervised learning methods to classify behaviours from waveform signals automatically (Brown *et al.* 2013, Walker *et al.* 2015).

Unsupervised learning clusters similar features of acceleration waveform signal patterns together, which are then partitioned and assigned to behaviours *a priori* (Valletta *et al.* 2017). In the first study to use an automated ML technique, Sakamoto *et al.* (2009) applied an unsupervised algorithm, k-means clustering or k-nearest- neighbour (KNN), to successfully quantify behaviours of free- ranging European shags (*Phalacrocorax aristotelis*), showing that unsupervised ML had potential for successful behavioural classification. The second method, supervised learning, is far more common for behaviour classification in animals (Brown *et al.* 2013). Supervised learning workflows involve labelling a subset of ground-truthed data behaviours to train a classification algorithm. The remaining data is used as a test-set, to automatically make predictions and test the performance of the classifier (Brown *et al.* 2013, Gao *et al.* 2013, Valletta *et al.* 2017).

1.4.1 Choosing an appropriate machine learning method

Unsupervised ML algorithms are often the preferred method for behavioural classification (Prasad *et al.* 2006, Bidder *et al.* 2014, Sur *et al.* 2017). Unsupervised methods such as KNN have an advantage in identifying unknown behaviours which may not have been directly observed in the ground-truthing stage (Brown *et al.* 2013). Unsupervised learning methods are simpler to implement than supervised methods as they do not require extensive ground-truthing experiments. However, these clustering techniques may identify fewer or more clusters in the data than there are behaviours, and clusters may not be assigned to behaviours with the certainty gained from ground-truthing (Brewster *et al.* 2018). One of the main benefits of supervised learning is that the observed behaviours correspond directy to ground-truthed data (Sakamoto *et al.* 2009). There are several supervised ML tehcniques that can be used to classify behaviour. For example, Nathan *et al.* (2012) compared five different supervised ML models, aiming to establish a general protocol for classification of animal behaviour. Acceleration waveform signals ground- truthed to behavioural ethograms from griffon vultures (*Gyps fulvus*), were classified using various techniques; support vector mechanisms (SVM), classification and regression trees (CART), random forest (RF), artificial neural networks (ANN) and linear discriminant analysis (LDA). All methods were found to have high accuracy (80-90 %), yet there are benefits and drawbacks to each method. Some were considered more practical to implement than others, such as CART (a hierarchical set of decision rules), and RF, a combination of multiple decision trees (Breiman 2001, Cutler *et al.* 2007), performed better than the other methods (Nathan *et al.* 2012). Both CART and RF methods have been used successfully for classification of behaviour in various taxa. Shamoune-Baranes *et al.* (2012) used CART as their chosen model for behaviour classification for oystercatchers (*Haematopus ostralegus*) and RF models have successfully predicted behavioural classes in a range of animals with high performance (Wang *et al.* 2015, Sur *et al.* 2017, Valletta *et al.* 2017, Walton *et al.* 2018).

Despite the popularity of RF models for behaviour classification, this ML method has been referred to as a ‘black-box’ algorithm, where unlike CART for example, the results of each internal decision are unknown (Breiman 2001, Prasad *et al.* 2006, Nathan *et al.* 2012).

Choosing an appropriate ML method is dependent on each individual study, with recent studies having combined the strengths of multiple supervised ML techniques to improve classification of animal behaviour (Ladds *et al.* 2017, Brewster *et al.* 2018). These ‘super learners’ are labour intensive and time consuming when compared to the constituent ML models. Statistical software is becoming increasingly available to improve classification of animal behaviour from sensor data using both supervised and unsupervised ML, e.g. Ethographer (Sakamoto *et al.* 2009), AcceleRator (Resheff *et al.* 2014) and Framework4 (Walker *et al.* 2015). Ultimately the choice of ML analysis arises from how the data are intended to be collected, validated and interpreted (Resheff *et al.* 2014).

1.4.2 Training machine learning algorithms with predictor variables

The predictor variables with which supervised ML algorithms are trained reflect the characteristics of the raw data in relation to each class of behaviour (Nathan *et al.* 2012, Wang *et al.* 2015, Valletta *et al.* 2017). Analogous to the manual methods for inferring behaviour from accelerometer data, there is no fixed protocol for which or how many predictor variables are extracted to use for training the ML algorithm. Furthermore, it has been argued that predictor variable selection may confuse the relationship between movement and behaviour (Bidder *et al.* 2014). As few as 15 to as many as 147 different statistics have been calculated as predictor variables for ML studies (Shamoune-Baranes *et al.* 2012, Ladds *et al.* 2017). Common statistical features, characteristic of the data, that are used as predictor variables include mean, minimum, maximum, variance, standard deviation, kurtosis etc. in all three axes for both static and dynamic acceleration (Nathan *et al.* 2012, Tanha *et al.* 2012). Variables related to the animal (e.g. sex, length, weight) and the environment (e.g. depth, temperature) are also commonly added to the summary (Ladds *et al.* 2017). One common predictor variable used for classification of behaviour from accelerometer data is the metric Overall Dynamic Body Acceleration (ODBA), a simple metric describing the intensity of body movement of an individual (Wilson *et al.* 2006, Gleiss *et al.* 2011b).

1.5 Classifying shark behaviour from accelerometers using machine learning

Elasmobranchs, including sharks, occupy the upper trophic level of marine ecosystems and are valuable indicator species for environmental health. Therefore, it is of great value to consider their behavioural ecology (Cortés 1999, Sims 2003). Accelerometer devices can be attached relatively easily to the dorsal fin of a shark (as per Figure 1.5). As with all fish, every repetitive lateral beat of the tail (causing forward propulsion) corresponds to an oscillatory pattern in the dynamic sway acceleration axes waveform signal (Figure 1.6), from where behaviour can be quantified (Wilson *et al.* 2015). Other behaviours such as vertical movements, can be inferred from changes in pitch from the surge acceleration axes (Nakamura *et al.* 2011). Accelerometers have previously been used to study different aspects of behaviour in sharks, including swimming patterns (Whitney *et al.* 2007, Gleiss *et al.* 2009b, Nakamura *et al.* 2011, Gleiss *et al.* 2013, Wilson *et al.* 2015), energetics (Gleiss *et al.* 2009a, Gleiss *et al.* 2010, Lear *et al.* 2017, Bouyoucos *et al.* 2018), mating behaviour (Whitney *et al.* 2010), and travel cost (Gleiss *et al.* 2011b, Payne *et al.* 2016), providing important detailed information describing the behavioural ecology of many species.

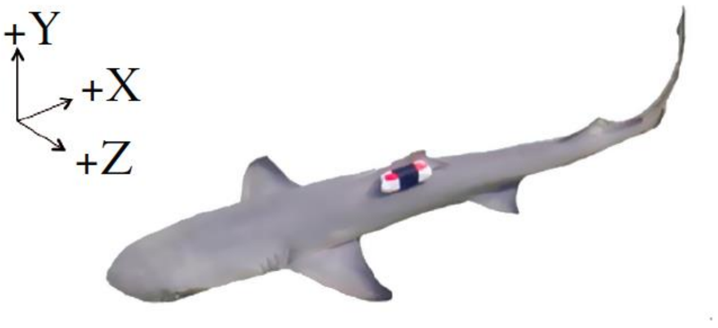

Figure 1.5 Schematic diagram of a lemon shark (*Negaprion brevirostris*) equipped with a dorsally mounted accelerometer tag. Arrows show the direction of the three acceleration axes; X (surge), Y (heave) and Z (sway). Reproduced from Wilson *et al.* (2015).

Both unsupervised and supervised ML techniques have been used to classify shark behaviour from accelerometer data. Unsupervised classification of accelerometer data has previously allowed for the successful identification and quantification of various swimming behaviours to create time-activity budgets for white tip reef sharks (*Triaenodon obesus*) (Whitney *et al.* 2007) and lemon sharks (*Negaprion brevirostris*) (Gleiss *et al.* 2009a, Wilson *et al.* 2015). Unsupervised learning was also used to differentiate between complex transitional swimming behaviours helping to explain social dynamics in lemon sharks from accelerometer data (Wilson *et al.* 2015). In the first study to investigate mating behaviour in sharks, Whitney *et al.* (2010) used both supervised and unsupervised ML methods to classify between swimming, resting and mating events in nurse sharks (*Ginglymostoma cirratum*; Figure 1.6). Where more complex behaviours, such as foraging, were not directly observed for validation by ground- truthing, they were hypothesised to explain any unlabelled accelerometer data. More comprehensive supervised ML studies of shark behaviour have incorporated depth, temperature and ODBA alongside behavioural classification, revealing that reef sharks displayed higher nocturnal activity levels in warmer waters (Leos-Barajas *et al.* 2017). The most recent classification study of shark behaviour combined several supervised learning methods (e.g. super learner), to classify five behaviours in lemon sharks, modelling specific behaviours of interest against environmental variables (Brewster *et al.* 2018).

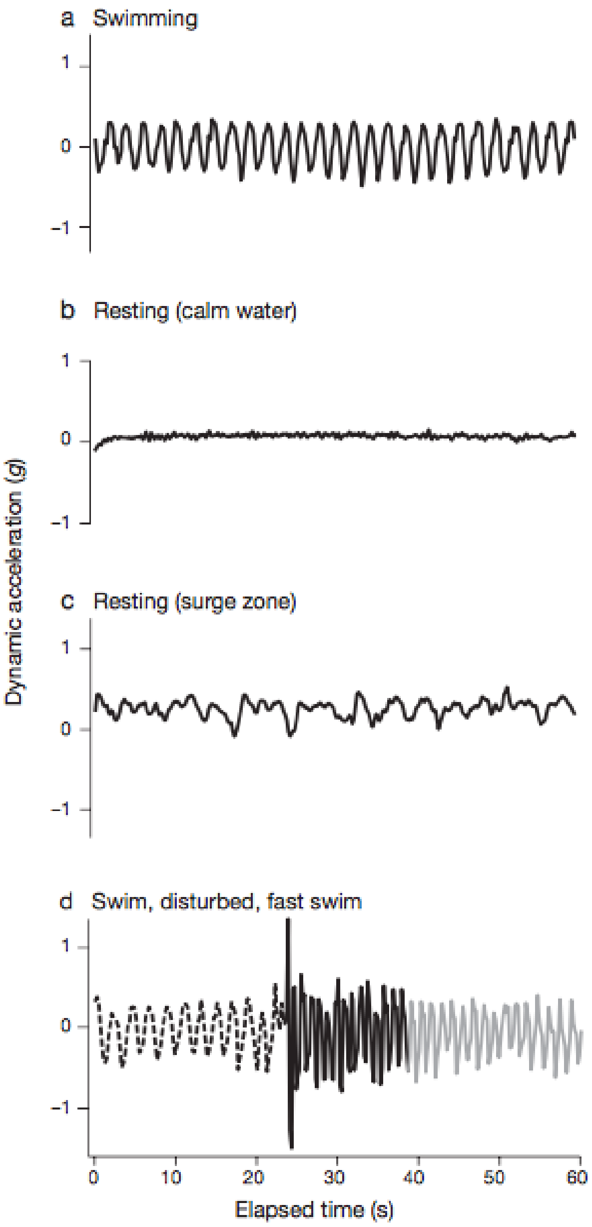

Figure 1.6 Accelerometer signal waveform patterns associated with different types of swimming behaviour in the dynamic sway axes for nurse sharks (*G. cirratum*). Reproduced from Whitney *et al.* (2010).

1.6 Challenges in behavioural classification

Whilst bio-logging tools such as accelerometers offer valuable insights into animal behaviour, there are substantial limitations to be considered, including issues with memory and battery capacity and complicated data analysis. The nature of archival data loggers is such that information is recorded directly to an on-board memory card with a limited capacity for data storage. Recording high resolution data from multiple sensors (accelerometer, magnetometer, gyroscope) produces millions of data points; therefore, devices use a substantial amount of memory and battery power (Shepard *et al.* 2008b). Despite technological advances, acceleration data can only be collected for brief periods of time depending on the size of the device. This provides a limited representation of how animals respond to the environmental conditions, which may be experienced over varying time scales (e.g. diel, tidal, lunar, seasonal, annual) (Whitney *et al.* 2012, Hays *et al.* 2016). The capacity of internal on-board memory cards could be increased to some extent in the future as technological improvements continue to be made, increasing battery life without making devices larger is also important, so as not to preclude the study of smaller animals (Cooke 2008). Both Wilson *et al.* (2008) and Shepard *et al.* (2008b) point out that it would be wise to pre-programme devices at the lowest possible sampling frequency to enable the collection of more data over extended durations in order to record over ecologically meaningful periods of time.

One method currently employed to extend device memory is to programme sensors to record data in bursts, or epochs instead of continually, reducing the amount of data logged (Brown *et al.* 2013). It has also been suggested that on-board processing might be useful in order to identify and record only when particular behaviours of interest occur (Noda *et al.* 2013). However, these methods for conserving memory and battery are not always suitable for behaviour classification, as many behaviours are transitional or short lived, therefore we cannot be sure that each recorded burst represents just one particular behaviour without an overlap into a different behaviour (Resheff *et al.* 2014, Kröschel *et al.* 2017, Ladds *et al.*

2017, Sur *et al.* 2017). The impact of sampling frequency on behavioural classification from accelerometer data has only recently been assessed in a variety of animals (Broell *et al.* 2013, Wang *et al.* 2015, Sur *et al.* 2017, Walton *et al.* 2018). Sur *et al.* (2017) subsampled raw accelerometer data from eagles (*Aquila chrysaetos*) at various frequencies, finding that as sampling frequency was reduced, complex behaviours were misclassified, concluding that sampling frequency can only be reduced to a certain level. Nevertheless, future research must establish best practice for reducing sampling frequency so as to maximise memory and battery capacity without sacrificing performance in behavioural classification.

As a result of technological advances, multiple movement sensors can now be built into one data logging device, providing a complete reconstruction of an animals' total body movement (Shepard *et al.* 2008b, Sherub *et al.* 2017). Whilst accelerometers are useful for the quantification of behaviours by studying body acceleration in up to three dimensions, these devices cannot always provide body orientation information regarding angular rotation (turning direction) or angular velocity (turning speed) (Fourati *et al.* 2011). This can be overcome by using bio-logging data collected from alternative movement sensors such as magnetometers and gyroscopes. Magnetometers measure angular rotation, defining body orientation relative to the Earth's magnetic field lines independently from gravity and dynamic body movement (Walker *et al.* 2015, Williams *et al.* 2015). Similar to accelerometers, tri-axial magnetometers can indicate active and inactive behaviours through measurement of angular rotation (Figure 1.7; Wilson *et al.* 2008). This has allowed for the study of activity patterns using magnetometers in various aquatic animals including turtles, pinnipeds and cetaceans (Hochscheid *et al.* 1999, Davis *et al.* 2003, Mitani *et al.* 2003, Ware *et al.* 2011).

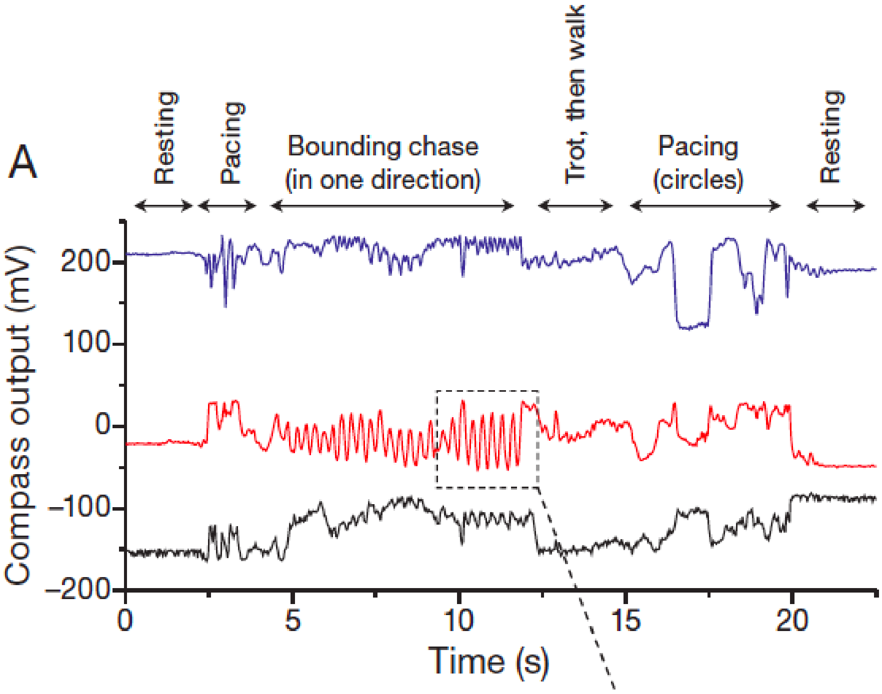

Figure 1.7 Example of how behaviour can be inferred from magnetometer data (compass output) for a cheetah (*Acinonyx jubatus*). The large changes in compass output values result from changes in body orientation during running behaviour. Reproduced from Wilson *et al.* (2008).

Another movement sensor capable of providing a direct measure of orientation, is the tri-axial gyroscope, measuring angular velocity or rotational turning speed about each axis. Behaviours can be visualised from gyroscope data in the same way that they are represented in accelerometer or magnetometer data (Walton *et al.* 2018). Data can be recorded simultaneously in three channels from each of the three movement sensors described. This tri-sensor combination of movement data was successfully used to improve monitoring of different behaviours in dogs (Fourati *et al.* 2011). Meanwhile in the aquatic environment, combining all three types of movement sensors improved the determination of complex fine scale behaviours in teleost fishes (Noda *et al.* 2014), further demonstrating that the integration of multiple movement sensors could improve classification of fine scale behaviours.

1.7 Current study and research aims

This extensive review has demonstrated that ML provides an automated method for processing the large data sets produced by high sampling frequencies of accelerometers, thus overcoming some of the challenges arising from manual data analysis for behaviour classification. However, the crucial problem remains that battery and memory capacity of most commercially available devices are exceeded very quickly when sampling at frequencies assumed to be high enough to distinguish fine scale behaviour; so too, with the incorporation of multiple sensors. This limits our ability to record data from these devices over biologically meaningful time scales. This problem could be reduced by only recording data with the sensors that are most informative for classification of behaviour, and by reducing sampling frequency without sacrificing classifier performance.

The broad objective for this thesis is to develop a “best-practice” framework for the programming of bio-loggers for the classification of shark behaviour. This will allow future researchers to refine their programming of bio-logging tags, to gather more data for every tag deployed for every wild shark tagged. The results of this thesis will in turn allow for the monitoring of shark behaviour for longer periods in the wild. This will promote the uptake of these technologies and generate valuable knowledge in the discipline of ecology in sharks, resulting in improved long-term conservation and management practices. To achieve this objective, supervised ML techniques were used, translating ground- truthed data into behavioural classes and by developing an algorithm for automatic classification of the data. One data set, previously collected and presented by (Brewster *et al.* 2018) was re- analysed to optimise sampling frequency. A new data set, collected for this thesis, was analysed to assess the impact of choice of movement sensors on classification.

The following research questions will be addressed in two separate chapters:

Chapter 2. How does sampling frequency influence the classification of shark behaviour?

Chapter 3. Which movement sensors are most important for the classification of shark behaviour?

Chapter 2. Assessing the effect of sampling frequency on the classification of shark behaviour from accelerometers

2.1 Introduction

Accelerometers offer valuable insights into animal behaviour, however there are substantial limitations which must be considered regarding issues with memory and battery capacity Recording high resolution data (16- bit, >20 Hz) produces millions of data points; therefore, archival data loggers that record data directly to an internal memory card, have a limited capacity for data-storage. Despite technological advances, acceleration data can only be collected for brief periods of time depending on the size of the device. This provides a limited representation of how animals respond to the environmental conditions, which may be experienced over varying time scales (for example diel, tidal, lunar, seasonal, annual; Whitney *et al.* 2012, Hays *et al.* 2016). To enable the collection of data over ecologically meaningful time scales (without increasing the size of devices thus precluding smaller animals), accelerometers must be pre-programmed at the lowest possible sampling frequency without sacrificing classification of behaviours (Cooke 2008, Wilson *et al.* 2008, Shepard *et al*. 2008b).

A wide range of sampling frequencies have been used to relate behaviours to acceleration data, from as low as 2 Hz (Wang *et al.* 2015) to 100 Hz (Broell *et al.* 2013, le Roux *et al.* 2017), with higher sampling frequencies allowing for finer determination of behaviours (Nathan *et al.* 2012). In general, sampling frequency should be at least twice that of the minimum frequency of the most rapid body movement of interest (Halsey *et al.* 2009, Graf *et al.* 2015), which is referred to as the Nyquist criterion (Chen and Bassett Jr 2005, Brown *et al.* 2013). In this chapter, the first objective of this thesis is addressed, where I assess the influence of different sampling frequencies on classifier performance.

2.2 Method

The data set used for this chapter was collected by and presented by Brewster *et al.* (2018). In order to assess the effects of sampling frequency on behaviour classification from this data, further analysis was conducted for this thesis chapter. Detailed methods are as follows.

2.2.1 Study site and species

The Bimini Islands, Bahamas (25°44’N, 79°16’W), are approximately 85 km due east of Miami, Florida, USA. Two small mangrove fringed islands form the main study site of the Bimini Biological Field Station Foundation, with the area well documented as a nursery habitat for juvenile lemon sharks (*Negaprion brevirostris*) (Morrissey and Gruber 1993b, Chapman *et al.* 2009). The lemon shark, listed as near threatened by the International Union for the Conservation of Nature (IUCN), is a carcharinid elasmobranch found in the warm-temperate and tropical waters of the West-Eastern Atlantic and Eastern Pacific oceans (Sundström 2015). The lemon shark was selected as the study species due to its abundance, size and propensity for captive studies at the study site (Chapman *et al.* 2009).

2.2.2 Shark capture

Juvenile lemon sharks (*n* =4) were captured using a gillnet set perpendicular to the southern shoreline of the South Bimini Island (Table 2.1). Nets were checked at regular 15-minute intervals and individuals were checked for previous capture by identification of a sub-dermally implanted passive integrated transponder (PIT; Destron Fearing Inc) (Gruber *et al.* 2001). Juvenile sharks were transported to a nearby purpose built 10 x 6 m rectangular pen which was erected in the shallows on neighbouring sandflats. The size range for captive sharks was 79.2- 85.2 cm total length (TL) (Table 2.1).

Table 2.1 Juvenile lemon sharks (*N. brevirostris*) used during captive trials for development of a behavioural ethogram

PIT Tag ID	Sex	Total Length (TL) (cm)	Weight (kg)
985121031792723	Female	82.6	3.75
4C4A2D3A12	Female	80.5	3.10
4C3A6C313A	Male	79.2	3.15
4C3B312275	Male	85.2	3.75

2.2.3 Tagging equipment

Sharks were equipped with Cefas G6a+ triaxial acceleration data loggers (accelerometers) (40 mm x 28 mm x 17 mm) which recorded acceleration at 30 Hz (Cefas Inc, Lowestoft, UK). Tags were coupled with epoxy resin to acoustic transmitters (9 mm x 25 mm). Tags were attached to individual sharks through two holes, 1.5 mm in diameter, made in the base of the first dorsal fin with a hypodermic needle (Figure 2.1). Nylon monofilament was looped through pre-drilled holes in the accelerometer package and through the corresponding holes in the fin. 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). Tag packages were a maximum of 0.79 (±0.15) % of the animal's mass and 6.94 (±1.36) % of the cross-sectional area of the dorsal fin to which the tag was attached to.

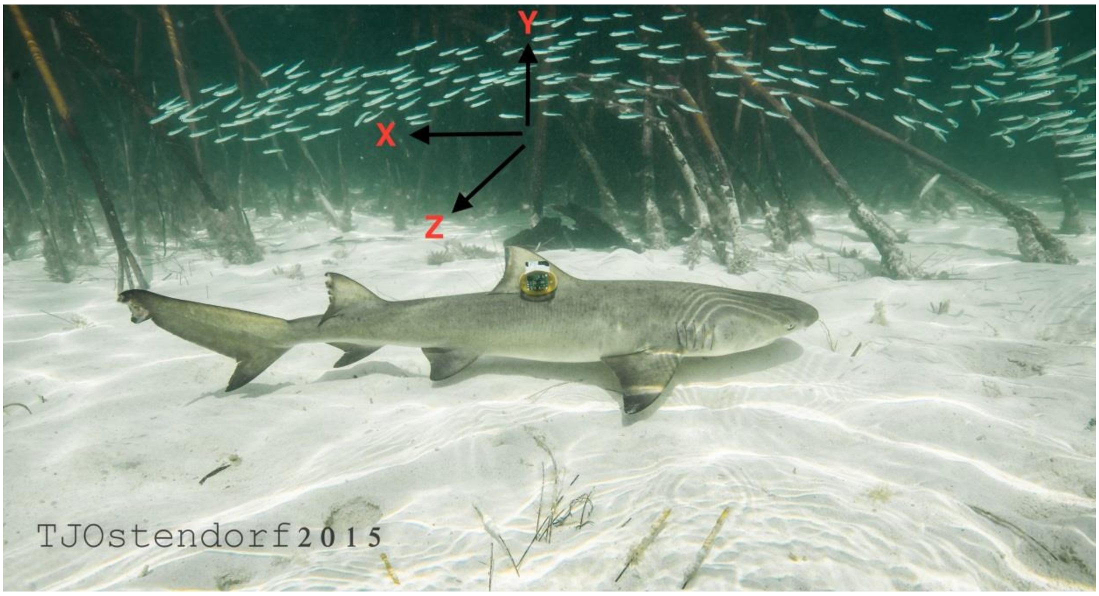

Figure 2.1 Juvenile lemon shark (*N.brevirostris*) equipped with a dorsally mounted accelerometer tag. Arrows show the direction of the three acceleration axes, X, Y and Z. (Photo credit: T.J.Ostendorf, Bimini Biological Field Station, 2015)

2.2.4 Captive trials and ethograms

The pen holding the captive sharks was constructed in a rectangular shape to minimise repetitive circular swimming patterns previously observed in captive trials using circular pens (Gleiss *et al.* 2009a). The pen was constructed using a plastic open-meshed fencing material, ensuring that ambient environmental conditions were experienced by the captive shark (e.g. salinity, temperature, tidal cycle, lunar cycle; Guttridge *et al.* 2009). Sharks were fed to satiation every third day except during trials, with tag packages attached after a minimum period of two days within the pen to allow for recovery from capture and acclimation to captivity.

To validate acceleration signatures against behaviours, accelerometer tagged animals were directly observed (i.e. ground- truthing). Observations were conducted from 3 m high wooden towers built adjacent to the pens. Each second of the semi-captive observation period was classified as one of five discrete behaviours (swimming, resting, burst swimming, chafing, and head shaking [Table 2.2]) by recording initiation and cessation times for each behaviour as it was performed. A catalogue of observed behavioural events, or ethogram (Sakamoto *et al.* 2009) was produced for each shark. Burst behaviour was induced to gain further replicates, by throwing dive weights into the water or making large movements to the side of the pen near the tagged sharks. Other behaviours were performed of the shark's own volition.

Table 2.2 Description of observed behaviours in semi- captive lemon sharks (*N.brevirostris*)

Behaviour Class	**Definition**
Swim	Steady lateral undulatory locomotion (Maia *et al.* 2012).
Rest	Lying motionless on sea floor (Whitney *et al.* 2012).
Chafe	Roll motion where dorsal side contacts surface or substrate in effort to remove unwanted parasites or foreign bodies (Myrberg Jr and Gruber 1974).
Burst	Fast-start rapid swim in response to a stimulus (Domenici and Blake 1991, Sundström *et al.* 2001).
Headshake	Side to side movement of the head associated with prey manipulation or predation behaviours (Motta *et al.* 1997).

2.2.5 Data analysis

To determine which sampling frequency was most suitable for characterising lemon shark behaviours, raw acceleration data, collected at 30 Hz, were re-sampled to different frequencies using the resample function in IGOR Pro version 7.06 (WaveMetrics Inc, Lake Oswego, Oregon, USA). Data were re-sampled as a factor of the original sampling frequency using decimation by omission. This re-sampling method involved systematically deleting every 'nth' point of the original data (data points were omitted) according to the new sampling frequency (Broell *et al.* 2013, Sur *et al.* 2017). For example, when resampling 30 Hz data down to 15 Hz, a decimation rate of two results in every second data point being omitted from the original data set (Table 2.3). This method of re-sampling precludes certain sampling frequencies from analysis, such as 20 Hz and 25 Hz, as the new resampled frequencies were dictated by the original sampling frequency. Another method in the resample function enabling sampling frequency conversion uses a combination of both interpolation (up-sampling) and decimation. Whilst this non-systematic re-sampling function may have enabled me to assess other sampling frequencies, this method calculates imagined data points from an estimation between known data points which may not accurately reflect real recorded data points. The decimation by omission method used here ensures that the actual recorded data remained within each re-sampled data set.

Table 2.3 Decimation rates for re-sampled accelerometer data. Raw acceleration data (30 Hz) were resampled using decimation by omission, whereby every nth data point was omitted from the new sampling frequency

Sampling Frequency (Hz)	Decimation Rate
30	N/A
15	2
10	3
5	6
3	10
1	30

Static acceleration, representing body posture in relation to Earth's gravitational field, and dynamic acceleration, representing body movement, were separated in all three axes (X, Y, and Z) using a 3-second box smoother (Shepard *et al.* 2008a). Overall dynamic body acceleration (ODBA) was then 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 from the sway (z) axis, representing TBF and acceleration amplitude (Sakamoto *et al.* 2009). Predictor variables (*n* =44; Supplementary material, Table S1), were then calculated and extracted to form a statistical summary of these features of the acceleration data (extracted as per previous studies; Nathan *et al.* 2012, Shamoune-Baranes *et al.* 2012, Sur *et al.* 2017). To enable time matching of the predictor variables to the observed behaviours which were recorded on a per- second basis, the predictor variables were calculated from one second time series averages of the acceleration derived metrics (Table S1). This procedure was repeated to create a set of behaviour labelled-predictor variables from the ground-truthed accelerometer data for each sampling frequency.

2.2.6 Machine learning and random forest classification

The choice of RF as a classification model for this data set was due to its relative simplicity when compared to more sophisticated classifiers. Classification performance is often related to the choice of ML algorithm itself (Nathan *et al.* 2012, Brown *et al.* 2013, Ladds *et al.* 2017). Combining multiple supervised ML techniques improves overall classification performance (Ladds *et al.* 2017). A voting ensemble (VE) classifier for the prediction of behaviours of lemon sharks from accelerometer data was successfully developed from five supervised ML base models, using the same data set used for this chapter (Brewster *et al.* 2018). Despite the VE classifier performing better than the component base models for behavioural classification, this was owing to very careful and time-consuming development (Brewster *et al.* 2018). RF algorithms are efficient for large data sets, adept in predicting and assessing different internal features of the data and reduce the overfitting tendencies of other ML methods (Breiman 1999, 2001). RF was therefore chosen here as an uncomplicated and simply executed ML model, where the sole purpose of analysis was to assess the qualitative change in classifier performance based on sampling frequency.

RF models are an ensemble classifier—whereby multiple (*ntree* as set by the user) unpruned classification or regression trees are grown—incorporating three steps. First, for each tree the observational data is randomly bootstrapped (with replacement) so that 63 % of the data is used to train the tree. Secondly, 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. Lastly, the predictions from each tree are aggregated and the observation is assigned to the class with the majority vote. The data remaining from the bootstrapping are used by the model for internal cross-validation to calculate classification errors or Out-Of-Bag (OOB) errors (Breiman 2001). Whilst supervised ML methods such as decision trees can be prone to overfitting (significant changes in predictions from small changes in the data), RF learns from data sets whilst minimising overall class error and overfitting (Breiman 2001, Valletta *et al.* 2017).

RF classification was conducted using the ‘randomForest’ package in R (Liaw and Wiener 2002). For each sampling frequency, the ground-truthed acceleration data was first randomly split into two sets whilst maintaining class ratios, a training set and a testing set as per other studies classifying animal behaviour from accelerometer data (Ladds *et al.* 2017, Sur *et al.* 2017). Seventy per cent of the data were used for training the model, and the classification accuracy was cross validated and tested on the remaining 30 % of the data. Whilst data were initially 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 were tested (500, 1000 and 1500). The value resulting in least computational time and the lowest OOB errors was selected for the model (*ntree*=1000). The number of predictor variables at each node, *mtry*, was chosen by using the square root of the total number of predictor variables (n =44) (Verikas *et al.* 2011). The ‘randomForest’ package only uses whole numbers in the model, therefore *mtry* was rounded to 7. RF models are unable to handle missing values, therefore predictor variables such as standard deviation, skewness and kurtosis were omitted from the 1 Hz data set. The total predictor variables for 1 Hz was reduced to n =23, therefore *mtry* =5, which did not change OOB error or classification results.

A data set is imbalanced when it contains one or more classes with more observations (majority classes) than the remaining (minority) classes (Chawla *et al.* 2002, Chawla *et al.* 2003). RF models aim to improve overall classification accuracy and reduce overall error rate, however with an imbalanced data set the model focuses on the predictive power of the majority class, to the detriment of the predictive accuracy for minority classes (Chen *et al.* 2004, Ganganwar 2012). There have been various suggestions for how to deal with imbalanced data sets for classification, assigning class weights and random or direct over- or under-sampling methods (Japkowicz 2000). As this data set was highly imbalanced (Figure 2.2), a stratified subset of the training data set was incorporated into the RF model to reduce. Other studies have created completely balanced data sets where each behaviour is represented by an equal number of measurements (le Roux *et al.* 2017), however these data sets included substantially more observations for the rarest behaviour class, and this was not reflective of how often each behaviour would be performed in the wild. For this study, selective direct under-sampling reduced the majority class (swim) in the training data set to make the majority class frequency closer to the rarest class frequency (burst) by a factor of ten. Despite some loss of potentially important data from the majority class, this method drastically reduced computational time as noted by Chen *et al.* (2004).

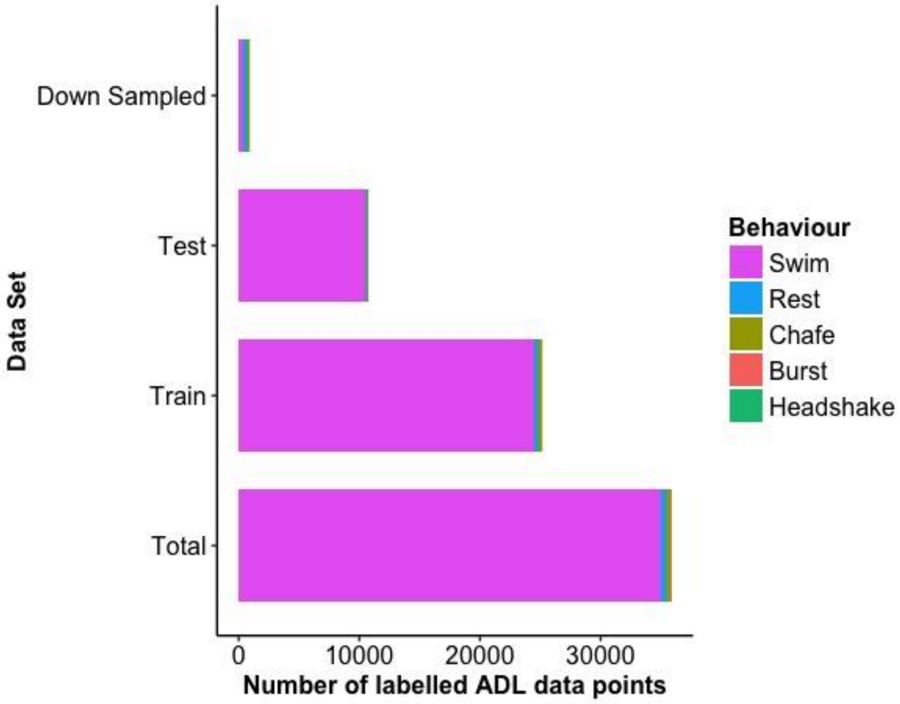

Figure 2.2 Number of per-second averaged accelerometer data points labelled with observed behaviours (swim, rest, chafe, burst and headshake) for lemon sharks (*N. brevirostris*). The total data set was imbalanced. Most of the data was labelled as swim class, which was reflected when the data was split into training and testing data sets. The training data was selectively under sampled to decrease the frequency of the majority classes (swim and rest).

The RF model was applied to the unseen test data set to form predictions and evaluate model performance. Evaluation metrics were calculated from a confusion matrix in the ‘caret’ package in R (Kuhn 2016). A confusion matrix is a table used to describe model performance, where rows are actual observed values and columns are model predicted values, represented by true positive (TP), false positive (FP), and false negative (FN) values (Breiman 1999, 2001). TP values occur when the 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. Evaluation metrics were calculated for both individual behavioural classes and overall for each sampling frequency, allowing for comparison between and within sampling frequencies. Metrics for each sampling frequency were calculated by averaging across all behavioural classes. Evaluation metrics were calculated from the confusion matrix according to Breiman (2002) as follows:

Accuracy — The overall percentage of behavioural classes predicted correctly.

$$(TP + TN)/(TP + TN + FP + FN) \qquad \text{Equation 1}$$

Recall — The proportion of predicted behaviours from a behaviour class that were correctly classified as that behaviour. This metric is also known as sensitivity.

$$R = TP/(TP + FN) \qquad \text{Equation 2}$$

Precision — The proportion of predicted behaviours from a class that were that behaviour. Precision can be poor if recall is accompanied by many predictions being incorrectly assigned to a behavioural class.

$$P = TP/(TP + FP) \qquad \text{Equation 3}$$

F_1 Score	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. $F_1 = 2PR/(P + R)$ Equation 4
Macro F_1	The mean of F_1 scores for all classes used to describe overall classifier performance for each sampling frequency $F_M = \Sigma^M F_1/M$ Equation 5 where M is the number of classes in the classification model.

Individual class F_1 score and overall F_M (macro-averaged F_1) were considered the best descriptors of classification performance for this data set.

Lastly, the relative importance of different predictor variables was estimated using the ‘varImpPlot’ function within the ‘randomForest’ package in R. The importance of predictor variables is calculated using the OOB data (Wang *et al.* 2015, Valletta *et al.* 2017). Mean decrease in accuracy shows how model performance decreases if a predictor variable is removed from the model and mean decrease in Gini Index shows the importance of a predictor variable based on the Gini Impurity Index for the calculation of splits in trees (Chen *et al.* 2004). This analysis (calculating evaluation metrics and predictor variable importance) was repeated for each sampling frequency to allow for direct comparison in model performance as sampling frequency was reduced.

2.3 Results

2.3.1 Ethogram and ground- truthed accelerometer data

Accelerometer data were collected for four lemon sharks during captive trials. From these four sharks, over 35,000 seconds of acceleration data were ground- truthed and labelled as one of five distinct behaviours (Table 2.4). Swimming and resting behaviours were observed to be performed most frequently, whilst burst, chafe and headshake behaviours were performed relatively infrequently in comparison. The proportion of time spent performing different behaviours varied between individuals and not all sharks performed all behaviours.

Table 2.4 Ethogram of distinct behaviours observed in accelerometer- equipped semi-captive juvenile lemon sharks (*N.brevirostris*) (*n* =4), showing the number of per second measurements and percentage occurrence of each behaviour for each shark

	Behaviour											
Shark ID	**Burst**	**%**	**Chafe**	**%**	**Headshake**	**%**	**Rest**	**%**	**Swim**	**%**	**Total**	**%**
1	4	0.01	139	0.47	57	0.19	82	0.28	29406	99.05	**29688**	**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**	**35003**	**97.57**	**35875**	**100.00**

Upon visual inspection of the waveform signal for dynamic sway acceleration for the five behaviours observed in lemon sharks, bouts of constant amplitude and frequency were indicative of swimming and resting, with higher positive and negative acceleration (g) values for swimming. Rest behaviour, defined as motionless behaviour, corresponds to acceleration values of near- zero. Higher signal amplitude and frequency was observed for chafe, burst and headshake behaviours, with headshake behaviours typically higher amplitudes for longer durations than burst and chafe (Figure 2.3). This visual inspection revealed obvious differences in waveform characteristics for each behaviour depending on the sampling frequency (Figure 2.3). Comparison of one headshake event between the original sampling frequency (30 Hz) and re-sampled frequency of 5 Hz reveals how the underlying acceleration signal is distorted and misrepresented as sampling frequency is reduced (Figure 2.4).

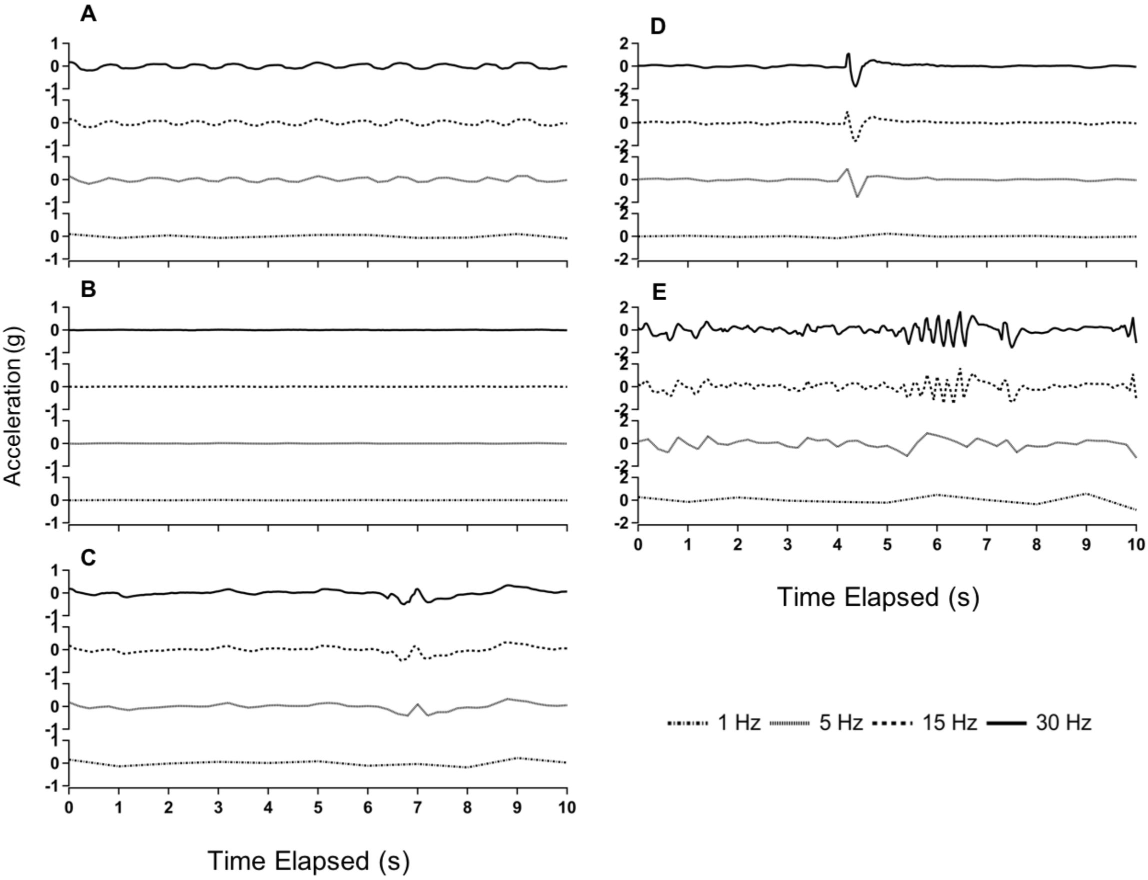

Figure 2.3 Examples of different representative acceleration plots for the dynamic sway (Z) acceleration for behaviours observed from a juvenile lemon shark (*N.brevirostris*). Raw acceleration data were resampled to show the representative change in acceleration waveform signal amplitude and frequency as sampling frequency is reduced for five observed behaviours (A) swim, (B) rest, (C) chafe, (D) burst and (E) headshake.

By exploring the same single headshake event, the effect of time-series averaging the accelerometer data into fixed one second time segments for each sampling frequency is revealed (Figure 2.5). The frequency and amplitude of the oscillations in the underlying acceleration signal are distorted across all sampling frequencies, causing similarities in the signal from which the predictor variables are extracted.

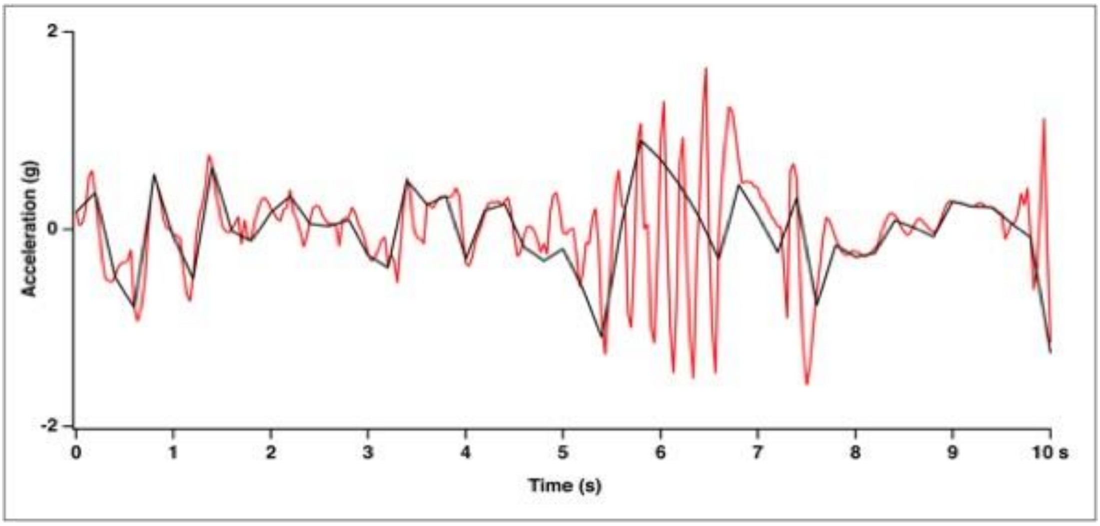

Figure 2.4 The higher frequency signal (red line, 30 Hz) becomes distorted at a lower sampling frequency (black line, 5 Hz). Decreasing sampling frequency decreases the number of data points in a given time. A higher frequency provides better representation of the underlying signal. For headshake behaviour in a juvenile lemon shark (*N.brevirostris*), the frequency of oscillations in the dynamic sway acceleration is less evident when sampling frequency is decreased.

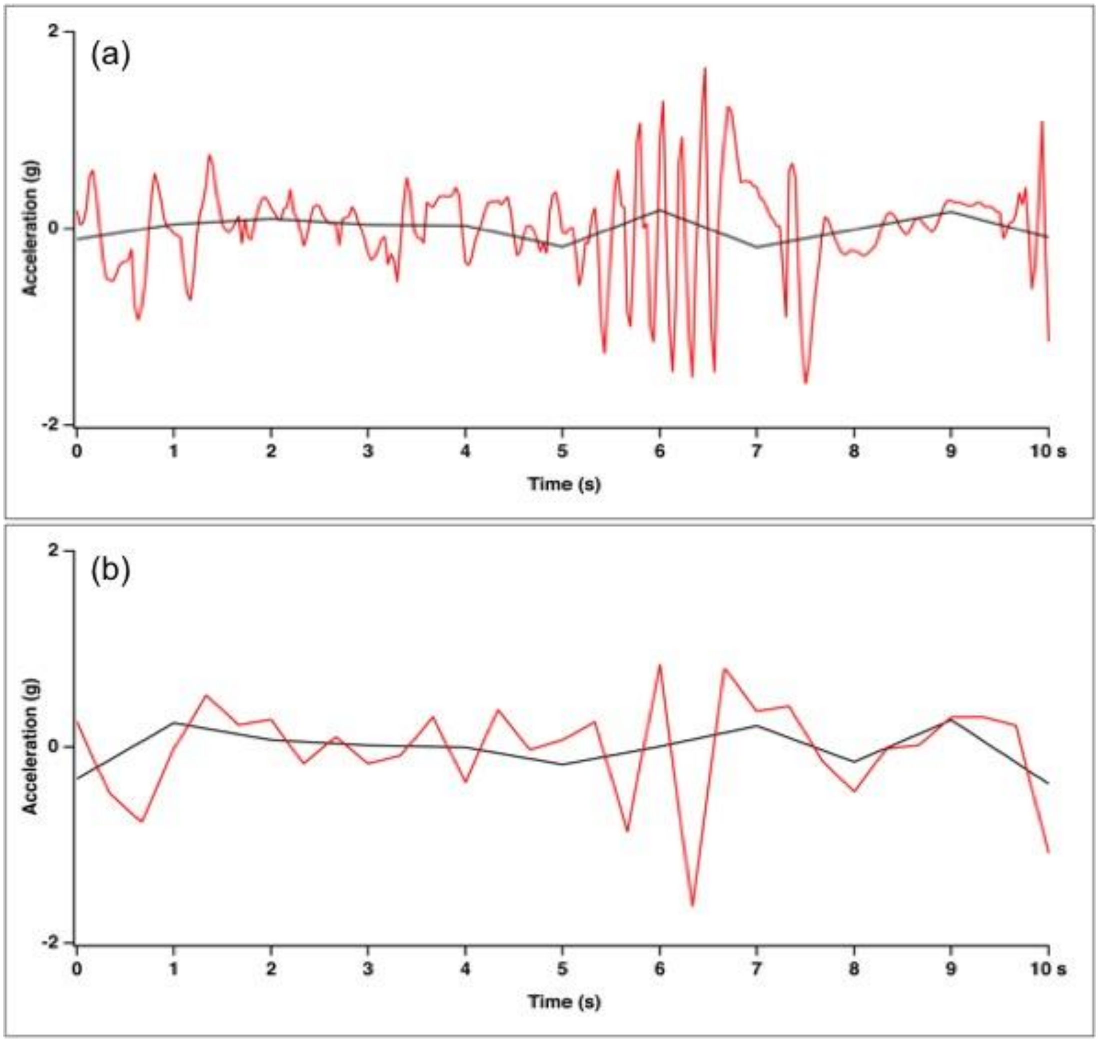

Figure 2.5 Time series averaging of accelerometer data into one second segments from which predictor variables are extracted has a distortive effect on the underlying signal. In this example for headshake behaviour in a juvenile lemon shark (*N. brevirostris*), the signal for dynamic sway acceleration at (a) 30 Hz (red line) becomes distorted when the data points within each second are averaged (black line). When sampling frequency is reduced to (b) 3 Hz the signal frequency of oscillations in the dynamic sway acceleration (red line) is notably decreased, however the time series average is similar to that for 30 Hz.

2.3.2 Overall classifier performance

Overall model accuracy for all sampling frequencies tested was high (>99 %) with a decrease in accuracy as sampling frequency was reduced below 3 Hz (Table 2.5). Results indicate that 30 Hz is the best sampling frequency for overall classification on the balanced data. The overall macro averaged F_1 score (F_M) for each sampling frequency was considered the most appropriate evaluation metric for such an imbalanced data set, as it combines F_1 score for all classes of behaviour to give a single value describing classifier performance across all classes. F_M was highest at 30 Hz and decreased when sampling frequency was reduced to 5 Hz or less (Figure 2.6). Little decrease in overall predictive power was observed until sampling frequency was reduced below 5 Hz (Figure 2.6). A total of 10,682 behavioural events were classified correctly at 30 Hz, whereas at 1 Hz 10,375 events were classified correctly (Table 2.6).

Table 2.5 Overall RF classifier accuracy for down sampled accelerometer data. Accuracy decreased more rapidly when sampling frequency was reduced to 3 Hz or below.

Sampling Frequency (Hz)	Overall Accuracy (%)
30	99.26
15	99.16
10	99.16
5	99.09
3	98.95
1	96.40

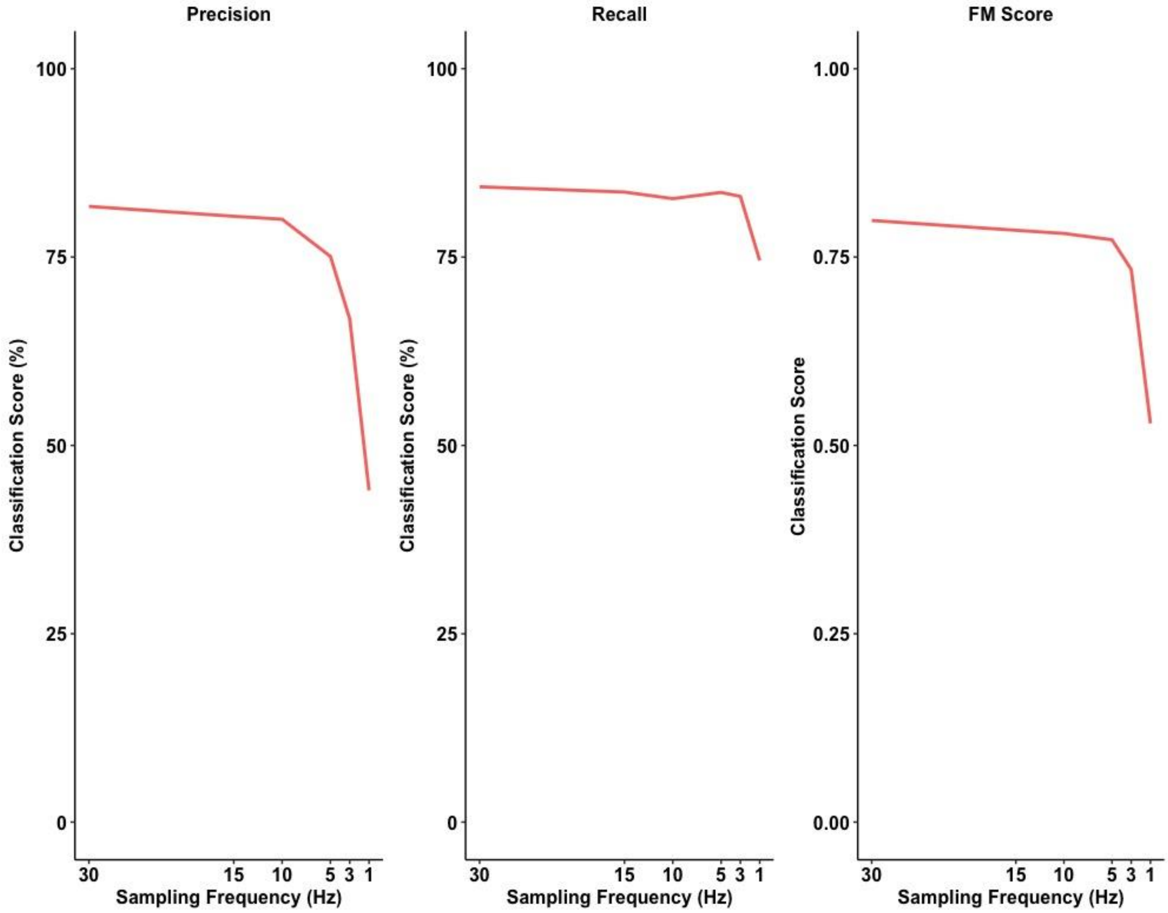

Figure 2.6 Overall RF model classification performance. As sampling frequency decreases, classification performance of the RF model decreases in all evaluation metrics except for a small peak in recall at 3 Hz and 5 Hz. Classification performance scores are good until sampling frequency is reduced to 5 Hz or below (> 75% precision, > 80% recall, > 0.75 F_M score). Precision decreases more than recall when sampling frequency is reduced. The F_M score reflects a combination of the change in recall and the more pronounced effect of reducing sampling frequency on precision.

2.3.3 Classifier performance for behaviour classes

Performance for individual behavioural classes varied according to both evaluation metric and sampling frequency (Figure 2.7; Table 2.6). Events from the swim and rest classes had the highest classification scores at all sampling frequencies; chafe, burst and headshake yielded lower classification scores overall (Table 2.6). The lowest scoring class was headshake (Figure 2.7).

The classification of events from the swim class was highest for all sampling frequencies (F_1 >0.982; Figure 2.7). As sampling frequency decreased, precision remained high (>0.998), however recall decreased very slightly when sampling frequency was below 5 Hz, due to swim events being increasingly labelled as other behaviour classes (Table 2.6).

Resting behaviour was also classified well by the model until sampling frequency was reduced to below 5 Hz (Figure 2.7). The highest precision and recall values were obtained at 5 Hz (0.931 and 1 respectively). Recall was >95 % for all sampling frequencies, whilst precision suffered at sampling frequencies below 3 Hz due to the model incorrectly predicting swim events from rest events. Rest events were never misclassified as chafe, burst or headshake events (Table 2.6).

The model classified chafe behaviours with high precision (≥0.760) at 30 Hz, decreasing with sampling frequency (2.7). Chafe behaviour was classified with high recall (>80 %), peaking at 3 Hz (0.963). This was at the expense of decreased precision at 3 Hz (0.690). The F_1 score for chafe events did not reflect this peak in recall due to the decrease in precision at 3 Hz (Figure 2.7). At 1 Hz, the model incorrectly classified more chafe events as swim events (Table 2.6).

Burst behaviour obtained precision scores as high as swimming when sampling frequency was ≥10 Hz (Figure 2.7). As sampling frequency was reduced <10 Hz, precision decreased, to 0.054 at 1 Hz (Figure 2.7, Table 2.6). This was due to the model incorrectly labelling events from the swim class as burst (Table 2.6). Lowest recall of all the classes was obtained for burst events regardless of frequency, with highest recall for this behaviour class achieved at 5 Hz (0.571). Low recall scores were due to the model incorrectly labelling burst events as chafe or headshake events even at 30 Hz. F_1 scores for burst events ranged from a minimum of 0.086 (1 Hz) to a maximum of 0.667 (≥5 Hz; Figure 2.7).

Headshakes achieved the lowest class performance of all behaviour classes across all sampling frequencies, with the highest F_1 at 30 Hz (0.535). Whilst recall was >0.672 at all sampling frequencies, this was at the expense of increasingly poor precision (Figure 2.7). Recall was lowest at 5 Hz due to incorrectly labelling headshake events as chafe events. Headshake scored the lowest precision of all behavioural classes (Figure 2.7). All classes except rest contained events that were incorrectly labelled as headshakes across all sampling frequencies (Table 2.6), with the most common false positive attributed to the swim class. However, the model classified headshake events with improved recall at both 3 Hz and 1 Hz (Figure 2.7). The model did not predict a burst event from the headshake class until sampling frequency was reduced to 1 Hz.

Classification performance for the minority classes (e.g. headshake and burst) was not significantly improved over using the original imbalanced training data set, however, the reduction in computational time was significant.

Table 2.6 Confusion matrix and performance metrics of RF model generated from the test set of ground- truthed data for each sampling frequency. Rows indicate actual observations and columns represent model predicted behaviours. Highlighted values show observations which were correctly classified by the model (TP; True Positives).

	SAMPLING FREQUENCY		PREDICTED BEHAVIOUR					PERFORMANCE METRICS			
		Class	Burst	Chafe	Headshake	Rest	Swim	Precision	Recall	F1	FM
ACTUAL BEHAVIOUR	30Hz	Burst	7	3	4	0	0	1.000	0.500	0.667	0.799
		Chafe	0	77	2	0	2	0.762	0.951	0.846	
		Headshake	0	7	27	0	0	0.403	0.794	0.535	
		Rest	0	0	0	129	3	0.921	0.977	0.949	
		Swim	0	14	34	11	10442	1.000	0.994	0.997	
	15Hz	Burst	7	3	4	0	0	1.000	0.500	0.667	0.786
		Chafe	0	76	2	0	3	0.752	0.938	0.835	
		Headshake	0	8	26	0	0	0.347	0.765	0.477	
		Rest	0	0	0	130	2	0.922	0.985	0.952	
		Swim	0	14	43	11	10433	1.000	0.994	0.997	
	10Hz	Burst	7	3	3	0	1	1.000	0.500	0.667	0.781
		Chafe	0	76	2	0	3	0.745	0.938	0.831	
		Headshake	0	9	24	0	1	0.333	0.706	0.453	
		Rest	0	0	0	132	0	0.923	1.000	0.96	
		Swim	0	14	43	11	10433	1.000	0.994	0.997	
	5Hz	Burst	8	4	1	0	1	0.800	0.571	0.667	0.773
		Chafe	0	76	2	0	3	0.710	0.938	0.809	
		Headshake	0	10	23	0	1	0.315	0.676	0.430	
		Rest	0	0	0	132	0	0.930	1.000	0.964	
		Swim	2	17	47	10	10425	1.000	0.993	0.996	
	3Hz	Burst	7	4	1	0	2	0.389	0.500	0.438	0.734
		Chafe	1	78	0	0	2	0.690	0.963	0.804	
		Headshake	0	5	24	0	5	0.393	0.706	0.505	
		Rest	0	0	0	131	1	0.868	0.992	0.926	
		Swim	10	26	36	20	10409	0.999	0.991	0.995	
	1Hz	Burst	3	5	3	0	3	0.054	0.214	0.086	0.529
		Chafe	0	67	1	0	13	0.409	0.827	0.547	
		Headshake	3	2	26	0	3	0.274	0.765	0.403	
		Rest	0	0	0	126	6	0.468	0.955	0.628	
		Swim	50	90	65	143	10153	0.998	0.967	0.982	

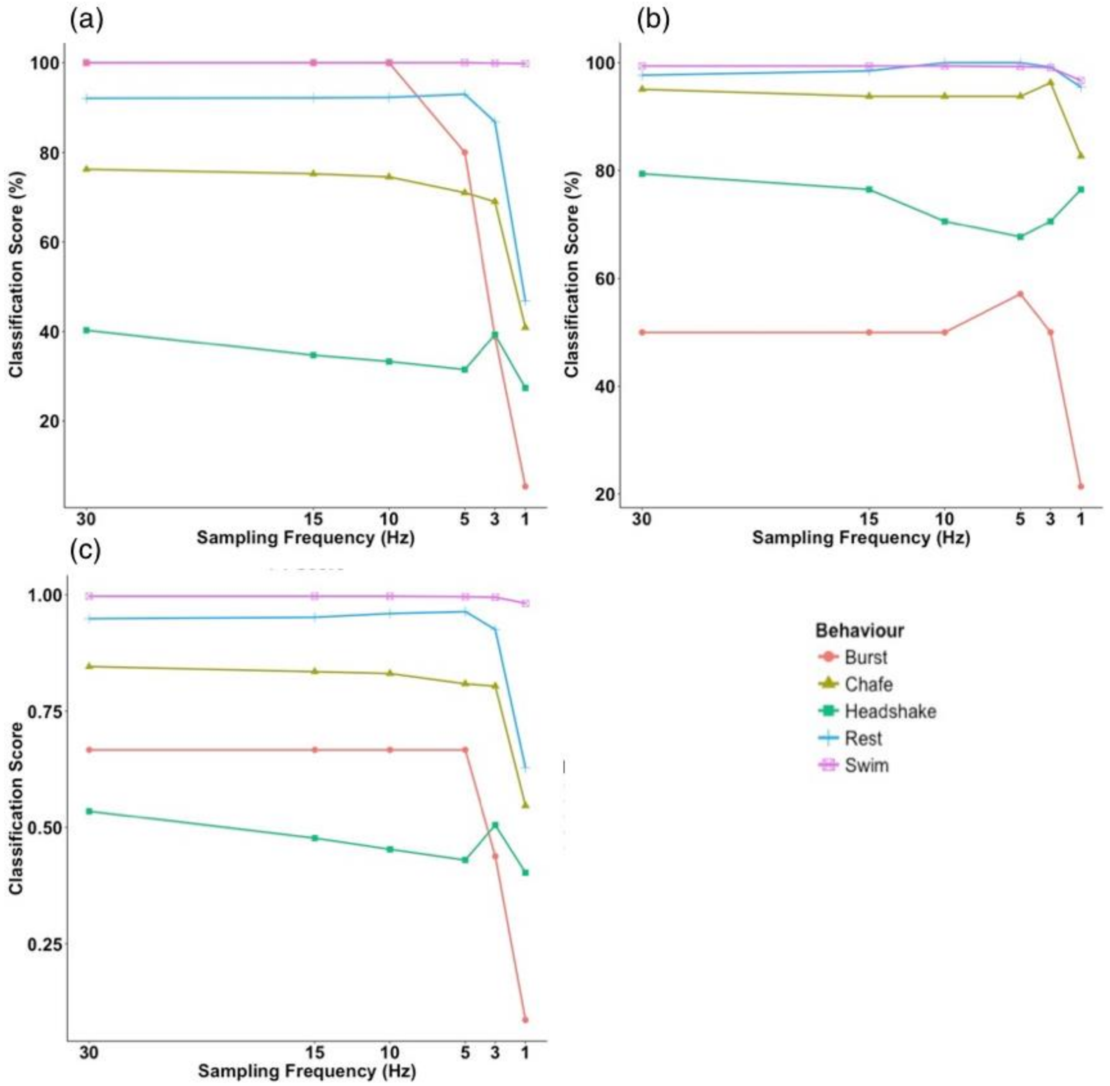

Figure 2.7 RF model performance for classification of behaviour of lemon sharks (*N.brevirostris*) (*n* =4) at different sampling frequencies. Performance is described by (a) Precision, (b) Recall and (c) F_1 Score for each observed behaviour from captive trials (burst, chafe, headshake, rest, swim). Performance varies by evaluation metric and by sampling frequency for each behavioural class.

2.3.3 Importance of predictor variables

The relative importance of the different predictor variables (refer to Supplementary Material, Table S1 for descriptions and explanations) varied between sampling frequencies. Amplitude was the most important for classification accuracy at all sampling frequencies except for 3 Hz, where Min_Y_Stat was the most important. The most important predictor variables for the Gini-Index were Avg_ODBA and max_ODBA for all sampling frequencies (Supplementary Material, Figure S1).

2.4 Discussion

To my knowledge this is the first study to evaluate the effects of sampling frequency from an accelerometer on behaviour classification in elasmobranchs. This study demonstrates the qualitative differences resulting from various sampling frequencies, and their effect on classifying lemon shark behaviour. Overall, best classification of lemon shark behaviour was achieved at 30 Hz ($F_M > 0.79$), which is comparable to results in sheep (32 Hz; Walton *et al.* 2018). However, optimisation of sampling frequency revealed that classification performance does not begin to drastically decrease until sampling frequency is decreased to 5 Hz, suggesting that this is an appropriate frequency for overall behavioural classification in lemon sharks of this size. Whilst 20 Hz was an appropriate sampling frequency for classification of flight behaviours in eagles (Sur *et al.* 2017), appropriate sampling frequency is scale dependant with regard to body size (Whitney *et al.* 2012, Brown *et al.* 2013). Broell *et al.* (2013) indicated that for smaller teleost fish, adequate sampling frequencies are usually above 32 Hz (Kawabe *et al.* 2003, Tsuda *et al.* 2006), whilst lower sampling rates are adequate for larger sharks who display lower frequency body movements (Whitney *et al.* 2007, Gleiss *et al.* 2009b, Whitney *et al.* 2010, Gleiss *et al.* 2011a), supporting the results presented in this study.

2.4.1 Classifier performance for individual behaviour classes

Whilst overall classification performance decreases as sampling frequency decreases, interpreting classification performance in terms of the specific behaviours of interest reveals more about the effects of sampling frequency. In this instance, performance of the RF model is dependent upon the complexity of the behaviour to be classified (as per McClune *et al.* 2014). The model classified swim and rest behaviours with high performance at all sampling frequencies ($F_1 > 0.9$ above 1 Hz). Active (such as swimming) and inactive (resting) behaviours can be successfully differentiated at a range of sampling frequencies, with 25 Hz having previously allowed for classification of these behaviours in teleost fish

(Brownscombe *et al.* 2014). Other studies have also had high success rates in identifying active and non-active behaviours in several elasmobranch species, including white tip reef sharks (Whitney *et al.* 2007) and lemon sharks (Gleiss *et al.* 2009a). Considering again that appropriate sampling frequency is scale dependant, it makes sense that positive identification of swimming and resting in large (TL >2.5 m) nurse sharks was also achieved at a lower sampling rate (5 Hz) than of teleost fish (Whitney *et al.* 2010), although classification in this study was a combination of both supervised and unsupervised learning, using data that had not been ground- truthed and is therefore not comparable to the supervised learning approach in this current study.

In contrast to the classification of swim and rest behaviour, the RF model struggled to classify more complex behaviours such as burst and headshake, with highest classification scores at 30 Hz ($F_1 \approx 0.70$ and 0.50 respectively). Broell *et al.* (2013) found a low overall accuracy of 60% for fast start and feeding events in great sculpin (*Myoxocephalus polyacanthoceaphalus*) at sampling rates between 20–30 Hz, concluding that high sampling frequencies are required to successfully classify these behaviours. This is demonstrated by (Horie *et al.* 2017) where sampling frequencies of 200 Hz were used to successfully identify feeding behaviours in red-spotted groupers (*Epinephelus akaara)* (F_1 >0.7). In addition, flight behaviours more complex than flapping and soaring in eagles were also poorly distinguished by a RF model even at the highest sampling frequency tested (40 Hz) (Sur *et al.* 2017). The overall objective of this analysis however was not to design a classifier best capable of distinguishing these behaviours. Whilst super-learners and ensemble classifiers (Ladds *et al.* 2017, Brewster *et al.* 2018) combine any number of base learners to improve classifier ability for distinguishing behaviours of interest, the purpose in this case was to show how classifier ability changes with sampling frequency. Although many ML models are suitable for behavioural classification (Nathan *et al.* 2012), RF was chosen to demonstrate this effect due to its relatively simple implementation and widespread use

(Nathan *et al.* 2012, Wang *et al.* 2015, Ladds *et al.* 2017, Sur *et al.* 2017, Walton *et al.* 2018).

2.4.2 Misclassification due to signal aliasing

Misclassification of complex behaviours (such as burst and headshake) has been attributed to aliasing, which is a distortion effect occurring in acceleration waveform signals as sampling frequency is reduced, causing different acceleration signals to become indistinguishable (Mallat 1999, Broell *et al.* 2013). Typically, swimming in sharks is represented by regular cyclic patterns in the sway acceleration, with each oscillation in the waveform signal representing the individual tail-beats contributing to forward propulsion (Gleiss *et al.* 2009a, Gleiss *et al.* 2009b). The appropriate sampling frequency reveals the underlying original acceleration signal and the general guideline is that the selected sampling frequency must exceed the Nyquist criterion - at least twice that of highest frequency of movement being classified (Brown *et al.* 2013). If this criterion is not exceeded, the original acceleration signal of interest is aliased. Higher movement behaviours, such as burst and headshake, feature greater tail-beat movements that result in high frequency oscillations in the representative waveform signal (Figure 2.3). It is hypothesised that these original waveform signals are increasingly aliased as sampling frequency is decreased (see Figure 2.4). This aliasing causes misclassification of these behaviours as they become indiscernible from each other (see Figure 2.3). This concept is demonstrated by the model's incorrect predictions of headshake events from all behaviour classes except rest, a behaviour that does not feature distinct tail-beat movements.

2.4.3 Misclassification of behaviours with similar acceleration characteristics

Poor classification of burst and headshake behaviours is due to acceleration waveform signals having very similar characteristics (McClune *et al.* 2014, Ladds *et al.* 2017, le Roux *et al.* 2017, Walton *et al.* 2018). Both of these behaviours in lemon sharks are characterised by high frequency and amplitude acceleration signals, from which the predictor variables are extracted for training the RF model. For example, burst events were misclassified as both chafe and headshake events even at 30 Hz; increasingly so as sampling frequency was decreased, and signals became more similar due to aliasing. Misclassification was also high between behaviours with similar acceleration signal characteristics in cats (Watanabe *et al.* 2005), sheep (le Roux *et al.* 2017, Walton *et al.* 2018) and seals (Ladds *et al.* 2017). Alternatively, misclassification could have been due to the variability within and inconsistencies between behaviour classes (Walker *et al.* 2015). Variable swimming speeds and postures were observed for similar behaviour classes, both within and between individual sharks which may have caused confusion within the RF model. As an aside, noise in all acceleration axes may be generated by different movements if the tag package is not securely attached to the animal (Shepard *et al.* 2008b). This may have a larger effect in behaviours accompanied by more vigorous body movements as observed in burst, chafe and headshake (Walton *et al.* 2018).

2.4.4 Misclassification of rare behaviours

In addition to the effects of sampling frequency, misclassification may be linked to the rarity of the burst and headshake behaviour events for training and testing the RF model (Shamoune-Baranes *et al.* 2012, Brown *et al.* 2013). Although little loss of overall predictive power was observed when classifying behaviour 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). For this study the behaviour class size imbalance was managed by selectively down sampling the majority behaviour class (swim), however there remained a limited number of observations for burst (*n* =48) and headshake (*n* =113) behaviour classes for training and testing the model. The effect of just one mislabelled data point therefore is much larger for these behaviour classes than for the behaviour classes which contain many thousands of observations (e.g. swim). The small class sizes for these behaviours were a direct result of the lengthy time required to obtain data for these behaviours from captive lemon sharks. As such, a larger sample size was not attainable, as per Brewster *et al.* (2018). For future studies, the ability to ground-truth a larger number of measurements for burst and headshake behaviours may result in improved classification performance across all sampling frequencies.

2.4.5 Effects of time series averaging

When interpreting model performance for more complex behaviours, it is important to note the peak in performance (described by the harmonic mean of precision and recall; F_1) at 3 Hz for headshake behaviours. This result was comparable to the classification of down sampled accelerometer data from eagles (Sur *et al.* 2017) where the decrease in RF classification performance was not linear with decreased sampling frequencies; instead, overall RF classifier performance, described by accuracy, was better at 10 Hz than at 20 Hz and 40 Hz. Again, this could be due to the distortive effects of aliasing (Figure 2.4). It is also possible that this could be related to the methodological effect of averaging the data into one second time series to be matched with behavioural observations (Figure 2.5). Behaviours characterised by high amplitude acceleration signals, such as burst or headshake, contain more extreme positive and negative value acceleration data points than swim or rest. Predictor variables for input to the RF model are extracted from the per second time series average. Each sampling frequency captures different acceleration data points after decimating the underlying signal, therefore the time series average for each sampling frequency also differs. This may have a more pronounced effect at low sampling frequencies if different extreme positive or negative data points remain in the decimated acceleration signal.

2.4.6 Evaluation metrics and predictor variables

The benefits of using the detailed evaluation metrics calculated from the confusion matrix (precision, recall, F_1 and F_M, previously described in full) to interpret RF model performance, rather than relying on accuracy as a performance descriptor for imbalanced data sets is highlighted by this study. Overall accuracy was above 96% for all sampling frequencies. Reducing the size of the majority classes (swim and rest) by using a stratified subset of the data made some attempt to overcome the imbalance at a data level, yet despite this the imbalance between the most common and rare classes was still evident in the high overall model accuracy. There are less naïve algorithmic level methods for dealing with imbalanced data sets. These methods either assign weight to the minority class (weighted RF) or alter class distribution (balanced RF), yet neither can yet be implemented in R software (Chawla *et al.* 2002, Chawla *et al.* 2003, Chen *et al.* 2004). This issue with accuracy was further overcome in this study by interpreting classifier performance from a detailed perspective by using F_M (overall) and F_1 (class) scores instead of accuracy alone.

Another perspective for describing RF model performance is by assessing the importance of the predictor variables to model accuracy. The importance of specific predictor variables could be due to complex interactions with other predictor variables (Liaw and Wiener 2002). Some behaviours may be identified by other features of acceleration in addition to the changes in tails beats which correspond to dynamic acceleration. In addition to dynamic acceleration in the sway axis, mating in mature nurse sharks was also characterised by changes in body orientation, represented by static acceleration in the pitch and sway axes (Whitney *et al.* 2010). This is where incorporation of other movement sensors that better measure body orientation might aid classification. Moreover, increasing the number of predictor variables improves model complexity and classification (Huynh *et al.* 2007, Nathan *et al.* 2012). This could help explain the very low classification performance at 1 Hz, because there were predictor variables removed from this data set due to the RF models' inability to handle n/a values. When the sampling frequency was decreased to 1 Hz, there was only one acceleration measurement per second, therefore there was no value for predictor variables such standard deviation, skewness or kurtosis.

2.4.7 Recommendations

This study has established that 5 Hz is an appropriate sampling frequency for successfully classifying swimming, resting and chafing behaviour in lemon sharks. Whilst classification performance was lower even at 30 Hz for both burst and headshake behaviour, there was no significant decrease in classifier performance for these behaviours until sampling frequency was less than 5 Hz. It is proposed that for the best possible determination of fine-scale complex behaviours, 30 Hz should be considered the minimum sampling frequency when programming devices. It is possible that detection of these behaviours may benefit from sampling frequencies >30 Hz, however such frequencies will further significantly reduce deployment durations. The ability to remotely classify lemon shark behaviours and study activity levels related to foraging behaviour and energy expenditure could greatly impact future conservation and management strategies. Fast swimming behaviours have been associated with predator avoidance and prey capture in fish (Domenici and Blake 1997), with headshaking behaviour linked to successful feeding events in sharks (Motta *et al.* 1997, Brewster *et al.* 2018). For this study, although the main aim was optimisation of sampling frequency, it is unfortunate that the model was not more successful in identifying fast swimming behaviours (burst and headshake) in lemon sharks. The VE classifier developed from this data set by Brewster *et al.* (2018) would improve the classification of these behaviours across all sampling frequencies and should be considered for a thorough demonstration of the effects of changing sampling frequency.

The results presented here are significant in terms of allowing future studies to set appropriate sampling frequencies for classifying lemon shark behaviours and potentially extending the duration of future deployments. Programming accelerometer devices at the lowest frequency possible for behaviour classification could drastically reduce the requirements for memory capacity, battery size and data processing time. However, a delicate trade-off was found between improving classification performance of behaviours in sheep and extending device battery life (Walton *et al.* 2018). Battery life was halved when sampling frequencies were increased from 16 Hz to 32 Hz 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), suggesting that any increase in battery capacity is not worth the minimal improvement in classifier performance. The effects of sampling frequency on the storage capacity of on board memory cards and processing time, to my knowledge, has not yet been fully evaluated. Technological limitations remain, meaning that the increased sampling frequencies recommended for the determination of complex behaviours in juvenile lemon sharks do not allow for deployments for more than a few days.

For future studies, there are a number of methodological factors that could be incorporated to improve classification of burst and headshake behaviours. A larger sample size for captive trials may increase the number of observations for more rare events (Wang *et al.* 2015). Increasing the number of predictor variables by selecting those related to the individual shark (e.g. sex, length, weight) or the environment (e.g. depth, temperature, diel cycle, tidal phase, season) may improve classification as these are all factors which affect the behaviour of lemon sharks (Morrissey and Gruber 1993a, Sundström *et al.* 2001, Guttridge *et al.* 2009, Lear *et al.* 2017, Leos-Barajas *et al.* 2017, Bouyoucos *et al.* 2018). Furthermore, the incorporation of other movement sensors alongside accelerometers, such as gyroscopes and magnetometers, has previously allowed for successful discrimination and classification of more complex behaviours in fish, turtles and birds (Noda *et al.* 2012, Noda *et al.* 2013, Kawabata *et al.* 2014, Noda *et al.* 2014, Williams *et al.* 2015, Williams *et al.* 2017).

2.5 Conclusion

In conclusion, this chapter provides a useful guide for future programming of devices for improving behavioural classification from accelerometer data for lemon sharks. Additional observations are required for more successful predictions of fine scale burst and headshake behaviours across all sampling frequencies. It is expected that rapid advancements in device technology will allow for increased deployment durations at the highest sampling frequency recommended. However, using the technology currently available, considering the optimum sampling frequency can help to optimise both recording duration and future classification of behaviour from lemon sharks.

Chapter 3. Evaluating the impact of movement sensors for the classification of shark behaviour

3.1 Introduction

Bio-logging tools such as accelerometers are useful for elucidating behaviours in free-ranging animals (Campbell *et al.* 2013). The use of only one sensor however, may not be the most beneficial for successful identification of behaviour (Noda *et al.* 2012, Ropert-Coudert *et al.* 2012, Noda *et al.* 2014). Whilst accelerometers are capable of measuring up to three-dimensional body acceleration (Wilson *et al.* 2006, Shepard *et al.* 2008b) they do not provide information regarding angular rotation or angular velocity (turning direction and turning speed respectively), which are important components of dynamic movements of animals (Noda *et al.* 2013). The potential inclusion of the movement sensors that measure body orientation separately from acceleration is significant for some fish behaviours, where the kinematics of the associated movement are important. These behaviours include fast-start behaviours such as feeding in white-streaked grouper *Epinephelus ongus* (Kawabata *et al.*, 2014) and single or double-bended escape movements, described by Domenici and Blake (1991), in Japanese amberjack *Seriola quinqueradiata* (Noda *et al.* 2014). For sharks, changes in body orientation (represented by static acceleration) has been used to distingusih behaviours including mating in nurse sharks (Whitney *et al.* 2010), diving in whale sharks (Gleiss *et al.* 2011a) and rolled swimming in hammerhead sharks (Payne *et al.* 2016). Presumably therefore, behavioural classification improves when accelerometers are combined with other movement sensors capable of more accurate calculations of orientation, such as a magnetometer (Williams *et al.* 2015, Williams *et al.* 2017) or a gyroscope (Noda *et al.* 2012, Noda *et al.* 2013, Kawabata *et al.* 2014, Noda *et al.* 2014).

Whilst behavioural classification could be improved by incorporating multiple sensors into one device, the collection of high resolution data from multiple channels introduces some problems (Walker *et al.* 2015). Recording high resolution data in three tri-axial sensors means that device memory and battery requirements are increased (Ropert-Coudert *et al.* 2012, Bidder *et al.* 2014, Noda *et al.* 2014). Consequences of this include reduced recording duration, precluding time scales that are biologically meaningful to sharks (e.g. diel and lunar cycles, seasons) (Guttridge *et al.* 2009). Here, the second objective of this thesis is addressed; I evaluate how three different tri-axial movement sensors (or combinations of these sensors) impact the performance of behavioural classification.

3.2 Method

The data set used for this chapter was collected through funding by the Save Our Seas Foundation. All data were collected in accordance with the ethical standards of the institution or practice at which the studies were conducted (Murdoch University Animal Ethics Committee, Permit Number R2927/17).

3.2.1 Study Site and Study Species

The St Joseph Atoll is located in the outer Amirantes Islands of the Republic of Seychelles, (5.43°S, 53.35°E), east of Mozambique, Africa (Figure 3.1a). The remote atoll, to the east of D'Arros Island, is made up of a ring of small islands surrounding a central lagoon approximately 4 km by 7 km across (Figure 3.1b). The study was conducted predominantly in the south-central part of the atoll, where sand flats on the edge of the lagoon were exposed at low tides.

The sicklefin lemon shark (*Negaprion acutidens*) is a carcharinid elasmobranch found in sub-tropical and tropical waters. The sicklefin lemon shark is assessed as vulnerable by the IUCN, with populations suffering rapid declines due to heavy fishing within its narrow habitat range (Pillans 2003). Sicklefin lemon sharks are studied far less than the lemon shark (*N. brevirostris*), however this species was selected because site fidelity and abundance are high at the study site (Filmalter *et al.* 2013).

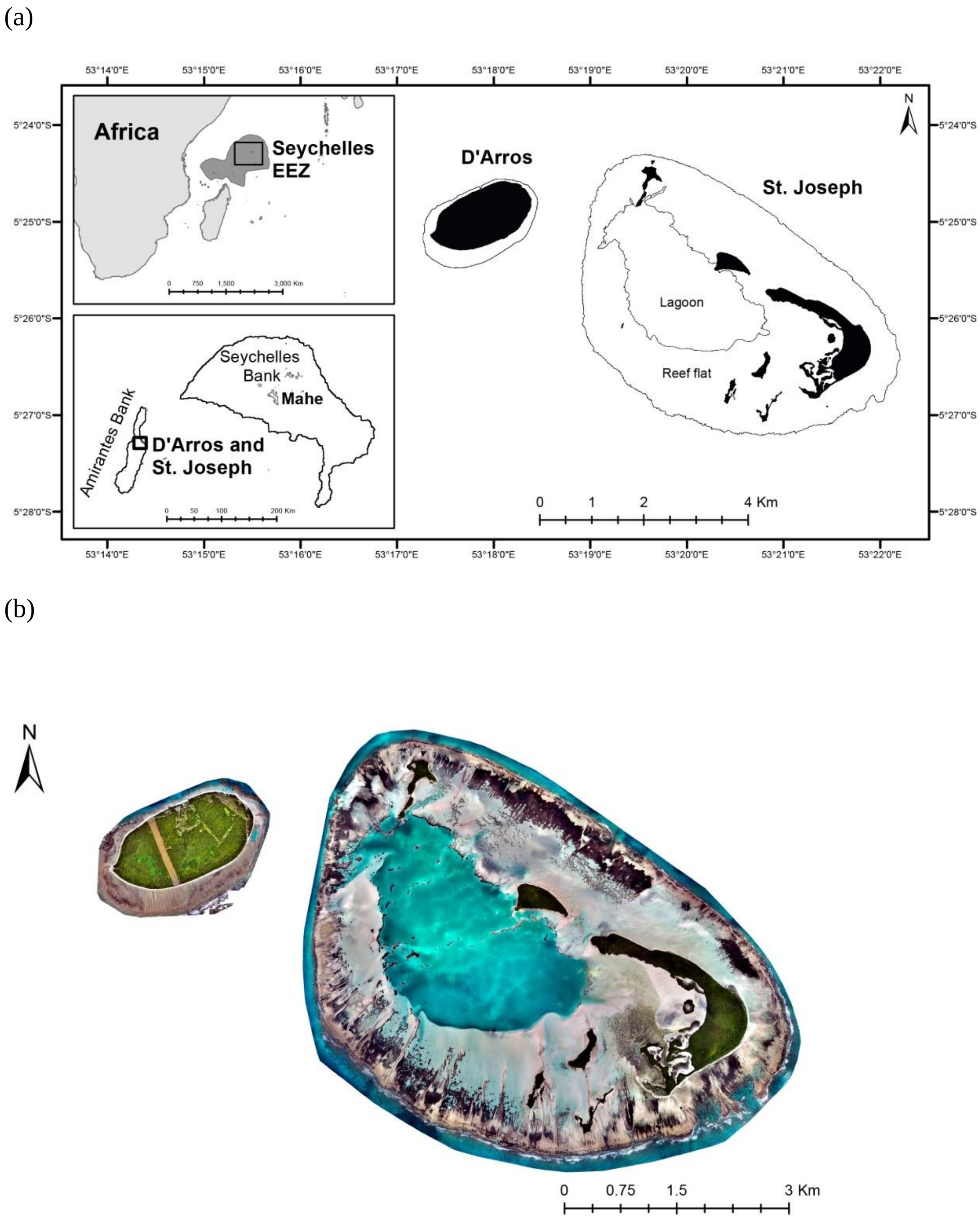

Figure 3.2 (a) Location of D'Arros Island and St Joseph Atoll, Republic of Seychelles. (b) The study site was located in the southern-central part of St Joseph Atoll, predominately on the sandflats at the edge of the lagoon (Credit: Ryan Daly, Save Our Seas Foundation, 2017).

3.2.2 Shark Capture

Sub adult sicklefin lemon sharks (*n* =4; Table 3.1) were captured using a baited hand line with a barbless circle hook and wire leader, cast from a small vessel. Once caught, sharks were brought to the side of the boat and safely restrained in a horizontal position, by holding the hand line at the bow of the boat and using a rope loosely placed around the caudal peduncle, then tied to the stern of the boat (Filmalter *et al.* 2013). Sharks were transported to a nearby purpose built rectangular pen (12 x 8 m) which was erected in close proximity to the capture location in the shallows on neighbouring sandflats. The TL of each shark was measured in centimetres and sex determined. Minimum TL was determined by size of the tag package to be attached to the first dorsal fin (Figure 3.2). There was no maximum TL, as it has been noted previously that sicklefin lemon sharks leave the atoll at sexual maturity, therefore most lemon sharks within the atoll are juvenile or sub-adult (Filmalter *et al.* 2013). The size range for captive sharks was 161–198 cm TL (Table 3.1).

Table 3.3 Sub-adult sicklefin lemon sharks (*N.acutidens*) used during semi-captive trials for development of an ethogram

Sex	Total Length (TL) (cm)
Female	198
Female	161
Male	183
Female	169

3.2.2 Tagging Procedure

Sharks were equipped with a CATS Diary tag (Customised Animal Tracking Solutions, www.cats.is). The multi-sensor data recorder was programmed to record acceleration, magnetic field strength and angular velocity at 20 Hz in three tri-axial sensors (accelerometer, magnetometer and gyroscope). The measurement ranges for each sensor were ±16 g, ±48 Gauss and ±2000 degrees second^{-1} respectively, with 16-bit resolution stored on an internal SD memory card. Tags were securely attached to the first dorsal fin on individual sharks using monofilament, plastic plates and crimps (Figure 3.2 and Figure 3.3), as per the method in Chapter 2.

Figure 3.3 The CATS Diary tag package was measured against the first dorsal fin of a sicklefin lemon shark (*N.acutidens*) caught by hand line and barbless hook in St Joseph Atoll. This was to assess suitability of the sharks dorsal fin size prior to tag attachment.

Figure 3.4 CATS Diary tag packages were attached to the first dorsal fin of subadult sicklefin lemon sharks (*N. acutidens*).

The accelerometer sensor measured body acceleration relative to gravity in three orthogonal axes; anterior-posterior (surge), dorso-ventral (heave) and lateral (sway). The magnetometer and gyroscope measured angular rotation and angular velocity about the same three axes respectively (Figure 3.4).

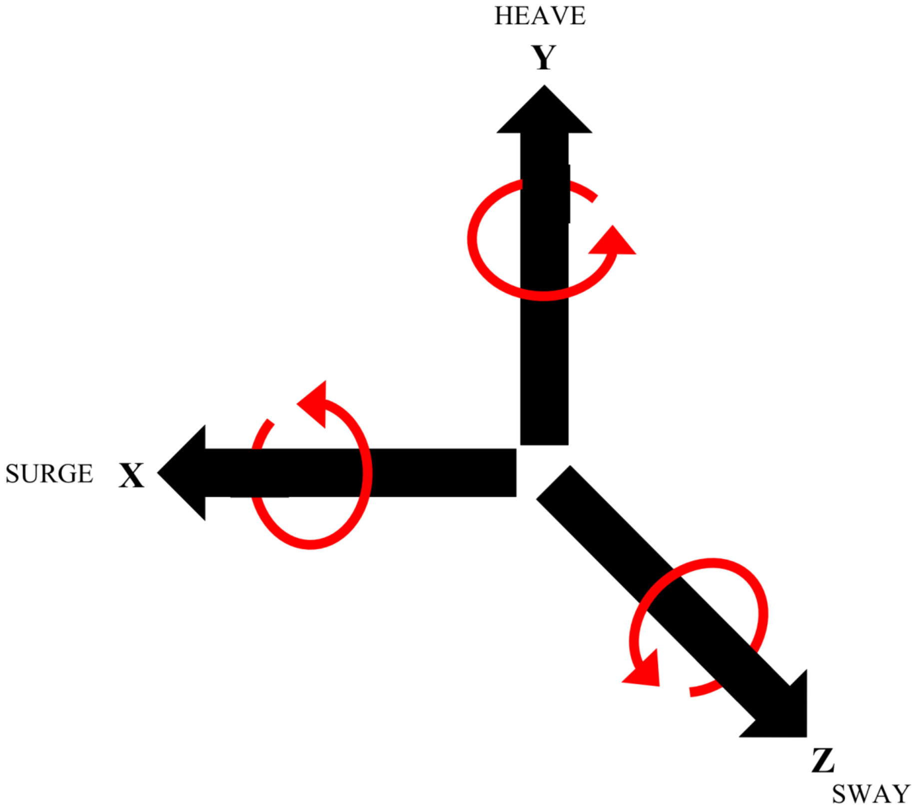

Figure 3.5 Schematic diagram of the three orthogonal axes measured by the accelerometer (black arrows) and the corresponding angular measurements (rotation and velocity) by the magnetometer and gyroscope respectively (red arrows).

3.2.3 Captive Trials and Ethograms

The holding pen for captive trials was constructed as per Brewster *et al.* (2018) and Chapter 2, ensuring that ambient environmental conditions were experienced by sharks during captive trials (Figure 3.5a). Tag packages were attached to sharks once within the pen, having allowed for initial capture recovery and acclimation to captivity. Acclimation was determined by regular, unstressed steady swimming movements and sharks eating, under their own volition, the thawed or fresh fish provided to them. Once acclimated, sharks were fed to satiation at libitum except during trials. To validate sensor data against behaviours, tagged sharks were directly observed during captive trials (ground- truthing). Sharks were observed from the slightly elevated sandbank adjacent to the pen (Figure 3.5b). Observations were supported by simultaneously recorded video footage captured from a hoisted camera (Hero3+ and Hero4, GoPro, www.gopro.com) on the sandbank and via unmanned aerial vehicle (UAV; Phantom 4, DJI, www.dji.com). Each second of the observation period was classified as one of four distinct behaviours: swimming, resting, burst swimming and headshaking (for definitions see Chapter 2, Table 2.1). These behaviours have been previously observed during captive trials with lemon sharks (Brewster *et al.* 2018). Again as per Chapter 2, rare behaviours such as burst were induced if they were not readily performed during captive trials to gain an adequate sample size for each observed behaviour. An ethogram of behaviours was produced (Sakamoto *et al.* 2009). Recorded observations included initiation and cessation times for each behaviour, using digital clocks that had been synchronised with tag packages during programming. Behavioural observations were subsequently time matched to the downloaded tag data to ground-truth the data for each shark. Observations were conducted across various tides and times of day, in varying weather conditions (sunny, overcast, raining). Observations were discounted from the labelled data set if the behaviour was likely to be a result of captivity and unlikely to be witnessed in the wild. For example, instances where the shark changed direction by turning 180 degrees alongside the net during steady swimming, or where one shark swam over another resting shark, potentially knocking the dorsally mounted tag. Following completion of the trials, sharks were re-captured inside the pen for removal of the tag package. Sharks were then monitored and released by opening a side panel of the pen.

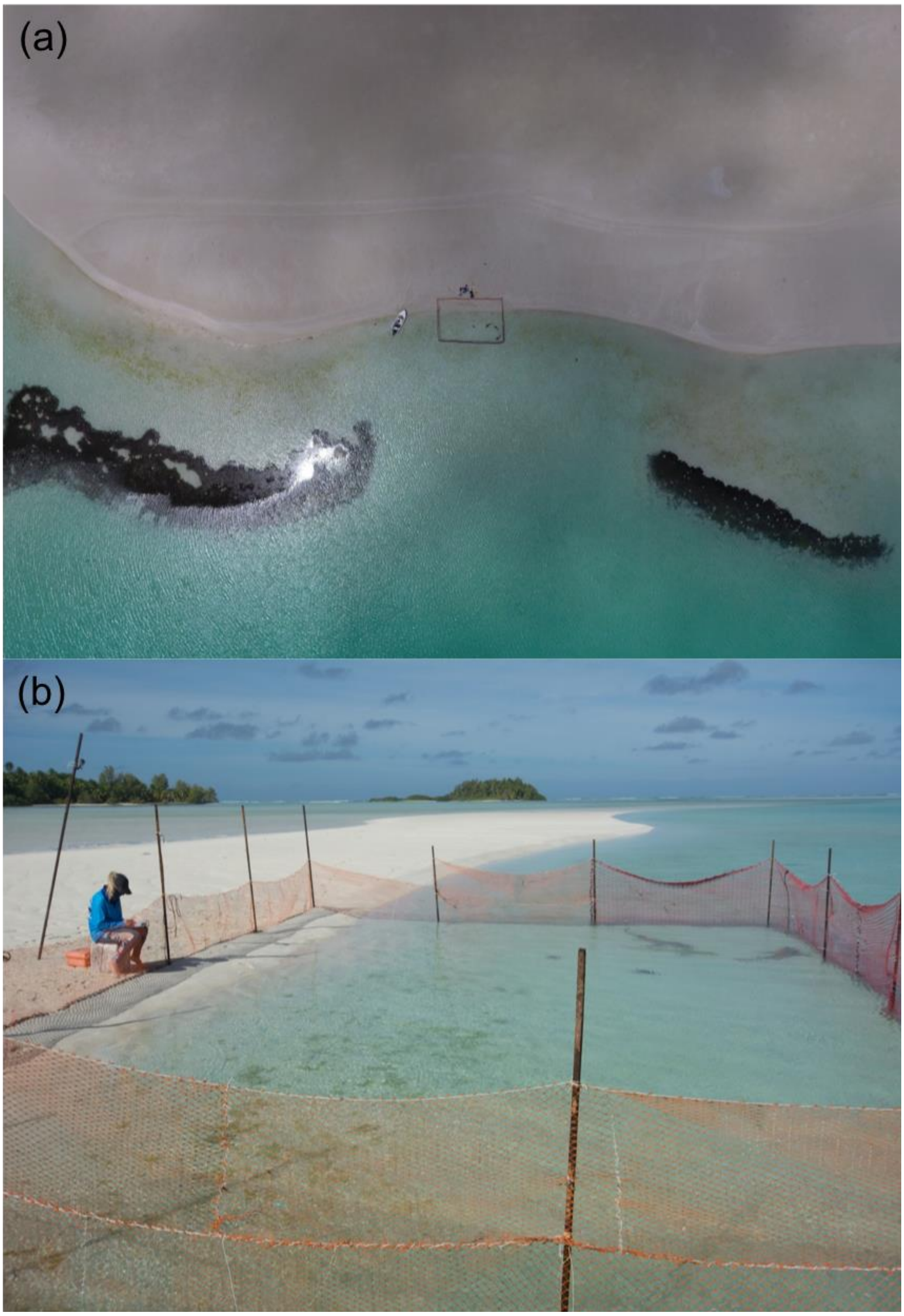

Figure 3.6 (a) Captive trials were conducted in a purpose built 12 x 8 m pen adjacent to a sandbank on the edge of the central lagoon in St Joseph Atoll (b) Direct observations during captive trials were conducted from the elevated sandbank and supported by video recorded from a mounted camera. (Photo credit Ryan Daly, Save Our Seas Foundation, 2017)

3.2.4 Data Analysis

Raw acceleration data was first analysed in IGOR Pro version 7.06 (WaveMetrics Inc, Lake Oswego, Oregon, USA). Static acceleration, representing body posture in relation to Earth's gravitational field, and dynamic acceleration, representing body movement, were separated in all three axes (x, y, and z) using a 3-second box smoother (Shepard et al., 2008) for both the accelerometer and the magnetometer. Overall dynamic body acceleration (ODBA) was then calculated by summing the absolute values of dynamic acceleration in each axis (Wilson et al., 2006; Shepard et al., 2008) for both the accelerometer and the magnetometer. The data from the gyroscope was treated as raw data in all three axes. Continuous wavelet transformation was used to derive waveform signal amplitude and frequency of the dominant cycle, from the dominant axis (representing lateral body movement) in each sensor to calculate TBF and tail-beat acceleration amplitude (Ethographer v2.0 package; Sakamoto *et al.* 2009). Predictor variables summarising the features of the data from each sensor were then extracted from fixed time series averages of one second segments (for method see Chapter 2; Supplementary Material Table S2, S3 and S4). Different combinations of predictor variables were then used to test which sensor or combination of sensors was best for the classification of behaviours for sicklefin lemon sharks (Table 3.2).

Table 3.2 Predictor variables were extracted from features of the data from each movement sensor. Different combinations of these predictor variables were used for training and testing the RF model.

Sensor Combination	**Number of predictor variables (*n*)**
Accelerometer	45
Magnetometer	39
Gyroscope	21
Accelerometer + Gyroscope	66
Accelerometer + Magnetometer	84
Accelerometer + Gyroscope + Magnetometer	105

3.2.5 Machine Learning and Random Forest

RF classification was again conducted using the ‘randomForest’ package in R (Liaw and Wiener 2002). The ground- truthed data were randomly allocated to train and test data sets, whilst maintaining class ratios, as per Chapter 2. Seventy per cent of the data were used for training the model, and the classification accuracy was cross validated and tested on the remaining 30 % of the data. The *ntree* value selected for the model was the same as Chapter 2 (*ntree*=1000). The number of predictor variables at each node, *mtry*, was again chosen by using the square root of the total number of predictor variables (*n*) as per Verikas *et al.* (2011). The total number of predictor variables varied with the sensor combination being tested (Table 3.2), therefore *mtry* varied accordingly with each iteration of the model.

RF models by nature aim to improve overall classification accuracy, yet this has a detrimental effect on the predictive performance for minority classes (Chen *et al.* 2004, Ganganwar 2012). As per Chapter 2, the data set was imbalanced, with the majority of the data labelled as swim (100096 observations) or rest (5612). The headshake and burst classes contained fewer observations as these behaviours occurred less frequently (45 and 125 respectively; see Table 3.3). Other studies have created balanced data sets where each behaviour is represented equally (le Roux *et al.* 2017), however these data sets included substantially more observations for the rarest behaviour class than in this data set. For example, Dutta *et al.* (2015) included 785 instances from each behavioural class to minimise the effects of bias in classification results. However, for this data set the minority class (headshake) contained only 45 instances. A balanced data set would not be reflective of the relative proportions of each behaviour as performed naturally in the wild. Because an RF model aims to replicate the relative class proportions contained in the training data, a model trained from a completely balanced data set may not classify new data with great accuracy. Using the same method as Chapter 2, a stratified subset of the training data set was incorporated into the RF model. The majority classes (swim and rest) were reduced to be closer in frequency to the minority class (headshake) by a factor of ten, by selective direct under-sampling in the training data set for each sensor combination.

RF model performance was evaluated by applying the model to the unseen test data (30 %) and generating a confusion matrix as per Chapter 2. Each RF model, representing the different sensor combinations, was trained and tested on the same data split to allow for direct comparison of performance. From the confusion matrix, evaluation metrics were calculated for the model overall and for individual behaviour classes, for each sensor combination. These included overall accuracy, recall, precision, F_1 score and macro-averaged F_1 (F_M). For full equations and descriptions of evaluation metrics refer to Chapter 2. Again, individual class F_1 score and F_M (overall class macro-averaged F_1 score) were considered the best descriptors of classification performance for this data set, especially where specific behaviours are of interest (Walton *et al.* 2018). The overall macro averaged F_1 score (F_M) combines class F_1 scores to give a single value describing classifier performance for each model (see Equation 5; Chapter 2).

3.3 Results

3.3.1 Ethogram and ground- truthed data

Multi-sensor data were collected for four sicklefin lemon sharks during captive trials. From these data sets, over 105,000 seconds of data were ground- truthed and labelled as one of four distinct behaviours (Table 3.3). Swimming and resting behaviours were observed most frequently, whilst burst and headshake behaviours were relatively rare. The proportion of time spent performing different behaviours varied between individual sharks and not all sharks performed all behaviours.

Table 3.3 Ethogram of distinct behaviours observed in sub adult sicklefin lemon sharks (*N.acutidens*) (*n* =4), showing the number of per second measurements and percentage occurrence of each behaviour for each shark

	Behaviour									
Shark ID	**Burst**	**%**	**Headshake**	**%**	**Rest**	**%**	**Swim**	**%**	**Total**	**%**
1	46	0.15	22	0.07	4208	13.34	27263	86.44	31539	**29.79**
2	19	0.08	0	0.00	222	0.88	25062	99.05	25303	**23.90**
3	19	0.06	12	0.04	313	1.07	28942	98.83	29286	**27.66**
4	41	0.21	11	0.06	869	4.40	18829	95.34	19750	**18.65**
Total	**125**	**0.12**	**45**	**0.04**	**5612**	**5.30**	**100096**	**94.54**	**105878**	**100.00**

Examples of each observed behaviour (swim, rest, burst and headshake) were visualised from the waveform signal for the dominant axis in each sensor (Figure 3.6). This visualisation demonstrated that regular bouts of constant amplitude and frequency were indicative of swimming and resting in all three sensors. Rest behaviour, defined as motionless behaviour, corresponds to acceleration and magnetometer values of near- zero with small fluctuations in the waveform signal. Higher waveform signal frequency and amplitude were typically representative of burst and headshake behaviours, with headshake behaviours typically eliciting higher frequencies amplitudes for longer durations than burst. This visual inspection revealed obvious differences in waveform signal characteristics for each behaviour depending on the sensor (Figure 3.6). As an aside, upon exploration of the data, regular oscillations in steady swimming behaviour were less evident for this data set than for the previous chapter, with noise visible in the waveform signal (Figure 3.7).

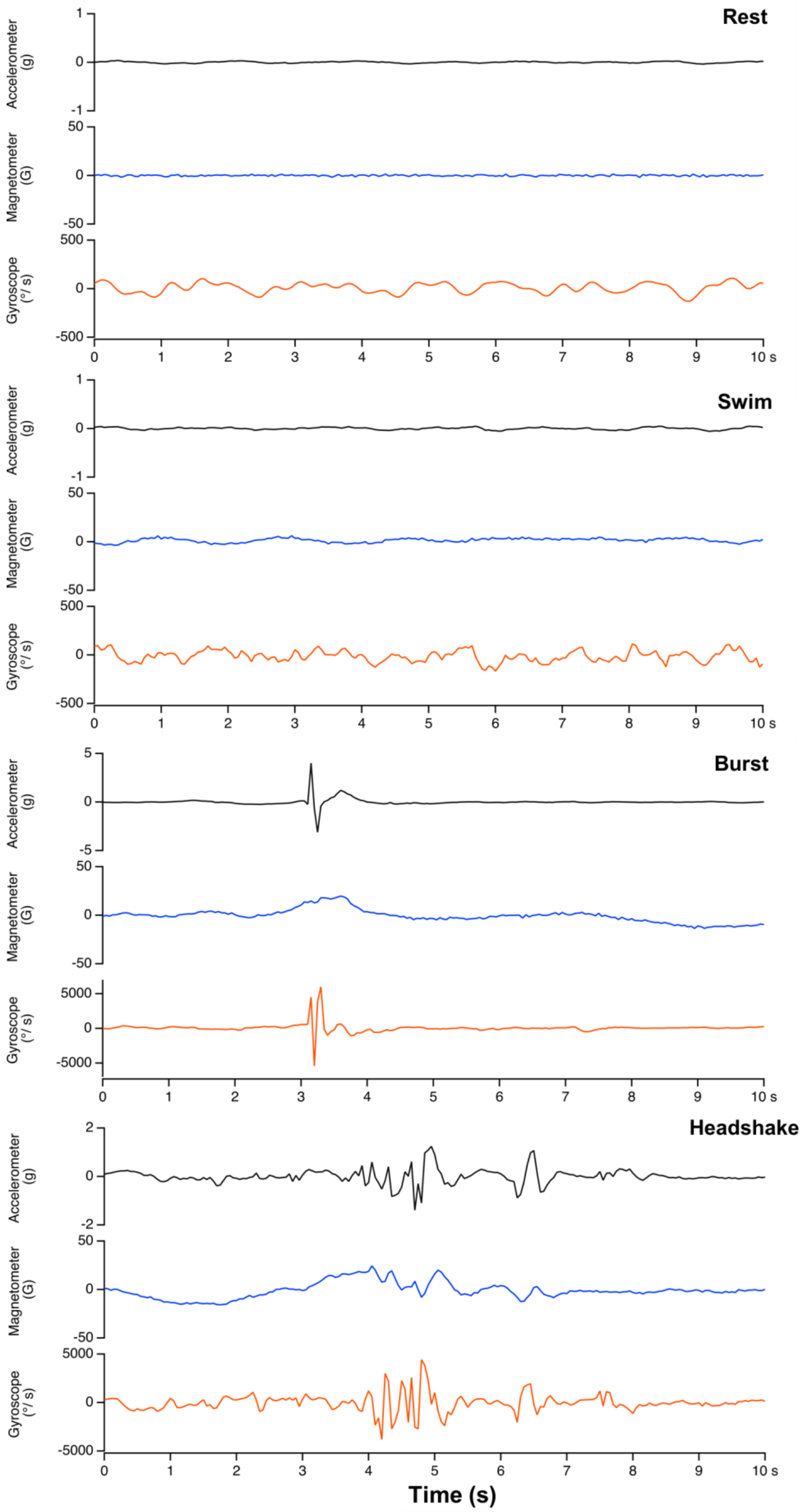

Figure 3.6 Time series plots showing the representative change in waveform signal amplitude and frequency in the dominant lateral axis of each sensor for four observed behaviours (swim, rest, burst and headshake) in a sicklefin lemon shark (*N. acutidens*). Note the different scales to facilitate visualisation of behaviours and different waveform signal features.

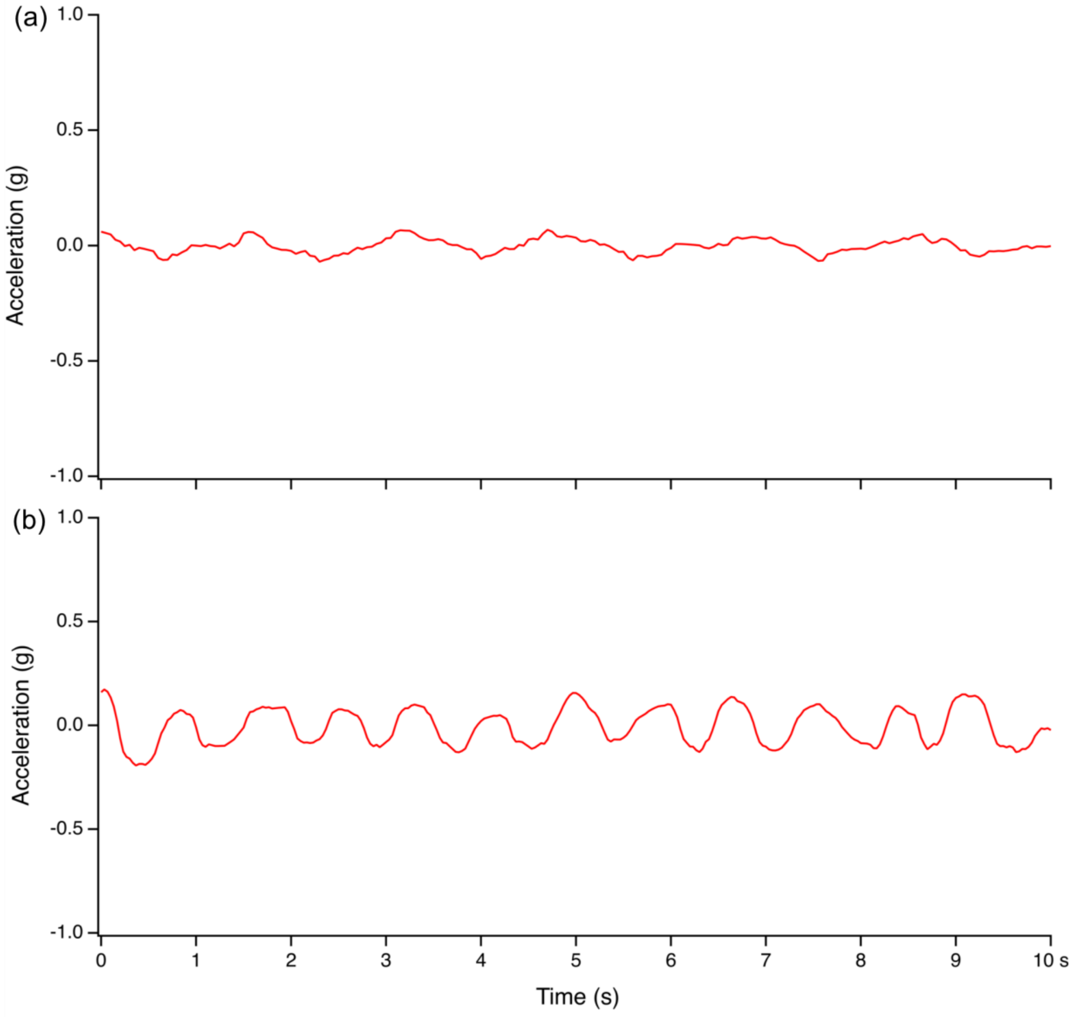

Figure 3.7 Time series plot showing representative waveform signals for dynamic sway acceleration for steady swimming in (a) a subadult sicklefin lemon shark (*N.acutidens*) and (b) a juvenile lemon shark (*N.brevirostris*) from the data collected by Brewster et al., (2018) used for analysis in Chapter 2. The oscillations in (a) are distorted and noise is present in the acceleration waveform signal, whilst in comparison the waveform signal for (b) displays clearly discernible regular oscillations with no apparent noise.

3.3.2 Overall model classification performance

Overall model accuracy for all movement sensor combinations was high (>96 %), with minimal overall difference in accuracy between sensors (Figure 3.8). Results indicate that combining all three tri-axial movement sensors is best for overall classification of behaviours observed in sicklefin lemon sharks. The range of F_M scores for all combinations of sensors was small (0.535 - 0.597; Table 3.4 and Figure 3.9c). The magnetometer performed with the lowest classification score according to this metric, whilst the tri-sensor combination achieved the highest score. The gyroscope was the best performing individual movement sensor (FM =0.579; Table 3.4 and Figure 3.9c). Dual-combinations of accelerometer-magnetometer and accelerometer-gyroscope performed equally well overall, and only marginally less than the highest performing tri-sensor combination (F_M =0.590; Table 3.4 and Figure 3.9c). Overall precision was lowest in the accelerometer and highest for both a combination of accelerometer and magnetometer and also the tri-sensor combination, however the range was small (Figure 3.9a). Overall recall was lowest for the magnetometer and highest for accelerometer and magnetometer combined (Figure 3.9b).

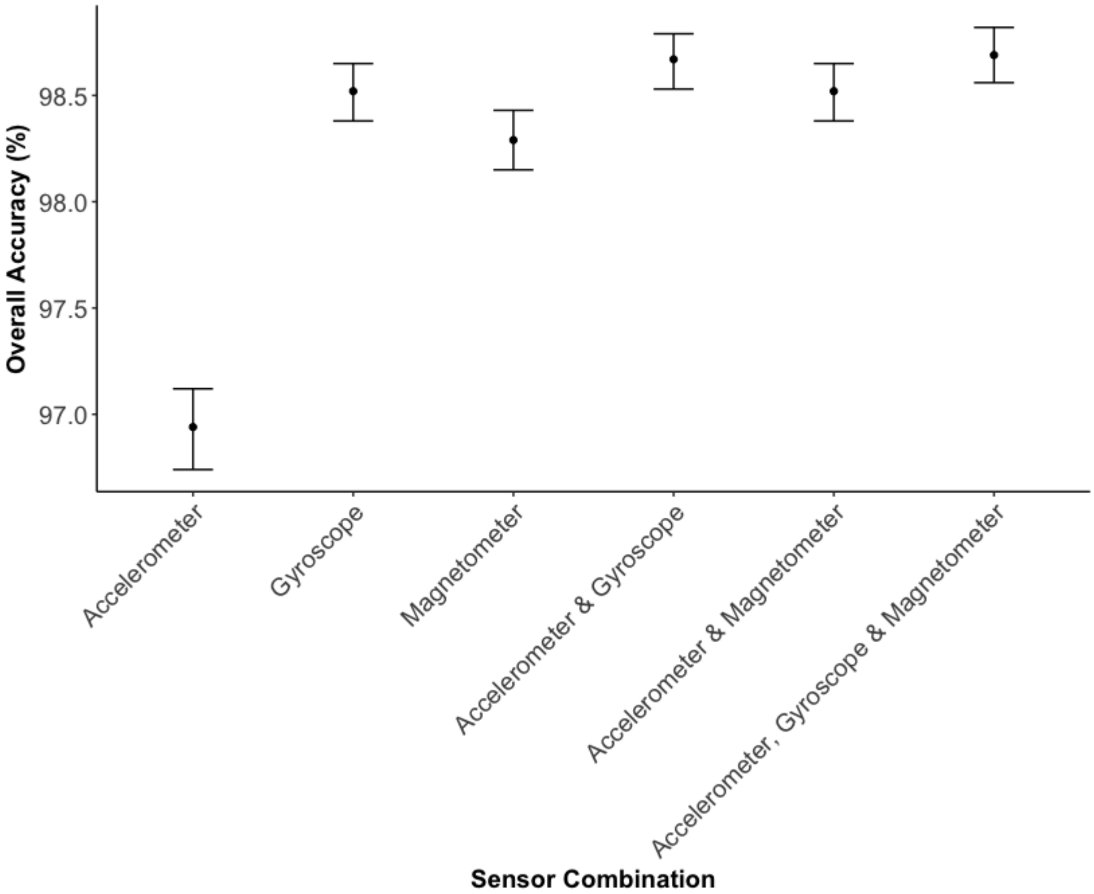

Figure 3.8 Overall RF classifier accuracy including 95% confidence intervals. Overall accuracy was >96.5 % for all individual movement sensors and each combination of sensors. Overall accuracy was lowest for the accelerometer and highest for the tri-sensor combination of accelerometer, gyroscope and magnetometer. The difference in overall accuracy between sensor combinations was small (<2 %).

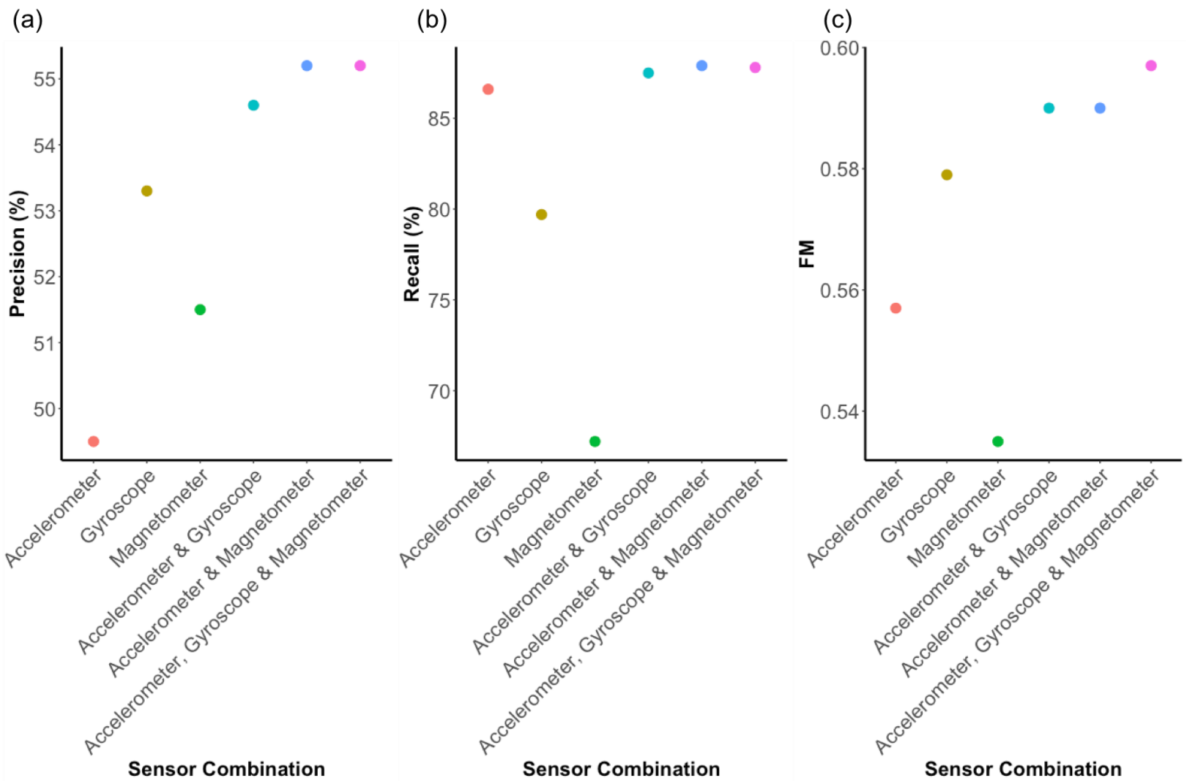

Figure 3.9 Overall RF model classification performance (a) Precision was lowest for the accelerometer and highest for a both a combination accelerometer and magnetometer and a combination of all three sensors. (b) Recall was lowest for the magnetometer and highest for b a combination of all three sensors. (c) F_M was lowest for magnetometer and highest for a combination of all three sensors. Individual movement sensors performed lower in all evaluation metrics than all combinations of movement sensors.

3.3.3 Individual behaviour class performance

Classifier performance for individual behavioural classes varied according to combination of movement sensors tested and evaluation metric and (Figure 3.10; Table 3.4). The tri-sensor combination was the best combination of sensors for classification of all behavioural classes except for burst, where the accelerometer performed best (F_1 =0.287, Table 3.4).

Swim was classified with highest scores across all individual movement sensors and combinations of sensors (F_1 >0.983; Table 3.4). Both precision and recall were >0.971 (Table 3.4). There was very little difference between lowest (F_1 =0.983) and highest (F_1 =0.993) classification scores of swim events, however the tri-sensor combination was the best for this class (Table 3.4 and Figure 3.10c).

Rest was also classified well by all individual sensors and combinations of sensors ($F_1 \geq 0.816$; Table 3.4). Confusion between swim and rest caused lower precision (0.726) and therefore F_1 score (0.816) for the accelerometer than the other sensors (Figure 3.10 and Table 3.4). The tri-sensor achieved best classification of rest events (Table 3.4).

Burst events were not classified as well as events from swim or rest classes ($F_1 \leq 0.287$). The magnetometer was the worst sensor for classifying burst events due to scoring lowest precision (≤ 0.250) and lowest recall (0.071) (Table 3.4 and Figure 3.10). Misclassification of burst events was due to confusion between burst and swim (Table 3.4). The accelerometer was the best sensor for classification of burst events (Table 3.4).

Headshakes achieved the lowest class performance of all behaviour classes across all sensor combinations ($F_1 \leq 0.160$; Figure 3.10c and Table 3.4). Recall was high (>0.750) for all combinations of sensors except the gyroscope (0.500) and magnetometer (0.071). Only one headshake event was correctly labelled by the magnetometer (Table 3.4). Precision however was low in all sensors and combinations of sensors (≤ 0.088; Figure 3.10a and 3.10b). Poor precision was a result of incorrect labels from all classes, however the most common false positive for headshake was from the swim class. Headshakes were best classified by the tri-sensor combination (Table 3.4).

Table 3.4 Confusion matrix and evaluation metrics of RF models generated from the test set of ground- truthed data for each sensor combination. Rows indicate actual observations and columns represent predicted behaviours. Highlighted values show observations which were correctly classified by the model (TP; True Positives).

	SENSOR COMBINATION		PREDICTED BEHAVIOUR				PERFORMANCE METRICS			
		Class	Burst	Headshake	Rest	Swim	Precision	Recall	F1 Score	Macro F1
ACTUAL BEHAVIOUR	**Accelerometer**	Burst	26	2	0	9	0.181	0.703	0.287	0.557
		Headshake	0	12	0	2	0.077	0.857	0.142	
		Rest	1	4	1570	109	0.726	0.932	0.816	
		Swim	117	137	592	29183	0.995	0.971	0.983	
	Magnetometer	Burst	25	0	0	12	0.081	0.676	0.144	0.535
		Headshake	4	1	0	9	0.067	0.071	0.069	
		Rest	1	0	1608	75	0.916	0.955	0.935	
		Swim	280	14	147	29588	0.997	0.985	0.991	
	Gyroscope	Burst	27	1	0	9	0.152	0.730	0.252	0.579
		Headshake	2	7	0	5	0.078	0.500	0.135	
		Rest	3	0	1637	44	0.902	0.972	0.936	
		Swim	145	82	178	29624	0.998	0.986	0.992	
	Accelerometer & Gyroscope	Burst	26	2	0	9	0.176	0.703	0.281	0.590
		Headshake	0	12	0	2	0.081	0.857	0.147	
		Rest	1	4	1602	77	0.928	0.951	0.940	
		Swim	121	131	124	29653	0.997	0.988	0.992	
	Accelerometer & Magnetometer	Burst	26	2	0	9	0.173	0.703	0.278	0.590
		Headshake	0	12	0	2	0.083	0.857	0.151	
		Rest	1	3	1604	76	0.926	0.952	0.939	
		Swim	123	128	127	29651	0.997	0.987	0.992	
	Accelerometer, Magnetometer & Gyroscope	Burst	26	2	0	9	0.177	0.703	0.282	0.597
		Headshake	1	12	0	2	0.088	0.857	0.160	
		Rest	2	2	1619	61	0.943	0.961	0.952	
		Swim	119	120	98	29692	0.998	0.989	0.993	

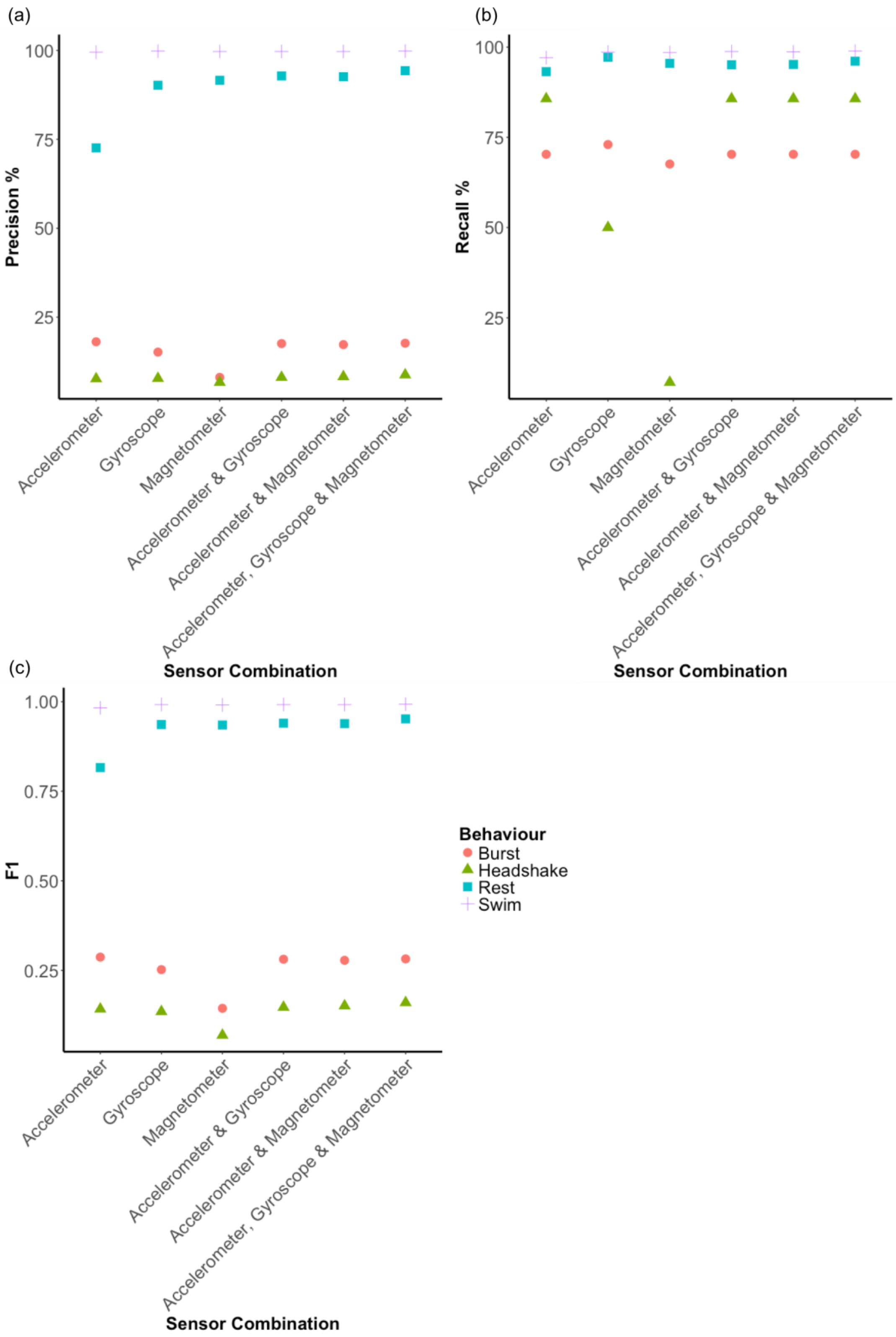

Figure 3.10 RF model class performance for classification of behaviour of sicklefin lemon sharks (*N.acutidens*) (*n* =4) for different movement sensors and combinations of sensors. Performance is described by (a) Precision, (b) Recall and (c) F_1 Score for each observed behaviour from captive trials (burst, headshake, rest, swim). Performance varies by evaluation metric and movement sensor for each behavioural class. Swim and rest achieve higher scores in all evaluation metrics than burst and headshake.

3.4 Discussion

This is the first study to my knowledge to evaluate which movement sensors are best for automatic classification of shark behaviour. This study provides evidence that overall classification of sicklefin lemon shark behaviours can be improved by a tri-sensor combination incorporating a tri-axial accelerometer, magnetometer and gyroscope. However, overall classifier performance for the tri-sensor combination was only marginally improved over both dual-sensor combinations, where overall behaviour was classified equally well by combining an accelerometer with either a magnetometer or gyroscope.

3.4.1 Addition of a magnetometer

A dual-combination of accelerometer and magnetometer might be the preferred choice for a number of reasons. Firstly, this combination improves classification of behaviour that does not involve significant dynamic body movement (resting). Whilst the model classified swim and rest behaviours well with all sensors and combinations of sensors (F_1> 0.816, Figure 10c), it is interesting to note that the accelerometer alone classified these behaviours with lowest performance, especially for resting behaviour. This result is somewhat surprising given that accelerometers are considered an appropriate sensor for the elucidation of both active (swimming) and inactive (resting) behaviour in various species of shark (Whitney *et al.* 2007, Gleiss *et al.* 2009a, Whitney *et al.* 2010, Wilson *et al.* 2015, Brewster *et al.* 2018). Williams *et al.* (2015) used unsupervised learning techniques to successfully classify active and passive flying from accelerometer data from large birds of prey, where passive flight was characterised by the absence of dynamic body movement, or flapping (Halsey *et al.* 2009). This behaviour class was not classified well from data recorded with the accelerometer, attributed to the lack of dynamic body movement (Williams *et al.* 2015). It was suggested that rotational measurements from the magnetometer might enable improved classification of passive flight due to its independence from dynamic body movement. Resting in sicklefin lemon sharks was characterised in this study by a lack of dynamic body movement, or absence of tail- beats (see Figure 3.6). Resting behaviour was indeed better classified with the magnetometer than the accelerometer. Body orientation calculated from fin-mounted accelerometer devices can be inexact, due to minor fin movements resulting from both wave surge and the flexibility of the fin to which the tag is attached. This was observed in resting nurse sharks (Whitney *et al.* 2010). Although these sharks were adults of a slightly larger size (TL 248- 263 cm) and the tag was mounted on the second dorsal fin, this supports the results presented here, where classification was improved when rotational measurements from the magnetometer are included in the model. All behaviours, regardless of dynamic body movement, were better classified with a dual combination of the two sensors.

In addition to the magnetometers independence from dynamic body movement, the improved overall classification using this dual-sensor combination may have been a result of the magnetometers insensitivity to noise. López *et al.* (2015) captured swimming movements in cetaceans with magnetometers, before later postulating more specifically that the tail- beats from caudal propulsion in other fish could be well distinguished using magnetometers (López *et al.* 2016). Accelerometer and magnetometer data has been collected in tandem from whale sharks (6 Hz), with the magnetometer enhancing the accelerometers ability to capture regular tail- beats in steady swimming (Williams *et al.* 2017). Although much smaller in size (<2 m) than whale sharks (>4 m), tail- beats arising from caudal propulsion in the sub-adult sicklefin lemon sharks were captured well in both sensors. The sharks in this study were often observed to be ‘finning’ or swimming with their dorsal fin (and therefore dorsally mounted tag) almost fully exposed at the surface of the shallow water during captive trials. This was a naturally occurring behaviour personally witnessed in wild, free swimming sharks at the study site, therefore not simply a result of captivity. Williams *et al.* (2017) noted that unwanted noise, owing to movement of water past the tag and the fin, caused distortion in the acceleration waveform signal from the dominant sway axis representing whale shark tail- beats. This same distortion was visible from the acceleration data for sicklefin lemon sharks (Figure 3.7), which is probably a result of this ‘finning’ in the shallow water at the study site (< 1 m). Although Dutta *et al.* (2015) attributed a high level of noise to the reduced performance of an ensemble classifier for cattle behaviour from accelerometer and magnetometer data, Williams *et al.* (2017) proposed that a combination of the two was best for capturing swimming behaviours in large sharks. The magnetometer captured body movements from whale sharks in all three axes well, resolving the issue with noise in the acceleration data. This provides further evidence for choosing this dual sensor combination for the classification of behaviour in sicklefin lemon sharks.

Another reason for choosing the accelerometer- magnetometer combination of sensors would be the opportunity to combine behavioural classification with dead reckoning. Dead reckoning reconstructs the three-dimensional movement path of an animal, based on combining sequential measurements of compass heading, speed and change in depth or height (Mitani *et al.* 2003, Wilson *et al.* 2008, Bidder *et al.* 2014, Bidder *et al.* 2015). Over-laying broad-scale spatial movement data with fine-scale behaviours gives us potentially valuable insights into where sharks perform certain behaviours within their habitat.

3.4.2 Addition of a gyroscope

The successful identification and classification of resting and swimming behaviours has been achieved using accelerometers with a range of sampling frequencies as low as 5 Hz (Chapter 2) (Whitney *et al.* 2007, Gleiss *et al.* 2009a, Whitney *et al.* 2010, Wilson *et al.* 2015, Brewster *et al.* 2018). These behaviours were classified well by the model developed here in all instances, regardless of sensor combination. In contrast however, overall classification performance for headshake and burst behaviours was poor, regardless of the sensor combination was used. Classifier performance again, as per Chapter 2, is inherently dependant on behaviour complexity (McClune *et al.* 2014).

The kinematics of fast start (burst) behaviours in fish involve highly agile and rapid turns at the beginning of forward propulsion, associated with attacking prey or predator avoidance behaviours (Domenici and Blake 1991, Domenici and Blake 1997, Broell *et al.* 2013). Burst and headshake behaviours in sicklefin lemon sharks were characterised by higher waveform signal frequency and amplitude in each movement sensor (Figure 3.6), indicating higher acceleration and similar changes in body orientation. Used alone, the gyroscope was the bet individual sensor for overall classification of behaviour in sicklefin lemon sharks. Furthermore, addition of predictor variables extracted from the raw gyroscope data to those from other movement sensors aided classification of all behaviours, therefore this sensor may be particularly beneficial for classification of complex burst and headshake behaviours.

Conventional smoothing and filtering methods used to separate gravity-based acceleration from dynamic body movements (Sato *et al.* 2003, Watanabe *et al.* 2005, Wilson *et al.* 2006, Shepard *et al.* 2008a) can be prone to error (Fourati *et al.* 2011). Gyroscopes can substantially reduce this error; especially for the quantification of high-resolution kinematics, as is required for investigations of bio-mechanics (Noda *et al.* 2012, Noda *et al.* 2013, 2014). However, recording duration in these studies was not necessarily important; fine scale behaviours were successfully described with devices deployed for limited recording durations of less than 2 hours, at scale appropriate high sampling rates (100- 500 Hz). The gyro method has not only been successful for accurately quantifying the kinematics of ecologically important predation and escape behaviours in fish but has even enabled fine scale identification of prey type captured (Kawabata *et al.* 2014). In this example, the dual-sensor accelerometer-gyroscope device was pre-programmed to record only during periods corresponding to increased crepuscular feeding activity in white streaked grouper (*Epiniphelus ongus*), as the aim was to specifically identify prey type during feeding behaviours. The high sampling frequencies described for gyroscopes result in substantially larger memory and battery requirements and decreased recording durations (Noda *et al.* 2012, Noda *et al.* 2014). However, when the aim of a study is to simply quantify or characterise complex behaviours, rather than create time-activity budgets from classification, recording duration must only be sufficient enough to capture the specific behaviours of interest.

A more recent study focusing on simple classification of behaviours rather than quantification of kinematics for biomechanics, successfully classified standing, lying and walking behaviours in sheep from an accelerometer-gyroscope combination (Walton *et al.* 2018). According to Walker *et al.* (2015), successful classification is often a result of the predictor variables extracted from the raw data for input to the ML algorithm. For classification of sheep behaviours, the predictor variables were extracted from differences in the magnitude of the raw ground- truthed data from each sensor, without having used the common gyro-method, however it was not assessed whether behaviours could have been classified equally well with either sensor alone. In contract to the previous recommendations, the current study provides no evidence for a substantial increase in overall classification performance by incorporating other movement sensors over an accelerometer alone, which was the best sensor for burst behaviour. Given that the focus of this study was to simply establish best practice classification of shark behaviour, addition of a gyroscope for improved classification (as opposed the quantification of specific kinematics of behaviour) may not be worth the cost of the additional data and decreased deployment duration. In fact, given the minimal range between overall classification scores for the worst sensor (accelerometer, F_M =0.535) and the best (tri-sensor combination, F_M =0.597), a device programmed to record acceleration only, without utilising either of the other movement sensors would ensure absolute maximisation of memory and battery capacity without substantially decreasing classifier performance.

3.4.2 Misclassification of individual behaviour classes

Besides assessing movement sensors for behaviour classification, there are a number of reasons for misclassification of behaviours involving more complex kinematics (burst and headshake). Outlined in the first chapter of this thesis, misclassification of these behaviours was again likely a result of a combination of small class size (Brown *et al.* 2013, Brewster *et al.* 2018), similarities in waveform signal characteristics between behaviours (McClune *et al.* 2014, Walton *et al.* 2018) and variability within behaviours (Walker *et al.* 2015). The variability and inconsistency of some behaviours makes automatic labelling of behaviour a difficult task (Walker *et al.* 2015) and misclassification in some cases- even with multiple sensors- has been attributed to high variability (Noda *et al.* 2013). Swim was classified well by the model with all sensors and combinations of sensors, however high variability for this behaviour was demonstrated by Gleiss *et al.* (2009a) where a range of 0.4-1.2 Hz TBF and 0.002-0.16 *g* amplitude was found for steady swimming in lemon sharks from accelerometer data. Although these values for the kinematics of each behaviour were not directly assessed in this study, variation is presumably even larger for behaviours such as burst or headshaking which involve far more complex kinematics. Again, this is where class size affects classifier performance similarly to Chapter 2. The majority classes were selectively down sampled from a stratified subset to be closer in frequency to the rare class sizes, yet even then the model was trained incorporating a large degree of variability for the majority classes. There were few observations for both burst and headshaking behaviour involving far more complex kinematics and variable speeds and postures, so any extreme variability was not taken into account when training and testing the model. This explains why the model wrongly labelled some events from the swim as both headshake and burst behaviours, resulting in low precision for these classes. As per the previous chapter, classifier performance could have been improved by using a VE classifier such as that developed by Brewster *et al.* (2018). Despite the limitations noted previously, the main purpose of analysis for this study was achieved in that the results demonstrate the effect that the choice of movement sensor has on classification of sicklefin lemon shark behaviour.

Where the results from each chapter of this thesis are compared, it is evident that headshakes were classified better ($F_1 \approx 0.500$) by an accelerometer (30 Hz) in juvenile lemon sharks (Chapter 2) than by any movement sensor or combination of sensors (20 Hz) in larger sub-adult lemon sharks ($F_1 \leq 0.160$; Chapter 3). Larger sharks may have had the ability to handle and manipulate prey with less effort (both side to side head shakes and tail- beats) than smaller juvenile sharks (personal observations). The frequency of limb movements is scale related however (Whitney *et al.* 2012), warranting further investigation if classification of this particular behaviour is of interest.

3.5 Conclusion

The importance of studying animal behaviour over biologically meaningful time scales has been highlighted in the previous chapters. To summarise this chapter, the benefits of collecting high resolution movement data recorded from additional sensors must outweigh the increased memory and battery consumption of the device itself. Given current technological constraints, the multi sensor device programmed for three sensors to record in 9 channels (at sampling frequencies of 20 Hz) will record for up to 48 hours. This chapter has demonstrated that although a tri-sensor combination improves classification, when seeking to maximise recording durations an accelerometer is sufficient for simple classification of sicklefin lemon shark behaviour. Whilst the addition of other movement sensors has little effect on the classification of behaviour, the incorporation of a magnetometer, gyroscope or both of these sensors is beneficial when maximum recording duration is not critical. This is especially relevant if behaviour classification is intended to be combined with dead reckoning or biomechanics. In conclusion, the results presented in this chapter are significant in terms of allowing future studies to consider three tri-axial movement sensors (accelerometer, magnetometer and gyroscope) for improved overall classification of shark behaviour.

Chapter 4. General Discussion

4.1 Model application

The relevant limitations have been discussed in Chapter 2 and Chapter 3, however there are some broader concepts that must be acknowledged. Although the model developed for this study was designed with the purpose of analysis being to assess the impacts of sampling frequency and movement sensors on classification, it is important to note that if the model is to be applied to data collected from wild, free-swimming sharks, there are a number of caveats. Firstly, the data collected from both captive trials was limited by time, budget, equipment and human resources, as pointed out by (Cooke 2008), this is where care must be taken in bio-logging studies as such limitations narrow the applicability of results. In addition, the RF classifier developed here has several limitaitons, including that it is only applicable to sharks of a similar size and that the behaviours observed in captivity may not always translate directly into the same behaviours performed in the wild (Leos-Barajas *et al.* 2017, Brewster *et al.* 2018). This could be due to both the size of the captive study area and the presence of a human observer (Ladds *et al.* 2017). Whilst every care was taken to ensure that the captive trials took place in such a way that normal ambient environmental conditions were experienced by the captive sharks within their natural habitat, the pen size was far smaller than the habitat range of both juvenile lemon sharks and sub adult sicklefin lemon sharks. In addition, observations for ethograms were conducted as discretely and quietly as possible, however human activity around the pen may have had some effect on the incidence or repertoire of behaviours performed, although this should not have affected the kinematics of the behaviours performed. Moreover, the ethogram was developed only from those behaviours which were observed from juvenile sharks in captivity. Wild sharks may spend their time performing other behaviours that were not captured in captivity, including those related to social dynamics (Wilson *et al.* 2015), mating in sexually mature sharks (Whitney *et al.* 2010) and regurgitation (personal observations in the field). If these behaviours were to be recorded from a free-swimming deployment and presented to a model for which is has not been trained they will be misclassified. In spite of this, the VE classifier developed by Brewster *et al.* (2018) was tested against observations from wild juvenile lemon sharks and appeared to translate well from captive observations of behaviour.

4.2 Scope of behaviour classification

Behaviour classification is a hypothesis free approach to the study of animal behaviour (Valletta *et al.* 2017). That said, animal behaviour must be studied in both its natural environmental and physiological context, broadening the potential scope of animal behavioural studies (Whitney *et al.* 2012). The integration of multiple parameters in addition to body movement with behavioural classification means that it is possible to gain unique insights into what animals do, where and why (Whitney *et al.* 2012). This can be achieved by incorporating location (GPS, depth), environmental parameters (humidity, light intensity, temperature) and physiological data (heart rate, body temperature). Although these additional channels of data take up extra memory and battery of device, they are not required to be recorded in high resolution (e.g. 1 Hz). This “daily diary” of animal behaviour, developed by (Wilson *et al.* 2008), helps reveal temporal- spatial distributions and timing of specific behaviours, or time-activity budget, essential for addressing conservation issues (Shepard *et al.* 2008b, Wilson *et al.* 2008, Brown *et al.* 2013, Walker *et al.* 2015). Both Cooke (2008) and Shepard *et al.* (2008b) highlight that temporal-spatial knowledge of the occurrence of behaviours helps reveal critical species-specific information regarding habitat use (e.g. foraging, courtship, refuge and nursery areas, migration and social interactions) and even the potential sources and transmission of disease. This ecological information may contribute towards the management of conservation issues including the consequences environmental change, effects of anthropogenic disturbance and protected area planning.

Furthermore, the physiological context of behaviours can be examined by incorporating the study of energetics (Cooke 2008, Wilson *et al.* 2008). All behaviours have an energetic cost (Gleiss *et al.* 2011b). Respirometry, or the measurement of oxygen consumption enables us to correlate an animals' activity with its metabolic rate (Wilson *et al.* 2006, Halsey *et al.* 2009, Lear *et al.* 2017). Predictions of metabolic rate have been made for active and resting lemon sharks from accelerometer data, and it was suggested that the energetic cost of behaviours such as prey capture and mating could be estimated by pairing ODBA with behaviours visualised in the accelerometer data (Lear *et al.* 2017). More recently, in-situ energy expenditure was linked to time- activity budgets in lemon sharks (Bouyoucos *et al.* 2018), however this was not applied to data from free-swimming sharks. Another technique related to energetics is stable isotope analysis, a method that enables us to understand the diet, trophic position and role of an animal within an ecosystem (Hussey *et al.* 2012). If we have knowledge of an animal's diet, we can estimate the metabolic gain from classification of successful prey capture. The VE model developed by Brewster *et al.* (2018) enabled successful identification of when and where successful prey captures occurred. By combining classification of behaviour with the study of energetics, we can calculate and map not only a time-activity budget but also a time-energy budget (Lear *et al.* 2017). This physiological information regarding energetics is considered essential for informing conservation measures (Cooke 2008).

The results from this thesis show that behaviours in sharks can be classified with a RF classifier, and when both sampling frequency and choice of movement sensor are carefully considered, classification methods can be optimised to enable future long-term studies of behaviour in the natural environment. It is important to consider an animal's behaviour from a holistic perspective, incorporating both the environmental and physiological context. This thesis provides a valuable resource enabling future research to refine the programming of bio-logging tools, hence facilitating best practice for the classification of shark behaviour.

References

Abrahms, B.; Jordan, N.; Golabek, K.; McNutt, J.; Wilson, A., and Brashares, J. 2016. "Lessons from integrating behaviour and resource selection: activity-specific responses of African wild dogs to roads." *Animal Conservation* 19 (3):247-255.

Bidder, O. R.; Campbell, H. A.; Gomez-Laich, A.; Urge, P.; Walker, J.; Cai, Y.; Gao, L.; Quintana, F., and Wilson, R. P. 2014. "Love thy neighbour: automatic animal behavioural classification of acceleration data using the K-nearest neighbour algorithm." *PLoS One* 9 (2):e88609.

Bidder, O. R.; Walker, J. S.; Jones, M. W.; Holton, M. D.; Urge, P.; Scantlebury, D. M.; Marks, N. J.; Magowan, E. A.; Maguire, I. E., and Wilson, R. P. 2015. "Step by step: reconstruction of terrestrial animal movement paths by dead-reckoning." *Movement Ecology* 3 (1):23.

Bouyoucos, I. A.; Suski, C. D.; Mandelman, J. W., and Brooks, E. J. 2018. "In situ swimming behaviors and oxygen consumption rates of juvenile lemon sharks (Negaprion brevirostris)." *Environmental Biology of Fishes*:1-13.

Breiman, L. 1999. "Random Forests." *UC Berkeley TR567*.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45 (1):5-32.

Breiman, L. 2002. "Manual on setting up, using, and understanding random forests v3. 1." *Statistics Department University of California Berkeley, CA, USA* 1.

Brewster, L. R.; Dale, J. J.; Guttridge, T. L.; Gruber, S. H.; Hansell, A. C.; Elliott, M.; Cowx, I. G.; Whitney, N. M., and Gleiss, A. C. 2018. "Development and application of a machine learning algorithm for classification of elasmobranch behaviour from accelerometry data." *Marine Biology* 165 (4):62.

Broell, F.; Noda, T.; Wright, S.; Domenici, P.; Steffensen, J. F.; Auclair, J.-P., and Taggart, C. T. 2013. "Accelerometer tags: detecting and identifying activities in fish and the effect of sampling frequency." *Journal of Experimental Biology* 216 (7):1255-1264.

Brown, D. D.; Kays, R.; Wikelski, M.; Wilson, R. P., and Klimley, A. P. 2013. "Observing the unwatchable through acceleration logging of animal behavior." *Animal Biotelemetry* 1 (1):20.

Brownscombe, J. W.; Gutowsky, L. F.; Danylchuk, A. J., and Cooke, S. J. 2014. "Foraging behaviour and activity of a marine benthivorous fish estimated using tri-axial accelerometer biologgers." *Marine Ecology Progress Series* 505:241-251.

Campbell, H. A.; Gao, L. L.; Bidder, O. R.; Hunter, J., and Franklin, C. E. 2013. "Creating a behavioural classification module for acceleration data: using a captive surrogate for difficult to observe species." *Journal of Experimental Biology* 216 (24):4501-4506.

Chapman, D. D.; Babcock, E. A.; Gruber, S. H.; Dibattista, J. D.; Franks, B. R.; Kessel, S. A.; Guttridge, T.; Pikitch, E. K., and Feldheim, K. A. 2009. "Long-term natal site-fidelity by immature lemon sharks (Negaprion brevirostris) at a subtropical island." *Molecular Ecology* 18 (16):3500-3507.

Chawla, N. V.; Bowyer, K. W.; Hall, L. O., and Kegelmeyer, W. P. 2002. "SMOTE: synthetic minority over-sampling technique." *Journal of Artificial Intelligence Research* 16:321-357.

Chawla, N. V.; Lazarevic, A.; Hall, L. O., and Bowyer, K. W. 2003. "SMOTEBoost: Improving prediction of the minority class in boosting." European Conference on Principles of Data Mining and Knowledge Discovery.

Chen, C.; Liaw, A., and Breiman, L. 2004. "Using random forest to learn imbalanced data." *University of California, Berkeley* 110:1-12.

Chen, K. Y., and Bassett Jr, D. R. 2005. "The technology of accelerometry-based activity monitors: Current and future." *Medicine and Science in Sports and Exercise* 37 (11 SUPPL.):S490-S500.

Cooke, S. J. 2008. "Biotelemetry and biologging in endangered species research and animal conservation: relevance to regional, national, and IUCN Red List threat assessments." *Endangered Species Research* 4 (1-2):165-185.

Cooke, S. J.; Hinch, S. G.; Wikelski, M.; Andrews, R. D.; Kuchel, L. J.; Wolcott, T. G., and Butler, P. J. 2004. "Biotelemetry: a mechanistic approach to ecology." *Trends in Ecology and Evolution* 19 (6):334-43.

Cortés, E. 1999. "Standardized diet compositions and trophic levels of sharks." *ICES Journal of Marine Science* 56 (5):707-717.

Cutler, D. R.; Edwards, T. C.; Beard, K. H.; Cutler, A.; Hess, K. T.; Gibson, J., and Lawler, J. J. 2007. "Random forests for classification in ecology." *Ecology* 88 (11):2783-2792.

Davis, R. W.; Fuiman, L. A.; Williams, T. M.; Horning, M., and Hagey, W. 2003. "Classification of Weddell seal dives based on 3-dimensional movements and video-recorded observations." *Marine Ecology Progress Series* 264:109-122.

Domenici, P., and Blake, R. 1997. "The kinematics and performance of fish fast-start swimming." *Journal of Experimental Biology* 200 (8):1165-1178.

Domenici, P., and Blake, R. W. 1991. "The kinematics and performance of the escape response in the angelfish (Pterophyllum eimekei)." *Journal of Experimental Biology* 156 (1):187-205.

Dutta, R.; Smith, D.; Rawnsley, R.; Bishop-Hurley, G.; Hills, J.; Timms, G., and Henry, D. 2015. "Dynamic cattle behavioural classification using supervised ensemble classifiers." *Computers and Electronics in Agriculture* 111:18-28.

Filmalter, J.; Dagorn, L., and Cowley, P. 2013. "Spatial behaviour and site fidelity of the sicklefin lemon shark Negaprion acutidens in a remote Indian Ocean atoll." *Marine Biology* 160 (9):2425-2436.

Fourati, H.; Manamanni, N.; Afilal, L., and Handrich, Y. 2011. "Posture and body acceleration tracking by inertial and magnetic sensing: Application in behavioral analysis of free-ranging animals." *Biomedical Signal Processing and Control* 6 (1):94-104.

Ganganwar, V. 2012. "An overview of classification algorithms for imbalanced data sets[J]." *International Journal of Emerging Technology and Advanced Engineering* 2 (4):42-47.

Gao, L.; Campbell, H. A.; Bidder, O. R., and Hunter, J. 2013. "A Web-based semantic tagging and activity recognition system for species' accelerometry data." *Ecological Informatics* 13:47-56.

Gleiss, A. C.; Dale, J. J.; Holland, K. N., and Wilson, R. P. 2010. "Accelerating estimates of activity-specific metabolic rate in fishes: Testing the applicability of acceleration data-loggers." *Journal of Experimental Marine Biology and Ecology* 385 (1-2):85-91.

Gleiss, A. C.; Gruber, S. H., and Wilson, R. P. 2009a. "Multi-channel data-logging: towards determination of behaviour and metabolic rate in free-swimming sharks." In *Tagging and tracking of marine animals with electronic devices*, 211-228. Springer.

Gleiss, A. C.; Norman, B.; Liebsch, N.; Francis, C., and Wilson, R. P. 2009b. "A new prospect for tagging large free-swimming sharks with motion-sensitive data-loggers." *Fisheries Research* 97 (1-2):11-16.

Gleiss, A. C.; Norman, B., and Wilson, R. P. 2011a. "Moved by that sinking feeling: variable diving geometry underlies movement strategies in whale sharks." *Functional Ecology* 25 (3):595-607.

Gleiss, A. C.; Wilson, R. P., and Shepard, E. L. C. 2011b. "Making overall dynamic body acceleration work: on the theory of acceleration as a proxy for energy expenditure." *Methods in Ecology and Evolution* 2 (1):23-33.

Gleiss, A. C.; Wright, S.; Liebsch, N.; Wilson, R. P., and Norman, B. 2013. "Contrasting diel patterns in vertical movement and locomotor activity of whale sharks at Ningaloo Reef." *Marine Biology* 160 (11):2981-2992.

Graf, P. M.; Wilson, R. P.; Qasem, L.; Hackländer, K., and 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 (8):e0136751.

Gruber, S. H.; Marignac, J. R., and Hoenig, J. M. 2001. "Survival of juvenile lemon sharks at Bimini, Bahamas, estimated by mark–depletion experiments." *Transactions of the American Fisheries Society* 130 (3):376-384.

Guttridge, T.; Gruber, S.; Gledhill, K.; Croft, D.; Sims, D., and Krause, J. 2009. "Social preferences of juvenile lemon sharks, Negaprion brevirostris." *Animal Behaviour* 78 (2):543-548.

Halsey, L.; Shepard, E.; Quintana, F.; Laich, A. G.; Green, J., and Wilson, R. 2009. "The relationship between oxygen consumption and body acceleration in a range of species." *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* 152 (2):197-202.

Hammond, T. T.; Springthorpe, D.; Walsh, R. E., and Berg-Kirkpatrick, T. 2016. "Using accelerometers to remotely and automatically characterize behavior in small animals." *Journal of Experimental Biology* 219 (11):1618-1624.

Hays, G. C.; Ferreira, L. C.; Sequeira, A. M.; Meekan, M. G.; Duarte, C. M.; Bailey, H.; Bailleul, F.; Bowen, W. D.; Caley, M. J., and Costa, D. P. 2016. "Key questions in marine megafauna movement ecology." *Trends in Ecology and Evolution* 31 (6):463-475.

Hochscheid, S.; Godley, B.; Broderick, A., and Wilson, R. P. 1999. "Reptilian diving: highly variable dive patterns in the green turtle Chelonia mydas." *Marine Ecology Progress Series*:101-112.

Horie, J.; Mitamura, H.; Ina, Y.; Mashino, Y.; Noda, T.; Moriya, K.; Arai, N., and Sasakura, T. 2017. "Development of a method for classifying and transmitting high-resolution feeding behavior of fish using an acceleration pinger." *Animal Biotelemetry* 5 (1):12.

Hussey, N.; MacNeil, M.; Olin, J.; McMeans, B.; Kinney, M.; Chapman, D., and Fisk, A. 2012. "Stable isotopes and elasmobranchs: tissue types, methods, applications and assumptions." *Journal of Fish Biology* 80 (5):1449-1484.

Hussey, N. E.; Kessel, S. T.; Aarestrup, K.; Cooke, S. J.; Cowley, P. D.; Fisk, A. T.; Harcourt, R. G.; Holland, K. N.; Iverson, S. J.; Kocik, J. F.; Mills Flemming, J. E., and Whoriskey, F. G. 2015. "Aquatic animal telemetry: A panoramic window into the underwater world." *Science* 348 (6240).

Huynh, T.; Blanke, U., and Schiele, B. 2007. "Scalable recognition of daily activities with wearable sensors." LoCA.

Japkowicz, N. 2000. "Learning from imbalanced data sets: a comparison of various strategies." AAAI workshop on learning from imbalanced data sets.

Jule, K. R.; Lea, S. E., and Leaver, L. A. 2009. "Using a behaviour discovery curve to predict optimal observation time." *Behaviour* 146 (11):1531-1542.

Kawabata, Y.; Noda, T.; Nakashima, Y.; Nanami, A.; Sato, T.; Takebe, T.; Mitamura, H.; Arai, N.; Yamaguchi, T., and Soyano, K. 2014. "Use of a gyroscope/accelerometer data logger to identify alternative feeding behaviours in fish." *Journal of Experimental Biology* 217 (18):3204-3208.

Kawabe, R.; Kawano, T.; Nakano, N.; Yamashita, N.; Hiraishi, T., and Naito, Y. 2003. "Simultaneous measurement of swimming speed and tail- beatsactivity of free-swimming rainbow trout Oncorhynchus mykiss using an acceleration data-logger." *Fisheries Science* 69 (5):959-965.

Khan, A.; Hammerla, N.; Mellor, S., and Plötz, T. 2016. "Optimising sampling rates for accelerometer-based human activity recognition." *Pattern Recognition Letters* 73:33-40.

Kröschel, M.; Reineking, B.; Werwie, F.; Wildi, F., and Storch, I. 2017. "Remote monitoring of vigilance behavior in large herbivores using acceleration data." *Animal Biotelemetry* 5 (1):10.

Kuhn, M. 2016. "A Short Introduction to the caret Package." accessed 07 January. https://cran.r-project.org/web/packages/caret/vignettes/caret.pdf

Ladds, M. A.; Thompson, A. P.; Kadar, J.-P.; Slip, D.; Hocking, D., and Harcourt, R. 2017. "Super machine learning: improving accuracy and reducing variance of behaviour classification from accelerometry." *Animal Biotelemetry* 5 (1):8.

Laich, A. G.; Wilson, R. P.; Quintana, F., and Shepard, E. L. 2008. "Identification of imperial cormorant Phalacrocorax atriceps behaviour using accelerometers." *Endangered Species Research* 10:29-37.

le Roux, S. P.; Marias, J.; Wolhuter, R., and Niesler, T. 2017. "Animal-borne behaviour classification for sheep (Dohne Merino) and Rhinoceros (Ceratotherium simum and Diceros bicornis)." *Animal Biotelemetry* 5 (1):25.

Lear, K. O.; Whitney, N. M.; Brewster, L. R.; Morris, J. J.; Hueter, R. E., and Gleiss, A. C. 2017. "Correlations of metabolic rate and body acceleration in three species of coastal sharks under contrasting temperature regimes." *Journal of Experimental Biology* 220 (Pt 3):397-407.

Leos-Barajas, V.; Photopoulou, T.; Langrock, R.; Patterson, T. A.; Watanabe, Y. Y.; Murgatroyd, M., and Papastamatiou, Y. P. 2017. "Analysis of animal accelerometer data using hidden Markov models." *Methods in Ecology and Evolution* 8 (2):161-173.

Liaw, A., and Wiener, M. 2002. "Classification and regression by randomForest." *R news* 2 (3):18-22.

López, L. M. M.; de Soto, N. A.; Miller, P., and Johnson, M. 2016. "Tracking the kinematics of caudal-oscillatory swimming: a comparison of two on-animal sensing methods." *Journal of Experimental Biology* 219 (14):2103-2109.

López, L. M. M.; Miller, P. J.; de Soto, N. A., and Johnson, M. 2015. "Gait switches in deep-diving beaked whales: biomechanical strategies for long-duration dives." *Journal of Experimental Biology* 218 (9):1325-1338.

Maia, A. M.; Wilga, C. A., and Lauder, G. V. 2012. *Biomechanics of locomotion in sharks, rays, and chimaeras*. 2 ed. Vol. 1, *Biology of sharks and their relatives*. CRC Press.

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., and Scantlebury, M. 2014. "Tri-axial accelerometers quantify behaviour in the Eurasian badger (Meles meles): towards an automated interpretation of field data." *Animal Biotelemetry* 2 (1):5.

Mitani, Y.; Sato, K.; Ito, S.; Cameron, M. F.; Siniff, D. B., and Naito, Y. 2003. "A method for reconstructing three-dimensional dive profiles of marine mammals using geomagnetic intensity data: results from two lactating Weddell seals." *Polar Biology* 26 (5):311-317.

Morrissey, J. F., and Gruber, S. H. 1993a. "Habitat selection by juvenile lemon sharks, Negaprion brevirostris." *Environmental Biology of Fishes* 38 (4):311-319.

Morrissey, J. F., and Gruber, S. H. 1993b. "Home range of juvenile lemon sharks, Negaprion brevirostris." *Copeia*:425-434.

Motta, P.; Tricas, T., and Summers, R. 1997. "Feeding mechanism and functional morphology of the jaws of the lemon shark Negaprion brevirostris (Chondrichthyes, Carcharhinidae)." *Journal of Experimental Biology* 200 (21):2765-2780.

Myrberg Jr, A. A., and Gruber, S. H. 1974. "The behavior of the bonnethead shark, Sphyrna tiburo." *Copeia*:358-374.

Nakamura, I.; Watanabe, Y. Y.; Papastamatiou, Y. P.; Sato, K., and Meyer, C. G. 2011. "Yo-yo vertical movements suggest a foraging strategy for tiger sharks Galeocerdo cuvier." *Marine Ecology Progress Series* 424:237-246.

Nathan, R.; Spiegel, O.; Fortmann-Roe, S.; Harel, R.; Wikelski, M., and 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." *Journal of Experimental Biology* 215 (Pt 6):986-96.

Noda, T.; Kawabata, Y.; Arai, N.; Mitamura, H., and Watanabe, S. 2013. "Monitoring escape and feeding behaviours of cruiser fish by inertial and magnetic sensors." *PLoS One* 8 (11):e79392.

Noda, T.; Kawabata, Y.; Arai, N.; Mitamura, H., and Watanabe, S. 2014. "Animal-mounted gyroscope/accelerometer/magnetometer: In situ measurement of the movement performance of fast-start behaviour in fish." *Journal of Experimental Marine Biology and Ecology* 451:55-68.

Noda, T.; Okuyama, J.; Koizumi, T.; Arai, N., and Kobayashi, M. 2012. "Monitoring attitude and dynamic acceleration of free-moving aquatic animals using a gyroscope." *Aquatic Biology* 16 (3):265-276.

Payne, N. L.; Iosilevskii, G.; Barnett, A.; Fischer, C.; Graham, R. T.; Gleiss, A. C., and Watanabe, Y. Y. 2016. "Great hammerhead sharks swim on their side to reduce transport costs." *Nature Communications* 7:12289.

Pillans, R. 2003. "Negaprion acutidens (sharptooth lemon shark)." accessed 01 May. http://www.iucnredlist.org/details/41836/0

Prasad, A. M.; Iverson, L. R., and Liaw, A. 2006. "Newer classification and regression tree techniques: bagging and random forests for ecological prediction." *Ecosystems* 9 (2):181-199.

Resheff, Y. S.; Rotics, S.; Harel, R.; Spiegel, O., and Nathan, R. 2014. "AcceleRater: a web application for supervised learning of behavioral modes from acceleration measurements." *Movement Ecology* 2 (1):27.

Robert, B.; White, B.; Renter, D., and Larson, R. 2009. "Evaluation of three-dimensional accelerometers to monitor and classify behavior patterns in cattle." *Computers and Electronics in Agriculture* 67 (1-2):80-84.

Ropert-Coudert, Y.; Grémillet, D.; Kato, A.; Ryan, P. G.; Naito, Y., and Le Maho, Y. 2004. "A fine-scale time budget of Cape gannets provides insights into the foraging strategies of coastal seabirds." *Animal Behaviour* 67 (5):985-992.

Ropert-Coudert, Y.; Kato, A.; Grémillet, D., and Crenner, F. 2012. "Bio-logging: recording the ecophysiology and behaviour of animals moving freely in their environment." *Sensors for Ecology: Towards Integrated Knowledge of Ecosystems* 1:17-41.

Rutz, C., and Hays, G. C. 2009. New Frontiers in Biologging Science. The Royal Society.

Sakamoto, K. Q.; Sato, K.; Ishizuka, M.; Watanuki, Y.; Takahashi, A.; Daunt, F., and Wanless, S. 2009. "Can ethograms be automatically generated using body acceleration data from free-ranging birds?" *PloS One* 4 (4):e5379.

Sato, K.; Mitani, Y.; Cameron, M. F.; Siniff, D. B., and Naito, Y. 2003. "Factors affecting stroking patterns and body angle in diving Weddell seals under natural conditions." *Journal of Experimental Biology* 206 (9):1461-1470.

Shamoune-Baranes, J.; Bom, R.; van Loom, E. E.; Ens, B. J.; Oosterbeek, K., and Bouten, W. 2012. "From sensor data to animal behaviour:an oystercatcher example." *Plos One* 7 (5).

Shepard, E. L. C.; Wilson, R. P.; Halsey, L. G.; Quintana, F.; Laich, A. G.; Gleiss, A. C.; Liebsch, N.; Myers, A. E., and Norman, B. 2008a. "Derivation of body motion via appropriate smoothing of acceleration data." *Aquatic Biology* 4:235-241.

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., and Myers, A. E. 2008b. "Identification of animal movement patterns using tri-axial accelerometry." *Endangered Species Research* 10:47-60.

Sherub, S.; Fiedler, W.; Duriez, O., and Wikelski, M. 2017. "Bio-logging, new technologies to study conservation physiology on the move: a case study on annual survival of Himalayan vultures." *Journal of Comparative Physiology A* 203 (6-7):531-542.

Sims, D. 2003. "Tractable models for testing theories about natural strategies: foraging behaviour and habitat selection of free-ranging sharks." *Journal of Fish Biology* 63 (s1):53-73.

Sundström, L. F. 2015. "Negaprion brevirostris.", accessed 01 May. http://dx.doi.org/10.2305/IUCN.UK.2015.RLTS.T39380A81769233.en

Sundström, L. F.; Gruber, S. H.; Clermont, S. M.; Correia, J. P.; De Marignac, J. R.; Morrissey, J. F.; Lowrance, C. R.; Thomassen, L., and Oliveira, M. T. 2001. "Review of elasmobranch behavioral studies using ultrasonic telemetry with special reference to the lemon shark, Negaprion brevirostris, around Bimini Islands, Bahamas." In *The behavior and sensory biology of elasmobranch fishes: an anthology in memory of Donald Richard Nelson*, 225-250. Springer.

Sur, M.; Suffredini, T.; Wessells, S. M.; Bloom, P. H.; Lanzone, M.; Blackshire, S.; Sridhar, S., and Katzner, T. 2017. "Improved supervised classification of accelerometry data to distinguish behaviors of soaring birds." *PloS One* 12 (4):e0174785.

Sutherland, W. J. 1998. "The importance of behavioural studies in conservation biology." *Animal behaviour* 56 (4):801-809.

Tanha, J.; Van Someren, M.; de Bakker, M.; Bouteny, W.; Shamoun-Baranesy, J., and Afsarmanesh, H. 2012. "Multiclass semi-supervised learning for animal behavior recognition from accelerometer data." 2012 IEEE 24th International Conference on Tools with Artificial Intelligence (ICTAI).

Tsuda, Y.; Kawabe, R.; H., T.; Y., M.; T., H.; K., Y., and K., N. 2006. "Monitoring the spawning behaviour of chum salmon with an acceleration data logger." *Ecology of Freshwater Fish* 15 (3):264-274.

Valletta, J. J.; Torney, C.; Kings, M.; Thornton, A., and Madden, J. 2017. "Applications of machine learning in animal behaviour studies." *Animal Behaviour* 124:203-220.

Verikas, A.; Gelzinis, A., and Bacauskiene, M. 2011. "Mining data with random forests: A survey and results of new tests." *Pattern Recognition* 44 (2):330-349.

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., and Maguire, I. E. 2015. "Prying into the intimate secrets of animal lives; software beyond hardware for comprehensive annotation in 'Daily Diary'tags." *Movement Ecology* 3 (1):29.

Walton, E.; Casey, C.; Mitsch, J.; Vázquez-Diosdado, J. A.; Yan, J.; Dottorini, T.; Ellis, K. A.; Winterlich, A., and Kaler, J. 2018. "Evaluation of sampling frequency, window size and sensor position for classification of sheep behaviour." *Royal Society Open Science* 5 (2):171442.

Wang, Y.; Nickel, B.; Rutishauser, M.; Bryce, C. M.; Williams, T. M.; Elkaim, G., and Wilmers, C. C. 2015. "Movement, resting, and attack behaviors of wild pumas are revealed by tri-axial accelerometer measurements." *Movement Ecology* 3 (1):2.

Ware, C.; Friedlaender, A. S., and Nowacek, D. P. 2011. "Shallow and deep lunge feeding of humpback whales in fjords of the West Antarctic Peninsula." *Marine Mammal Science* 27 (3):587-605.

Watanabe, S.; Izawa, M.; Kato, A.; Ropert-Coudert, Y., and Naito, Y. 2005. "A new technique for monitoring the detailed behaviour of terrestrial animals: a case study with the domestic cat." *Applied Animal Behaviour Science* 94 (1):117-131.

Whitney, N. M.; Papastamatiou, Y. P., and Gleiss, A. C. 2012. *Integrative multi-sensor tagging: emerging techniques to link elasmobranch behavior, physiology and ecology*. 2 ed. Vol. 1, *Biology of Sharks and Their Relatives*: CRC Press.

Whitney, N. M.; Papastamatiou, Y. P.; Holland, K. N., and Lowe, C. G. 2007. "Use of an acceleration data logger to measure diel activity patterns in captive whitetip reef sharks, Triaenodon obesus." *Aquatic Living Resources* 20 (4):299-305.

Whitney, N. M.; Pratt Jr, H. L.; Pratt, T. C., and Carrier, J. C. 2010. "Identifying shark mating behaviour using three-dimensional acceleration loggers." *Endangered Species Research* 10:71-82.

Williams, H.; Shepard, E.; Duriez, O., and Lambertucci, S. A. 2015. "Can accelerometry be used to distinguish between flight types in soaring birds?" *Animal Biotelemetry* 3 (1):45.

Williams, H. J.; Holton, M. D.; Shepard, E. L.; Largey, N.; Norman, B.; Ryan, P. G.; Duriez, O.; Scantlebury, M.; Quintana, F., and Magowan, E. A. 2017. "Identification of animal movement patterns using tri-axial magnetometry." *Movement Ecology* 5 (1):6.

Wilmers, C. C.; Nickel, B.; Bryce, C. M.; Smith, J. A.; Wheat, R. E., and Yovovich, V. 2015. "The golden age of bio-logging: how animal-borne sensors are advancing the frontiers of ecology." *Ecology* 96 (7):1741-1753.

Wilson, A. D. M.; Brownscombe, J. W.; Krause, J.; Krause, S.; Gutowsky, L. F. G., and Brooks, E. J. S. J. C. 2015. "Integrating network analysis, sensortags, and observation to understand shark ecology and behaviour." *Behavioural Ecology* 26 (6):1577-1586.

Wilson, R. P.; Shepard, E., and Liebsch, N. 2008. "Prying into the intimate details of animal lives: use of a daily diary on animals." *Endangered Species Research* 4 (1-2):123-137.

Wilson, R. P.; White, C. R.; Quintana, F.; Halsey, L. G.; Liebsch, N.; Martin, G. R., and Butler, P. J. 2006. "Moving towards acceleration for estimates of activity-specific metabolic rate in free-living animals: the case of the cormorant." *Journal of Animal Ecology* 75 (5):1081-1090.

Yoda, K.; Naito, Y.; Sato, K.; Takahashi, A.; Nishikawa, J.; Ropert-Coudert, Y.; Kurita, M., and Le Maho, Y. 2001. "A new technique for monitoring the behaviour of free-ranging Adelie penguins." *Journal of Experimental Biology* 204 (4):685-690.

Yoda, K.; Sato, K.; Niizuma, Y.; Kurita, M.; Bost, C.; Le Maho, Y., and Naito, Y. 1999. "Precise monitoring of porpoising behaviour of Adélie penguins determined using acceleration data loggers." *Journal of Experimental Biology* 202 (22):3121-3126.

Supplementary Material

Table S1 Predictor variables extracted for each sampling frequency, from each of the acceleration axes (X, Y & Z) used to train the RF ML algorithm.

Predictor variable	Label	Definition
Average Static	avg_X_Static avg_Y_Static avg_Z_Static	1 second means for static acceleration representing body posture in each axis
Average Dynamic	avg_X_Dynamic avg_Y_Dynamic avg_Z_Dynamic	1 second means for dynamic acceleration representing body movement in each axis
Overall Dynamic Body Acceleration	avg_ODBA	sum of absolute dynamic body acceleration in each axis
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
Minimum	min_X_Static min_Y_Static min_Z_Static min_X_Dynamic min_Y_Dynamic min_Z_Dynamic min_ODBA	minimum per one second values for static and dynamic acceleration in each axis and ODBA
Maximum	max_X_Static max_Y_Static max_Z_Static Max_X_Dynamic max_Y_Dynamic max_Z_Dynamic max_ODBA	maximum per one second values for static and dynamic acceleration in each axis and ODBA
Kurtosis	kurt_X_Static kurt_Y_Static kurt_Z_Static kurt _X_Dynamic kurt_Y_Dynamic kurt_Acc_ZDynamic kurt_ODBA	measure of weight of tailedness relative to normal distribution for static and dynamic acceleration in each axis and ODBA
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
Cycle	Cycle	Inverse frequency of the dominant cycle
Amplitude	Amp	Amplitude of dominant cycle

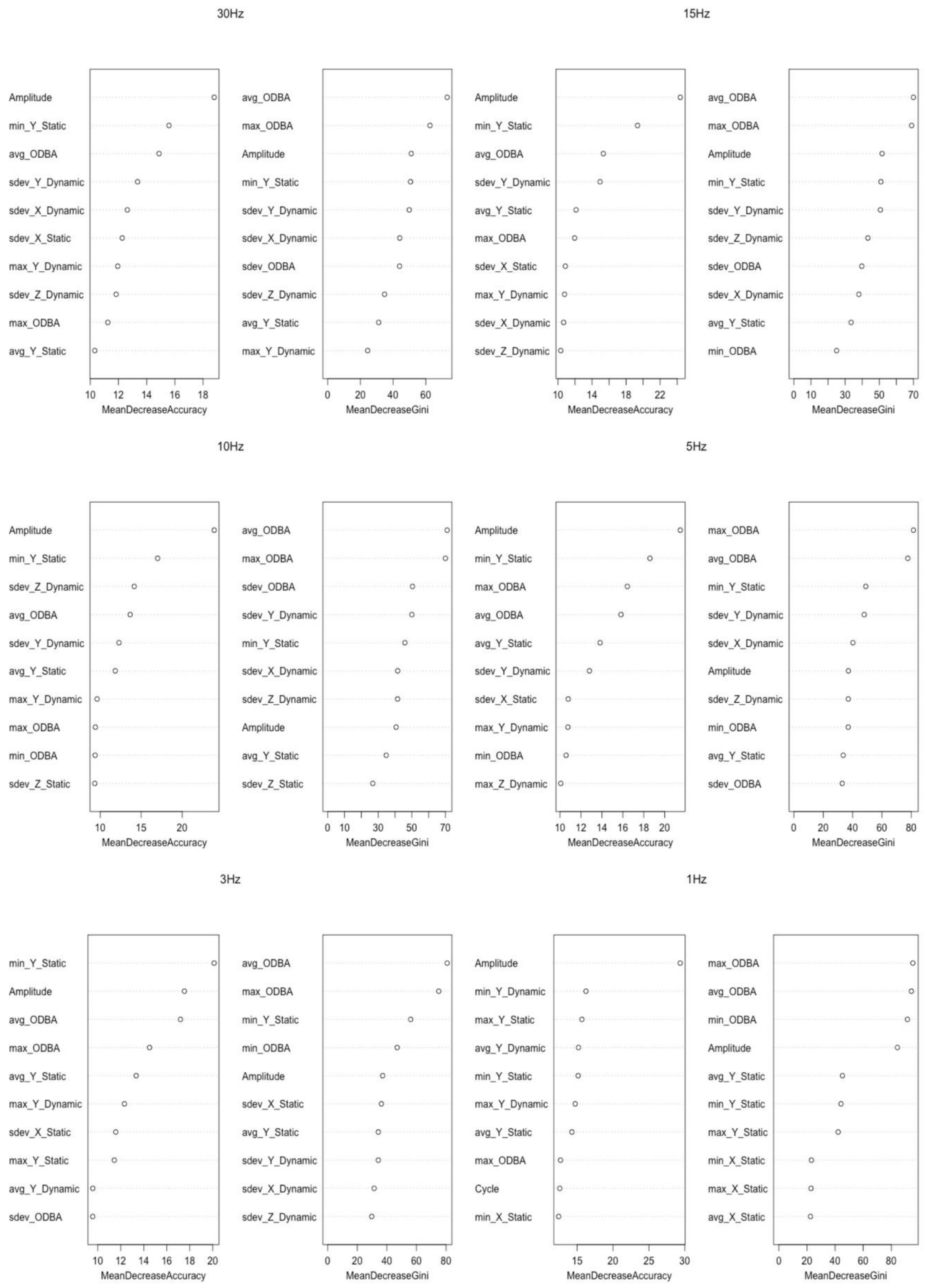

Figure S1 Variable importance plots for top ten important predictor variables from RF classification of accelerometer data used for predicting behaviour collected from lemon sharks (*N.brevirostris*). Higher values of mean decrease in accuracy and mean decrease in Gini index indicate higher importance for that predictor variable to the classification model.

Table S2 Predictor variables extracted from features of the accelerometer data (*n* =45)

Predictor variable	**Label**	**Definition**
Average Static	avg_Acc_XStatic avg_Acc_YStatic avg_Acc_ZStatic	1 second means for static acceleration representing body posture in each axis
Average Dynamic	avg_Acc_XDynamic avg_Acc_YDynamic avg_Acc_ZDynamic	1 second means for dynamic acceleration representing body movement in each axis
Overall Dynamic Body Acceleration	avg_Acc_ODBA	sum of absolute dynamic body acceleration in each axis
Standard deviation	sdev_Acc_XStatic sdev_Acc_YStatic sdev_Acc_ZStatic sdev_Acc_XDynamic sdev_Acc_YDynamic sdev_Acc_ZDynamic sdev_Acc_ODBA	standard deviation of static and dynamic acceleration in each axis and ODBA
Minimum	min_Acc_XStatic min_Acc_YStatic min_Acc_ZStatic min_Acc_XDynamic min_Acc_YDynamic min_Acc_ZDynamic min_Acc_ODBA	minimum per one second values for static and dynamic acceleration in each axis and ODBA
Maximum	max_Acc_XStatic max_Acc_YStatic max_Acc_ZStatic max_Acc_XDynamic max_Acc_YDynamic max_Acc_ZDynamic max_Acc_ODBA	maximum per one second values for static and dynamic acceleration in each axis and ODBA
Kurtosis	kurt_Acc_XStatic kurt_Acc_YStatic kurt_Acc_ZStatic kurt_Acc_XDynamic kurt_Acc_YDynamic kurt_Acc_ZDynamic kurt_Acc_ODBA	measure of weight of tailedness relative to normal distribution for static and dynamic acceleration in each axis and ODBA
Skewness	skew_Acc_XStatic skew_Acc_YStatic skew_Acc_ZStatic skew_Acc_XDynamic skew_Acc_YDynamic skew_Acc_ZDynamic skew_Acc_ODBA	measure of asymmetry about mean for static and dynamic acceleration in each axis and ODBA
TBF	Acc_TBF	Frequency of the dominant cycle
Cycle	Acc_Cycle	Inverse frequency of the dominant cycle
Amplitude	Acc_Amp	Amplitude of dominant cycle

Table S3 Predictor variables extracted from features of the magnetometer data (*n* = 39)

Predictor variable	**Label**	**Definition**
Average Static	avg_Mag_XStatic avg_Mag_YStatic avg_Mag_ZStatic	1 second means for static representing body posture in each axis
Average Dynamic	avg_Mag_XDynamic avg_Mag_YDynamic avg_Mag_ZDynamic	1 second means for dynamic representing body movement in each axis
Standard deviation	sdev_Mag_XStatic sdev_Mag_YStatic sdev_Mag_ZStatic sdev_Mag_XDynamic sdev_Mag_YDynamic sdev_Mag_ZDynamic	standard deviation of static and dynamic acceleration in each axis
Minimum	min_Mag_XStatic min_Mag_YStatic min_Mag_ZStatic min_Mag_XDynamic min_Mag_YDynamic min_Mag_ZDynamic	minimum per one second values for static and dynamic acceleration in each axis
Maximum	max_Mag_XStatic max_Mag_YStatic max_Mag_ZStatic max_Mag_XDynamic max_Mag_YDynamic max_Mag_ZDynamic	maximum per one second values for static and dynamic acceleration in each axis
Kurtosis	kurt_Mag_XStatic kurt_Mag_YStatic kurt_Mag_ZStatic kurt_Mag_XDynamic kurt_Mag_YDynamic kurt_Mag_ZDynamic	measure of weight of tailedness relative to normal distribution for static and dynamic acceleration in each axis
Skewness	skew_Mag_XStatic skew_Mag_YStatic skew_Mag_ZStatic skew_Mag_XDynamic skew_Mag_YDynamic skew_Mag_ZDynamic	measure of asymmetry about mean for static and dynamic acceleration in each axis
TBF	Mag_TBF	Frequency of the dominant cycle
Cycle	Mag_Cycle	Inverse frequency of the dominant cycle
Amplitude	Mag_Amp	Amplitude of dominant cycle

Table S4 Predictor variables extracted from features of the gyroscope data ($n = 21$)

Predictor variable	**Label**	**Definition**
Average	avg_Gyro_XRaw avg_Gyro_YRaw_ avg_Gyro_ZRaw	1 second means for static acceleration representing body posture in each axis
Standard deviation	sdev_Gyro_XRaw sdev_Gyro_YRaw sdev_Gyro_ZRaw	standard deviation of static and dynamic acceleration in each axis and ODBA
Minimum	min_Gyro_XRaw min_Gyro_YRaw min_Gyro_ZRaw	minimum per one second values for static and dynamic acceleration in each axis and ODBA
Maximum	max_Gyro_XRaw max_Gyro_YRaw max_Gyro_ZRaw	maximum per one second values for static and dynamic acceleration in each axis and ODBA
Kurtosis	kurt_Gyro_XRaw kurt_Gyro_YRaw kurt_Gyro_ZRaw	measure of weight of tailedness relative to normal distribution for static and dynamic acceleration in each axis and ODBA
Skewness	skew_Gyro_XRaw skew_Gyro_YRaw skew_Gyro_ZRaw	measure of asymmetry about mean for static and dynamic acceleration in each axis and ODBA
Cycle	Gyro_Cycle	inverse frequency of the dominant cycle
TBF	Gyro_TBF	frequency of the dominant cycle
Amplitude	Gyro_Amp	Amplitude of dominant cycle