RESEARCH ARTICLE

High accuracy at low frequency: detailed behavioural classification from accelerometer data

Jack Tatler[1,*], Phillip Cassey[1] and Thomas A. A. Prowse[2]

ABSTRACT

Accelerometers are a valuable tool for studying animal behaviour and physiology where direct observation is unfeasible. However, giving biological meaning to multivariate acceleration data is challenging. Here, we describe a method that reliably classifies a large number of behaviours using tri-axial accelerometer data collected at the low sampling frequency of 1 Hz, using the dingo (*Canis dingo*) as an example. We used out-of-sample validation to compare the predictive performance of four commonly used classification models (random forest, *k*-nearest neighbour, support vector machine, and naïve Bayes). We tested the importance of predictor variable selection and moving window size for the classification of each behaviour and overall model performance. Random forests produced the highest out-of-sample classification accuracy, with our best-performing model predicting 14 behaviours with a mean accuracy of 87%. We also investigated the relationship between overall dynamic body acceleration (ODBA) and the activity level of each behaviour, given the increasing use of ODBA in ecophysiology as a proxy for energy expenditure. ODBA values for our four 'high activity' behaviours were significantly greater than all other behaviours, with an overall positive trend between ODBA and intensity of movement. We show that a random forest model of relatively low complexity can mitigate some major challenges associated with establishing meaningful ecological conclusions from acceleration data. Our approach has broad applicability to free-ranging terrestrial quadrupeds of comparable size. Our use of a low sampling frequency shows potential for deploying accelerometers over extended time periods, enabling the capture of invaluable behavioural and physiological data across different ontogenies.

KEY WORDS: Accelerometer, Animal behaviour, Classification model, ODBA, Random forest

INTRODUCTION

The foundation of animal ecology is understanding how individuals interact with their abiotic and biotic environment. These interactions are increasingly being measured with bio-logging techniques, where biological data are recorded remotely from devices attached to animals. This approach has allowed researchers to answer questions on everything from hunting tactics of puma (Williams et al., 2014) to energy expenditure in cormorants (Gómez Laich et al., 2011) and diving behaviour in whales (Ishii et al., 2017).

Consequently, the ability to continuously 'observe' free-ranging animals has facilitated the development and exploration of entirely new theories (Wilmers et al., 2015).

Accelerometers are a valuable tool in bio-logging research as they provide quantitative measurements of animal behaviour and physiology where direct observation is not possible or logistically feasible. The use of accelerometers mitigates some of the major challenges associated with studying the behaviour of wild animals, such as extensive time investment, animal disturbance and observer bias. Accelerometers measure acceleration (gravitational and inertial) caused by animal movement in different planes, allowing the development of classification models calibrated to predict behavioural states such as resting, walking, swimming and eating (e.g. Pagano et al., 2017). Further, there is a strong linear relationship between body acceleration and energy expenditure in many taxa, which is of particular interest to ecophysiologists (Halsey and White, 2010; Wilson et al., 2006; Halsey et al., 2009). Although accelerometry has been used to study animal movement and behaviour for almost two decades (Yoda et al., 1999), recent methodological advancements have increased its accessibility and appeal to a broader scientific community.

Classifying animal behaviours to high-frequency acceleration data presents a suite of new and complex challenges. One approach is unsupervised machine learning, in which pattern-recognition algorithms identify different states directly from the accelerometer signatures. Unsupervised learning is intrinsically challenging so algorithms are frequently used to 'learn' the relationship between acceleration data and behaviour using a model-training dataset that is acquired from direct observation. The ability of the algorithm to interpret this relationship depends largely on the variables used to characterise the raw acceleration data. Several attempts to simplify or streamline this approach have been made, with varying success. Ladds et al. (2017) introduced a super-machine-learning method that identified six behaviours in four species of pinniped with approximately 73% accuracy. They used a high sampling frequency (25 Hz), large training dataset (∼90,000 individual data points) and a very large set of input variables (*n*=147). In contrast, when using fewer input variables and the relatively simple approach (*k*-nearest neighbour), McClune et al. (2014) classified four behaviours in Eurasian badgers (*Meles meles*) with an overall classification accuracy of 89%. In general, it is expected that the classification accuracy of a model will increase when using: (a) higher sampling frequencies; (b) more training data; and (c) broader behaviour categories (i.e. fewer behaviours to be classified). The consequence of following these criteria is not only increased computational time and difficulty, but loss of behavioural diversity and decreased deployment time on free-ranging animals due to memory constraints, i.e. the exact opposite of what researchers are aiming for. Reducing the sampling frequency would greatly increase deployment time (e.g. from days to months) whilst also decreasing computational effort. However, it is challenging to accurately

[1]School of Biological Sciences and Centre for Applied Conservation Science, University of Adelaide, Adelaide, SA 5005, Australia. [2]School of Mathematical Sciences, University of Adelaide, Adelaide, SA 5005, Australia.

*Author for correspondence (jack.tatler@adelaide.edu.au)

J.T., 0000-0002-8380-3612; P.C., 0000-0002-2626-0172

classify a broad range of behaviours using very low sampling rates. If we can create a simple model that overcomes the aforementioned hurdles, we will greatly improve integration with other fields, such as movement ecology and physiology.

One major weakness in applying machine-learning algorithms to acceleration data is that, for accurate and reliable identification of different behaviours, a period of observation is required to 'train' the algorithm. Therefore, it has only been possible to use this approach on species that can be observed whilst simultaneously recording their acceleration. Campbell et al. (2013) made an important step in overcoming this problem by demonstrating the potential of 'surrogacy', whereby a classification model was trained with behavioural observations from one species, and accurately predicted these behaviours in other species that possessed similar morphometrics.

In this study, we describe an approach to the classification of behaviours using accelerometer data collected at the very low sampling frequency of 1 Hz. We used the dingo (*Canis dingo* Meyer 1793), a medium-sized prototypical quadruped, as an example because it readily exhibits behaviours akin to its wild conspecifics. We used out-of-sample validation to compare the predictive performance of four commonly used classification models (random forest, *k*-nearest neighbour, support vector machine and naïve Bayes). We then tested the importance of predictor variables for the classification of each behaviour as well as overall model performance. We expected behaviours that were functionally similar, such as lateral and sternal recumbency, would produce similar acceleration signatures and thus be more difficult to classify accurately. Given the increasing use of overall dynamic body acceleration (ODBA) as a measure of activity and as a proxy for energy expenditure (Wilmers et al., 2015), we anticipated that ODBA would show a strong, positive relationship with intensity of movement.

MATERIALS AND METHODS
Data collection
Captive observations were conducted at Cleland Wildlife Park, Adelaide (34.9667°S, 138.6968°E) from August 2016–March 2017 under a University of Adelaide Animal Ethics permit (S-2015-177a). We used three captive-born adult male dingoes (c. 19 kg) that were kept on permanent display in a 2500 m^2 outdoor enclosure. We fitted each dingo with a tri-axial accelerometer (LISD2H, ST Microelectronics, USA) built into a custom-made GPS collar (Telemetry Solutions, Concord, CA, USA). The tri-axial accelerometer was programmed to sample changes in acceleration at 1 Hz (one sample per second) and orientated so that the *x*-, *y*- and *z*-axes recorded acceleration along the sway, heave and surge planes, respectively (Fig. 1). Dingo movement was recorded continuously with the accelerometer and visually with a camcorder at 30 frames s^{-1} (Sanyo Dual Camera Xacti CG10 HD) for eight sessions of ca. 30 min each. Behaviours directly observed from the video footage were manually annotated into a Microsoft Excel spreadsheet by one observer (J.T.) and matched to the corresponding accelerometer data via concurrent timestamps. We synchronised the accelerometer and camera clocks by setting them using the same laptop computer (internet time server) on the morning of each session. Prior to manual annotation, we consulted the timestamp from auxiliary footage (iPhone 8 also set to the internet time server, 30 frames s^{-1}) to confirm the syncing of our devices. Although a handling keeper was present at all times, the focal animal (only ever one dingo per session) was unrestricted and conducted behaviours largely *ad libitum*. Prior exposure to commercial dog collars ensured that the dingoes did not act
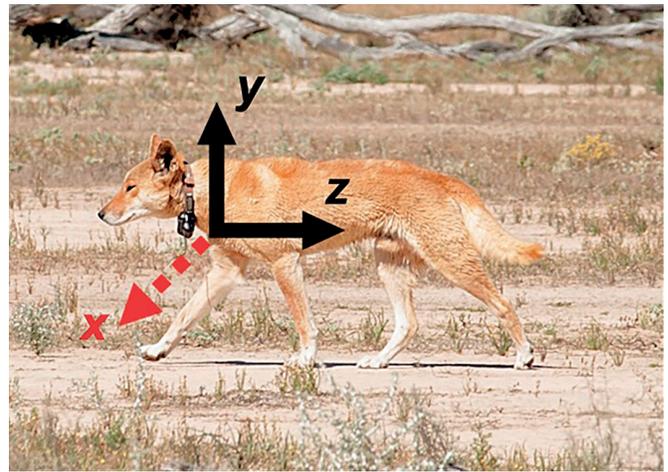

Fig. 1. Dingo wearing a global positioning system (GPS)-accelerometer collar. Arrows indicate axis orientation of the accelerometer (*x*, sway or side-to-side; *y*, heave or up-and-down; *z*, surge or back-and-forth). Photo credit: Casey O'Brien.

atypically during the sampling sessions (Hayley Wells, Cleland Wildlife Park, personal communication).

Determining behaviours from acceleration values requires a sampling frequency that is at least twice as fast as the observed behaviour (Nyquist sampling theorem). Thus, our core criteria for what constituted a behaviour was a repeated movement that consistently lasted two or more seconds. We observed 14 such behaviours and annotated them to 9360 accelerometer data points (equivalent to 156 min) on a per-second basis (Table S1). Quick transitory movements, between recognised behavioural states, were assigned to the behaviour (pre- or post-transition) that was mostly common across the 30 frames s^{-1}. We excluded any behaviours that had sample sizes <20 data points or were clearly observed to be influenced by physical interaction with the keeper. Based on direct observation of dingo movement, each behaviour was broadly assigned to an activity level: low, medium or high (Table 1).

Variable derivation
The ability of classification models to distinguish between behavioural states depends partly on the predictor variables used to characterise the raw acceleration signals. We adopted a comprehensive approach to selecting predictor variables by calculating an extensive list of derived variables (*n*=66) from the *x*-, *y*- and *z*-axes. These ranged from simple metrics such as the mean and standard deviation of an axis, to more complex, derived variables such as waveform length and signal magnitude area (Table 2). All predictor variables were calculated using a moving window centred on each data point (see detailed description given in 'Model evaluation').

Classification modelling
We used supervised machine-learning techniques to fit classification models that used different combinations of the predictor variables. In supervised learning, an algorithm is employed to learn the relationship between a given set of input and output variables (our predictor variables and manually assigned behaviours, respectively) so that, when provided with a new set of input variables, it can predict what the output variables will be. With the goal of finding a reliable method that would be straightforward to implement, we compared four supervised machine-learning

Table 1. Descriptions of all 14 behaviours that met the selection criteria

Activity	Behaviour	Description of movement
Low	Lat. lying inactive	Recumbent on flank with head down on the ground
	Stern. lying inactive	Recumbent on sternum with head down on the ground
	Lat. lying alert	Recumbent on flank with head and neck upright
	Stern. lying alert	Recumbent on sternum with head and neck upright
	Sitting	Sedentary, rump on ground, front legs held straight under body
	Standing	Sedentary, with all four legs held straight under the body
Medium	Drinking	Body in standing position, head lowered; rapid jaw movements
	Lat. lying groom	Recumbent on flank with rhythmic head movement
	Searching	Omni-directional movements, head low and focussed on ground
	Walking	Symmetric diagonal gait at a slow pace
	Collar discomfort	Standing/walking while jerking neck from side to side
High	Trotting	Symmetric diagonal gait; faster than a walk
	Playing	High-intensity interactions with conspecifics
	Running	Gallop-type movement at a very fast pace

Activity classes were assigned *post hoc*, based on direct observation of behaviours and intensity of movement. Lat., lateral; Stern., sternum.

algorithms using the R software environment for statistical and graphical computing (R Core Team 2016: http://www.R-project.org/). The *k*-nearest neighbour (*k*-NN; R library 'class'; Venables and Ripley, 2002) is a simple algorithm that employs a number of nearest neighbours (defined by the parameter *k*) to contribute to the classification of a sample. The majority of behaviours within *k* observations surrounding the data point being classified determines the behaviour of that data point (Coomans and Massart, 1982). Naïve Bayes (R library 'e1071', https://rdrr.io/rforge/e1071/) is a

probabilistic classifier that computes the conditional probabilities of a categorical class variable given independent predictor variables using Bayes' rule. A support vector machine (SVM: R library 'e1071') constructs an optimal hyperplane to separate patterns, or classes, in the data (Vapnik, 1999). Non-linear classification is achieved using kernel functions (chosen *a priori*), which nonlinearly map the input vectors into a very high-dimensional feature space. A random forest (RF; R library 'randomForest', Liaw and Wiener, 2002) is an ensemble method for classification in which a set of decision trees are constructed that are then used to classify a new instance according to the majority vote (Breiman, 2001). The number of decision trees needed generally increases with the number of predictor variables used. Each of these modelling approaches are widely used, computationally inexpensive, and represent different degrees of complexity to pattern recognition and classification.

Model evaluation

For each machine-learning algorithm, we evaluated a candidate set of models that ranged in complexity from a 'null' model containing just the *x*-, *y*- and *z*-axes (*n*=3 variables) to the most complex model (*n*=69 variables; Table S2). We tested six different moving windows for variable derivation (4, 8, 16, 32, 64 and 128 s). We also explored how the predictive ability of our four models was affected when we employed a different number of nearest neighbours (1, 3, 5, 7 and 9; *k*–NN), kernels (linear, radial and polynomial; SVM) and number of classification trees (500, 1000, 2500, 5000, 7500 and 10,000; RF). To measure the predictive performance of each model, we averaged the out-of-sample accuracy (see below) achieved across 10 repeated training-test splits, in each case using a random 90% of the data from each behaviour to train our model and the remaining 10% for testing.

For each test datum, predicted behaviours were labelled as true positive (TP) if they correctly matched the actual behaviour, true negative (TN) if they correctly identified as a different behaviour, false positive (FP) if the behaviour was incorrectly identified, and

Table 2. Descriptions of the predictor variables that were used to fit four different classification algorithms in an attempt to classify dingo behaviours

Parameter	Description
Axes	*x* (side to side), *y* (back and forth), *z* (up and down)
s.d.	Measures the spread of the signal for each axis or statistic
Magnitude	Magnitude of acceleration: $\sqrt{x^2 + y^2 + z^2}$
Signal magnitude area	A measure of movement intensity within all three axes: $SMA = \frac{1}{N}\left(\sum_{i=1}^{N}|x_i| + \sum_{i=1}^{N}|y_i| + \sum_{i=1}^{N}|z_i|\right)$
Waveform length	The total amount of variance within the signal through the cumulative measure of amplitude, frequency and duration: $WL = \frac{1}{N}\left(\sum_{i=1}^{N-1}|x_{i+1} - x_i| + \sum_{i=1}^{N}|y_{i+1} - y_i| + \sum_{i=1}^{N}|z_{i+1} - z_i|\right)$
Kurtosis	Measure of weight of the tails relative to a normal distribution
Skew	Measure of the lack of symmetry of the distribution
Correlation	Degree and type of pairwise relationship between the three axes
Dynamic body acceleration (DBA)	Static acceleration, calculated by subtracting a running mean from the raw acceleration data from each axis
Overall dynamic body acceleration	The absolute sum of DBA for each axis: $ODBA = |DBA \cdot x| + |DBA \cdot y| + |DBA \cdot z|$
Vectorial dynamic body acceleration	Vector of the DBA: $VeDBA = \sqrt{DBA \cdot x^2 + DBA \cdot y^2 + DBA \cdot z^2}$
Difference	The difference between successive values on each axis
Absolute values	Absolute value of the chosen parameters
Mean	Mean, calculated over a moving window, of the chosen parameters
Range	Minimum and maximum values calculated over the moving window

false negative (FN) if they incorrectly identified as a different behaviour. We evaluated the predictive ability of our models using three measures of accuracy: the true skill statistic (TSS; Eqn 1), Matthew's correlation coefficient (MCC; Eqn 2) and the *F*-measure (Eqn 3). The TSS was introduced by Allouche et al. (2006) as an improvement to the widely used kappa, whereby it not only accounts for both omission and commission errors, but is not affected by the sample size of each class:

$$\text{TSS} = \frac{\text{TP}}{(\text{TP} + \text{FN})} + \frac{\text{TN}}{(\text{TN} + \text{FP})} - 1. \quad (1)$$

Similarly, we chose MCC because it is a balanced measure of accuracy when sample sizes are unequal and has been more widely used in bioinformatics than TSS:

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP}) \times (\text{TP} + \text{FN}) \times (\text{TN} + \text{FP}) \times (\text{TN} + \text{FN})}}. \quad (2)$$

The TSS and MCC both return a value between −1 (total disagreement between prediction and observation) and +1 (perfect prediction), with a value of 0 indicating a prediction no better than random chance. We included the *F*-measure, calculated as the harmonic mean of precision and sensitivity, because it is a relatively simple statistic that has been widely used in the literature, thus placing our results in the context of previous studies:

$$F\text{-measure} = \frac{2 \times \text{TP}}{(2 \times \text{TP}) + \text{FP} + \text{FN}}. \quad (3)$$

We calculated three additional statistics that provide added insight into model performance. These were: precision, denoting the proportion of positive classifications that were correctly identified (Eqn 4); sensitivity, which is the probability that a behaviour will be correctly classified (Eqn 5); and specificity, the probability that a behaviour has been correctly classified as a different behaviour (Eqn 6):

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}; \quad (4)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}; \quad (5)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \quad (6)$$

Classification models produce a corresponding probability to their behavioural predictions, to which we apply a threshold criterion (0.1–0.9) that determines the rate of TP, TN, FP or FN. The threshold is usually used to fine-tune model parameters such as sensitivity and specificity. Therefore, it is important to choose a threshold based on the research questions and consequences associated with practical application of the model. Given that the intended practical application of this research is to predict behaviours of free-ranging animals, we chose a threshold that would maximise our TSS score whilst minimising the amount of unclassified data points.

Overall dynamic body acceleration

Dynamic body acceleration (DBA) was calculated by subtracting a running mean from each acceleration axis to give acceleration values occurring from inertia (i.e. movement). We chose a running mean of 4 s (i.e. four data points) because it was roughly half the length of our most active behaviour (running) and, thus, we minimise any loss of resolution for each behaviour. The absolute value for each axis (DBA) was summed to give ODBA, an overall value for dynamic acceleration. To determine whether ODBA differed between the observed behaviours, and whether there was a positive relationship between ODBA and activity level, we conducted an ANOVA and Tukey's test for paired comparisons. All analyses were conducted in the R software environment for statistical and graphical computing (http://www.R-project.org/).

RESULTS

Across the four machine-learning algorithms that we tested, the RF classification models produced superior out-of-sample validation scores (Fig. 2). The top 50 classification models were all achieved using the RF algorithm. Despite differing considerably in their complexity (Table S3), the predictive capacity of these models was similar (Δ mean TSS≤0.04). Our 'best' model, which ranked third overall, was selected due to its low number of predictor variables (26 of a possible 69) and classification trees (1000), in conjunction with returning the lowest range in TSS scores between the 14 behaviours (Table 3). This selected RF model predicted all 14 dingo behaviours with high accuracy (mean TSS=0.87).

Comparisons of different RF models indicated that the predictor variable set used had the strongest influence on the ability of RFs to predict dingo behaviours (Fig. 3). When models were fitted with just the *x*-, *y*- and *z*-axes (predictor variable set 1), they produced the lowest predictive accuracy. The RF models that were constructed using a moving window of ≥16 s were substantially better at classifying behaviours, whereas varying the number of classification trees had little effect on model accuracy (Fig. 3). The *z*-axis, which measured surge movement, was highly variable between behaviours and therefore was particularly important for classification (Fig. 4). Further, predictors that described the variation (s.d.) and characterised the distribution of sway movements proved to be valuable for classifying behaviours (Fig. 5). We attempted to refine the model further by excluding the variables that contributed least to the model
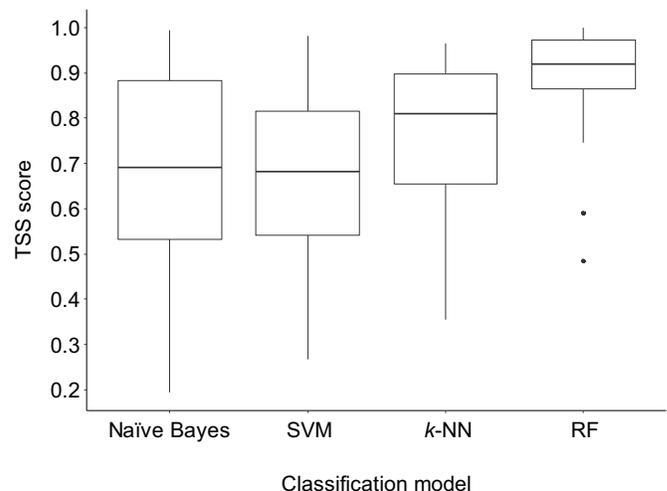
Fig. 2. The predictive ability [true skill statistic (TSS) score] of the best naïve Bayes, support vector machine (SVM), *k*-nearest neighbour (*k*-NN) and random forest (RF) models (no threshold). Each boxplot shows the range of TSS scores from the 14 behaviours classified by each algorithm. Lower and upper hinges correspond to the first and third quartiles, and the whiskers extend from the hinge to no more than 1.5× interquartile range (IQR). Outliers are represented individually by dots.

Table 3. Performance of our selected random forest (RF) model at predicting 14 different behaviours observed in captive dingoes

Behaviour	TSS	MCC	F-measure	Precision	Sensitivity	Specificity
Lat. lying inactive	0.99	0.99	0.99	0.99	0.99	1.00*
	[0.98, 0.99]	[0.98, 0.99]	[0.98, 0.99]	[0.98, 0.99]	[0.98, 0.99]	[1.00, 1.00]
Stern. lying inactive	0.90	0.94	0.94	0.99	0.90	1.00*
	[0.88, 0.92]	[0.93, 0.95]	[0.92, 0.95]	[0.98, 0.99]	[0.88, 0.92]	[1.00, 1.00]
Lat. lying alert	0.96	0.98	0.98	0.99	0.96	1.00*
	[0.95, 0.97]	[0.97, 0.98]	[0.97, 0.98]	[0.99, 1.00]	[0.95, 0.97]	[1.00, 1.00]
Stern. lying alert	0.96	0.97	0.97	0.98	0.97	1.00*
	[0.96, 0.97]	[0.97, 0.97]	[0.97, 0.98]	[0.98, 0.98]	[0.96, 0.97]	[1.00, 1.00]
Sitting	0.91	0.95	0.95	1.00*	0.91	1.00*
	[0.89, 0.92]	[0.94, 0.95]	[0.94, 0.95]	[0.99, 1.00]	[0.89, 0.92]	[1.00, 1.00]
Standing	0.85	0.84	0.91	0.88	0.93	0.92
	[0.85, 0.85]	[0.84, 0.85]	[0.90, 0.91]	[0.88, 0.89]	[0.93, 0.94]	[0.91, 0.92]
Drinking	0.97	0.97	0.96	0.96	0.98	1.00*
	[0.96, 0.99]	[0.95, 0.98]	[0.95, 0.98]	[0.94, 0.98]	[0.96, 0.99]	[1.00, 1.00]
Lat. lying groom	0.97	0.97	0.97	0.97	0.97	1.00*
	[0.95, 0.98]	[0.96, 0.98]	[0.96, 0.97]	[0.96, 0.98]	[0.95, 0.98]	[1.00, 1.00]
Searching	0.91	0.90	0.91	0.90	0.92	0.99
	[0.91, 0.91]	[0.90, 0.90]	[0.91, 0.91]	[0.90, 0.91]	[0.92, 0.93]	[0.99, 0.99]
Walking	0.76	0.76	0.81	0.81	0.81	0.95
	[0.76, 0.77]	[0.76, 0.76]	[0.81, 0.81]	[0.80, 0.81]	[0.81, 0.82]	[0.95, 0.95]
Collar discomfort	0.75	0.85	0.83	1.00	0.75	1.00
	[0.70, 0.80]	[0.82, 0.88]	[0.79, 0.87]	[1.00, 1.00]	[0.70, 0.80]	[1.00, 1.00]
Trotting	0.46	0.57	0.57	0.72	0.47	0.99
	[0.44, 0.48]	[0.55, 0.59]	[0.55, 0.59]	[0.70, 0.75]	[0.45, 0.49]	[0.99, 0.99]
Playing	0.88	0.91	0.91	0.95	0.89	1.00*
	[0.88, 0.89]	[0.91, 0.92]	[0.91, 0.92]	[0.94, 0.96]	[0.88, 0.89]	[1.00, 1.00]
Running	0.62	0.66	0.67	0.73	0.62	0.99
	[0.59, 0.64]	[0.65, 0.68]	[0.65, 0.69]	[0.72, 0.74]	[0.60, 0.65]	[0.99, 0.99]

We used six widely used techniques for quantitatively assessing the model's predictive ability, at a threshold of 0.3. Our selected model was fit using predictor variable set 4, a moving window of 64 s, and number of classification trees of 1000. 95% confidence intervals are presented in square brackets beneath each metric. TSS, true skill statistic; MCC, Matthew's correlation coefficient; Lat., lateral; Stern., sternum.
*Denotes values that scored 1.00 after rounding to two decimal places.

(mean decrease in accuracy ≤60%; n=8) but found a reduction in the ability of the model to identify dingo behaviours (mean values: TSS=0.85, MCC=0.87, F-measure=0.88, precision=0.91, sensitivity=0.86 and specificity=0.99).

Overall, our selected model performed better at classifying low-intensity, stationary behaviours. Specifically, all behaviours where

dingoes were lying down were identified with very high accuracy (TSS>0.90). In contrast, upright and more dynamic movements such as trotting and running were classified less well (TSS=0.46 and 0.62, respectively). We observed high specificity for each of our behaviours (0.92–1.00) and thus our selected model was robust to misclassification. Although the majority of behaviours exhibited a
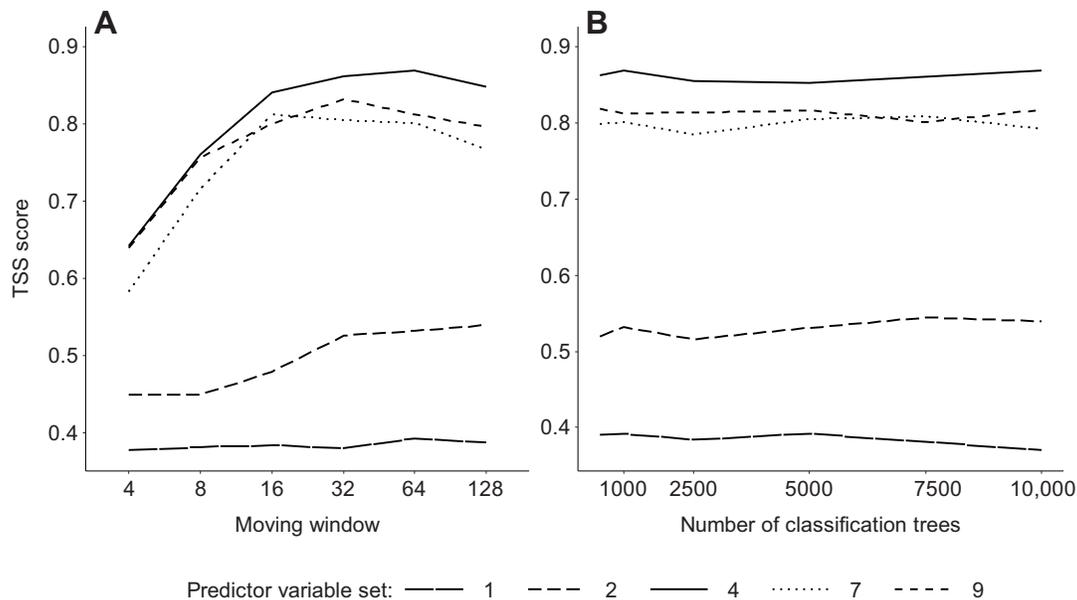

Fig. 3. The importance of the predictor variable set on the predictive ability. Mean TSS score from 10-fold out-of-sample validation of RF models when fit with (A) an increasing moving window and fixed number of classification trees (n_{tree}=1000; as per selected model), and (B) an increasing number of classification trees but a fixed moving window of 64 s (as per selected model).
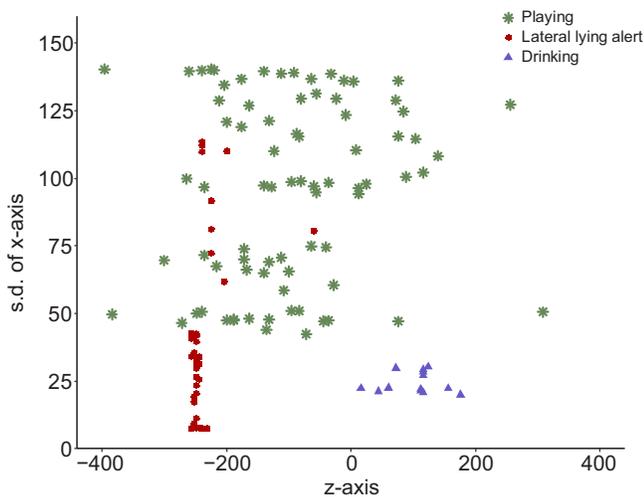
Fig. 4. Clustering of data points demonstrate the importance of predictor variables for classifying behaviours in our selected RF model. One behaviour from each activity level was chosen as an example: Lateral lying alert=low activity, drinking=medium activity and playing=high activity.

sensitivity above 0.90, low sensitivity for trotting and running indicated that the model had difficulty with positive classification of these behaviours (Table 3).

A threshold of 0.3 produced the model with the optimal balance between sensitivity and specificity (TSS score) and unclassified data points (Fig. S1). At this threshold, the overall number of incorrectly classified behaviours was very low, with higher classification errors occurring in the more active behaviours. Misclassifications produced by the model tended to confuse closely related behaviours; for example, 'trotting' most often misclassified as 'walking', and 'running' misclassified as 'trotting' (Table S4).
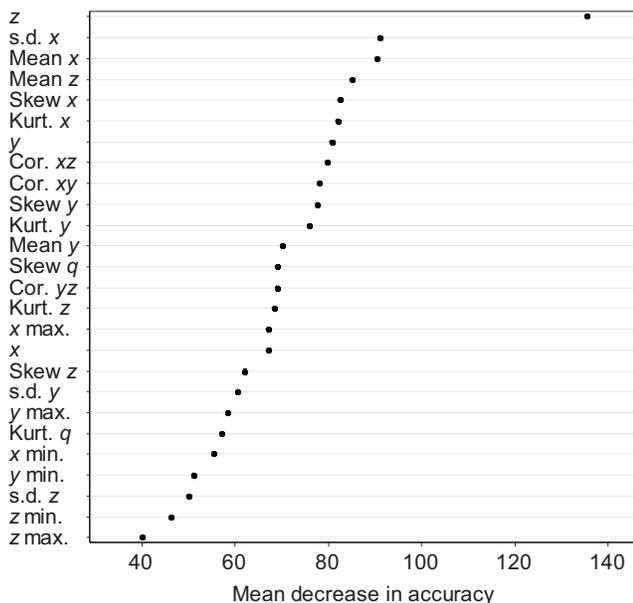

Fig. 5. Variable importance plot from our selected RF model. The importance plot provides a relative ranking of all 26 predictor variables that were used to fit our selected model, with larger values indicating variables that contributed more to the overall accuracy, or predictive ability, of the model. The mean decrease in accuracy for a variable is the normalised difference of the classification accuracy for the out-of-bag data for a given variable, and the classification accuracy for the out-of-bag data when the values of the given variable have been randomly permuted. Kurt., kurtosis; Cor., correlation.

Post hoc comparisons using Tukey's test indicated that the mean ODBA for each of our highly active behaviours was significantly greater than all behaviours except 'collar discomfort' (Table S5). When sorted by mean ODBA, all 14 behaviours grouped into their pre-assigned activity level, displaying a positive relationship between ODBA and animal activity (Fig. 6).

DISCUSSION

Accurate classification across a range of behaviours has been a great challenge for the majority of accelerometry studies. Our study is the first to use accelerometry to accurately classify a broad range of behaviours for an apex predator, at the very low sampling frequency of 1 Hz. Employing a comprehensive, yet strategic, approach to fitting and selecting our best model allowed us to address a number of challenges associated with translating raw acceleration data into a meaningful and biologically relevant format.

As expected, the choice of predictor variables influenced classification accuracy. Several other studies provide evidence for the importance of predictor variable selection. For example, Alvarenga et al. (2016) achieved an overall model accuracy of ~85% (across five behaviours) when using 44 derived predictor variables, whereas Martiskainen et al. (2009) used 28 relatively simple predictor variables and produced a mean accuracy across eight behaviours of ~94%. Although there was only a minor change in overall model accuracy between the larger predictor variable sets, the model's predictive ability for individual behaviours was influenced considerably by choice of predictor variables. Changes in acceleration across each axis depended largely on the type of movement and, therefore, a single axis (or predictor variable) may better capture the acceleration signature of one behaviour over another. Graf et al. (2015) reported the heave (up and down) axis to be particularly important for classifying different behaviours in Eurasian beavers (*Castor fiber*), while Alvarenga et al. (2016) found that, in sheep (*Ovis aries*), the surge (back and forth) axis contributed most to the model. Given that some predictor variables will be better suited to assist the classification model in distinguishing certain behaviours, *a priori* selection of predictor variables should be used, where decisions are driven by the behaviours of interest.

Acceleration during movement can change over very short time periods and is therefore commonly measured at infrasecond frequencies (between 8 and 100 Hz). Measuring acceleration at high frequencies increases computational effort for fitting classification models, limits the deployment period of accelerometers due to memory constraints, and may be unnecessary for behavioural classification in many instances. We are the first to show that a large number of distinct behaviours (14) can be classified using tri-axial accelerometer data derived from an unconventionally low sampling frequency of 1 Hz. There are few instances in the literature where an equally low sampling frequency was used, which is surprising given that several studies report only minor decreases in classification accuracy when down-sampling, for example, from 64 to 8 Hz (Wang et al., 2015), or 25 to 10 Hz (Alvarenga et al., 2016). However, there is a lower limit to sampling frequency that can be used to classify certain behaviours. In our study, the classification accuracy for our most active behaviours was not high. Given that the more active behaviours, such as running and playing, are swift movements performed over short time periods, our sampling frequency may not have allowed enough of the acceleration signature to be captured in order to adequately train the model. This issue is highlighted in small species, owing to their tendency for rapid movements of short duration (Hammond et al., 2016). Hammond et al. (2016) attached accelerometers to chipmunks (body mass c. 50 g) and found the lowest sampling frequency that
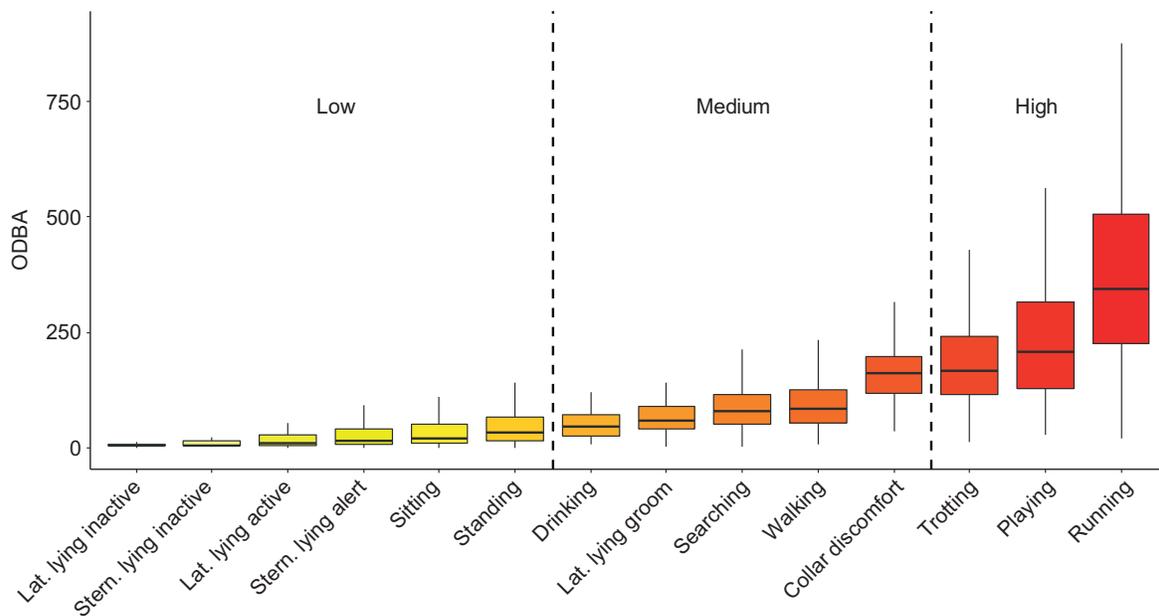
Fig. 6. Overall dynamic body acceleration (ODBA) values for each behaviour, calculated from a running mean of 4 s and displayed in ascending order of mean ODBA. Lower and upper hinges correspond to the first and third quartiles, respectively, and the whiskers extend from the hinge to no more than 1.5× IQR. Outliers were removed *post hoc* to better display the differences in the means. 'Low', 'Medium' and 'High' labels refer to the activity class assigned to the behaviours. Lat., lateral; Stern., sternum.

resulted in negligible decreases in model accuracy to be 20 Hz. Our study provides evidence that a very low sampling frequency can be used to classify a range of behaviours with high accuracy, in a medium-sized animal.

In our study, the expectation that functionally similar behaviours would most often misclassify as each other was only realised for the highly active behaviours. This is best explained by high intra-behaviour variation and inter-behaviour overlap within the axes. If classifying high-intensity behaviours with high accuracy is crucial, it may be necessary to increase the resolution of the acceleration signature by using a higher sampling frequency, but at a cost to deployment time. Our selected model performed extremely well (low misclassification rate) at classifying low-intensity functionally similar behaviours such as different resting postures. In studies where misclassification is particularly undesirable, it is common to group behaviours into broader classes, such as 'active' and 'inactive', which has the benefit of increasing classification accuracy but at the cost of behavioural diversity (e.g. Shamoun-Baranes et al., 2012). We chose a model that would identify a range of highly active behaviours, despite relatively low classification accuracy. We then managed our misclassification errors (model sensitivity) by choosing a threshold that would balance the number of unclassified and misclassified samples whilst retaining high overall model accuracy. If we were to apply our selected model to accelerometer data from free-ranging dingoes, we propose increasing the threshold from 0.3 to 0.5 for three reasons. Firstly, at 0.5 our overall model accuracy remains high at over 80%; secondly, our data will not be swamped by errors of omission; and, lastly, we are not overly concerned with a minor increase in misclassifications because the majority of behaviours will be misclassified to a functionally similar movement with comparable ODBA values. Overall, our methodological approach and resulting classification model is robust, and can readily be adapted to answer questions about different study systems.

Since Wilson et al. (2006) first presented evidence to suggest a positive correlation between ODBA and activity in cormorants, the use of accelerometry as a tool to remotely measure physiological traits such as energy expenditure and energy–time budgets has exploded. We also found a positive relationship between animal activity level and ODBA. Highly active behaviours exhibited significantly higher mean ODBA values than all low and medium behaviours. However, our results suggest that some caution should be taken when using ODBA as a proxy for energy expenditure. One of our highest ODBA scores came from the behaviour 'collar discomfort', which was typified by low overall body movement but acute movement of the accelerometer device due to quick side-to-side actions of the head. The result was ODBA values indistinguishable from 'trotting', an energetically demanding and ecologically important behaviour (Reilly et al., 2007). Energy and time budgets are of paramount importance for our understanding of how animals interact with their environment, especially for apex predators given their critical role in maintaining the structure of ecological communities (Fretwell, 1987). A recent study by Wang et al. (2015) used accelerometry to understand how an apex predator modulated their energy budget by examining foraging strategies (akin to 'searching' in our study), and in doing so they highlighted the potential benefits for conservation initiatives and human–wildlife conflict resolution. We extend the potential of future research by showing that classifying ecologically relevant behaviours whilst maintaining their aforementioned relationship with ODBA is possible even with acceleration data sampled at a very low frequency. The implication is that we can deploy accelerometers over much longer time periods to capture invaluable behavioural and physiological data across different life history stages of free-ranging animals.

Accelerometry is an exciting tool that is transforming the study of animal behaviour and physiology. The use of accelerometers to remotely classify behaviours of free-ranging animals has appreciable potential. However, prevailing methods limit our ability to establish meaningful ecological conclusions due to the challenge of classifying a diversity of behaviours over a significant

period of time. Our approach addresses these constraints and has applicability to free-ranging terrestrial quadrupeds of comparable size. We propose that our approach using the RF model can be directly applied to accelerometer data from other members of the family Canidae, given their shared body type and consistent style of locomotion (Flynn et al., 1988). Canids are a diverse lineage whose members are ecologically and economically important the world over; for example, red wolves (*Canis rufus*) are threatened with extinction (Kelly et al., 2008), gray wolves (*Canis lupus*) are keystone predators (Estes et al., 2011), and red foxes (*Vulpes vulpes*) are invasive pests that cause millions of dollars of damage each year in Australia alone (McLeod, 2004). Through building a classification model that exhibits high predictive performance at low frequency and across a large number of ecologically relevant behaviours, we increase the accessibility of accelerometer-based behavioural research and support much-needed integration with the fields of animal physiology and movement ecology.

Author contributions
Conceptualization: J.T., P.C.; Methodology: J.T., T.A.A.P.; Formal analysis: J.T., P.C., T.A.A.P.; Resources: J.T., P.C.; Writing - original draft: J.T.; Writing - review & editing: J.T., P.C., T.A.A.P.; Supervision: P.C., T.A.A.P.

Supplementary information
Supplementary information available online at
http://jeb.biologists.org/lookup/doi/10.1242/jeb.184085.supplemental

References
Allouche, O., Tsoar, A. and Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* **43**, 1223-1232.

Alvarenga, F. A. P., Borges, I., Palkovič, L., Rodina, J., Oddy, V. H. and Dobos, R. C. (2016). Using a three-axis accelerometer to identify and classify sheep behaviour at pasture. *Appl. Anim. Behav. Sci.* **181**, 91-99.

Breiman, L. (2001). Random Forests. *Mach. Learn.* **45**, 5-32.

Campbell, H. A., Gao, 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. *J. Exp. Biol.* **216**, 4501-4506.

Coomans, D. and Massart, D. L. (1982). Alternative k-nearest neighbour rules in supervised pattern recognition: Part 1. k-Nearest neighbour classification by using alternative voting rules. *Anal. Chim. Acta* **136**, 15-27.

Estes, J. A., Terbogth, J., Brashares, J. S., Power, M. E., Berger, J., Bond, W. J., Carpenter, S. R., Essington, T. E., Holt, R. D., Jackson, J. B. C. et al. (2011). Trophic downgrading of planet earth. *Science* **333**, 301-306.

Flynn, J. J., Neff, N. A. andTedford, R. H. (1988). *Phylogeny of the Carnivora*. Oxford: Oxford University Press.

Fretwell, S. D. (1987). Food chain dynamics: the central theory of ecology? *Oikos* **50**, 291-301.

Gómez Laich, A., Wilson, R. P., Gleiss, A. C., Shepard, E. L. C. and Quintana, F. (2011). Use of overall dynamic body acceleration for estimating energy expenditure in cormorants. *J. Exp. Mar. Biol. Ecol.* **399**, 151-155.

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**, e0136751.

Halsey, L. G. and White, C. R. (2010). Measuring energetics and behaviour using accelerometry in cane toads Bufo marinus. *PLoS ONE* **5**, e10170.

Halsey, L. G., Shepard, E. L. C., Quintana, F., GOMEZ Laich, A., Green, J. A. and Wilson, R. P. (2009). The relationship between oxygen consumption and body acceleration in a range of species. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **152**, 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. *J. Exp. Biol.* **219**, 1618-1624.

Ishii, M., Murase, H., Fukuda, Y., Sawada, K., Sasakura, T., Tamura, T., Bando, T., Matsuoka, K., Shinohara, A., Nakatsuka, S. et al. (2017). Diving behavior of Sei whales balaenoptera borealis relative to the vertical distribution of their potential prey. *Mamm. Study* **42**, 1-9.

Kelly, B. T., Beyer, A. and Phillips, M. K. (2008). *Canis rufus* [Online]. Available: http://www.iucnredlist.org/details/3747/0 [Accessed 09/02/2018].

Ladds, M., 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. *Anim. Biotelemetry* **5**, 8.

Liaw, A. and Wiener, M. (2002). Classification and Regression by randomForest. *R News* **2**, 18-22.

Martiskainen, P., Järvinen, M., Skön, J.-P., Tiirikainen, J., Kolehmainen, M. and Mononen, J. (2009). Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines. *Appl. Anim. Behav. Sci.* **119**, 32-38.

Mcclune, D. W., Marks, N. J., Wilson, R. P., Houghton, J. D. R., Montgomery, W. I., Mcgowan, N. E., Gormley, E. and Scantlebury, D. M. (2014). Tri-axial accelerometers quantify behaviour in the Eurasian badger (Meles meles): towards automated interpretation of field data. *Anim. Biotelemetry* **2**, 5.

Mcleod, R. (2004). *Counting the Cost: Impact of Invasive Animals in Australia 2004*. Cooperative Research Centre for Pest Animal Control. Canberra.

Pagano, A. M., Rode, K. D., Cutting, A., Owen, M. A., Jensen, S., Ware, J. V., Robbins, C. T., Durner, G. M., Atwood, T. C., Obbard, M. E. et al. (2017). Using tri-axial accelerometers to identify wild polar bear behaviors. *Endangered Species Res.* **32**, 19-33.

Reilly, S. M., Mcelroy, E. J. and Biknevicius, A. R. (2007). Posture, gait and the ecological relevance of locomotor costs and energy-saving mechanisms in tetrapods. *Zoology* **110**, 271-289.

Shamoun-Baranes, J., Bom, R., van Loon, E. E., Ens, B. J., Oosterbeek, K. and Bouten, W. (2012). From sensor data to animal behaviour: an oystercatcher example. *PLoS ONE* **7**, e37997.

Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE Trans. Neural Netw.* **10**, 988-999.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. New York, Springer.

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. *Mov. Ecol.* **3**, 2.

Williams, T. M., Wolfe, L., Davis, T., Kendall, T., Richter, B., Wang, Y., Bryce, C. M., Elkhaim, G. H. and Wilmers, C. C. (2014). Instantaneous energetics of puma kills reveal advantage of felid sneak attacks. *Science* **346**, 81-85.

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

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. *J. Anim. Ecol.* **75**, 1081-1090.

Yoda, K., Sato, K., Niizuma, Y., Kurita, M., Bost, C., LE Maho, Y. and Naito, Y. (1999). Precise monitoring of porpoising behaviour of Adelie penguins determined using acceleration data loggers. *J. Exp. Biol.* **202**, 3121-3126.