Ecological Informatics 60 (2020) 101152

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Accounting for unknown behaviors of free-living animals in accelerometer-based classification models: Demonstration on a wide-ranging mesopredator

Thomas W. Glass[a,b,*], Greg A. Breed[b,c], Martin D. Robards[a], Cory T. Williams[b,c], Knut Kielland[b,c]

[a] Wildlife Conservation Society, 3550 Airport Way, Suite 5, Fairbanks, AK 99709, USA
[b] Department of Biology and Wildlife, University of Alaska Fairbanks, PO Box 756100, Fairbanks, AK 99775, USA
[c] Institute of Arctic Biology, University of Alaska Fairbanks, PO Box 757000, Fairbanks, AK 99775, USA

ARTICLE INFO

Keywords:
Behavioral analysis
Accelerometer
Machine learning
Gulo gulo
Wolverine

ABSTRACT

Describing the behaviors of free-living animals is broadly useful for ecological and physiological research, but obtaining accurate records for difficult-to-observe species presents a considerable challenge. Tri-axial accelerometers are increasingly used for this purpose by exploiting behavioral observations from accelerometer-carrying animals to predict behaviors of unobserved conspecifics. We developed a modeling approach to predict behaviors of wolverines from collar-mounted accelerometers using Support Vector Machines. By applying a temporal smoothing function and setting a lower threshold for a-posteriori prediction probabilities, we improve the predictive performance of our model and simultaneously create a framework for explicitly accounting for behaviors unknown to the model, a problem that remains largely unaddressed in similar studies. We demonstrate that such an approach can achieve a model-averaged accuracy of 98.3%, with high predictive performance for the behaviors resting, running, scanning, tearing at food, and transferring items with the mouth, a behavior typically associated with caching food among captive wolverines. To illustrate the utility of this approach, we apply this model to a sample of seven free-living wolverines in Arctic Alaska.

1. Introduction

Describing the behaviors of free-living animals can provide important insights regarding a wide range of ecological processes. Taken alone, analysis of such behavioral records can be used to investigate temporal patterns in activity, including association among behaviors, yielding insights regarding circadian rhythms in the daily partitioning of behaviors or inter-individual differences in such temporal patterns (Garthe et al., 2003; Yoda and Ropert-coudert, 2007). When coupled with environmental and physiological information, behavioral analyses can address how such extrinsic and intrinsic factors influence behavioral decisions made by animals, including tradeoffs such as allocating time between foraging and antipredator behavior (Hamel and Côté, 2008; Studd et al., 2019; Switalski, 2003).

However, since documenting behavior has traditionally relied on direct observation, it is often a difficult or impossible task to assemble comprehensive records for remotely tracked free-living animals that have not been directly observed. Species that occupy areas that are remote or logistically difficult for human observers to access, such as under water, under snow, or in trees, present obvious challenges, as do species that range widely, travel quickly, or for which human observation alters behavior.

The rise of accelerometer-derived behavioral records promises to reduce these obstacles (Shepard et al., 2008). This process, whereby free-living animals are tagged with tri-axial accelerometers and the resulting data are used to predict the behaviors of the wearer, has been applied to a variety of marine (Battaile et al., 2015; Viviant et al., 2010; Whitney et al., 2010), and increasingly, terrestrial species (Hammond et al., 2016; McClune et al., 2014; Pagano et al., 2017; Wang et al., 2015). Resulting behavioral records have been used to investigate behaviors important to life history and fitness, including predation and mating events, and foraging strategies.

Using accelerometer data to classify behavior typically begins by building a predictive classification model based on observer-labeled accelerometer data. The labeled data are collected either by directly observing conspecific or surrogate species while wearing accelerometers (Campbell et al., 2013), or with the use of additional biologgers, such as video cameras, affixed to free-living individuals (Nakamura et al., 2015; Pagano et al., 2017; Watanabe and Takahashi, 2012). A classification model, such as a statistical learning classifier (Tatler et al., 2018) or decision tree analysis (Bellsolá, 2019) can then be applied to the labeled data to train and evaluate candidate models,

* Corresponding author at: Wildlife Conservation Society, 3550 Airport Way, Suite 5, Fairbanks, AK 99709, USA.
E-mail address: Trglass2@alaska.edu (T.W. Glass).

https://doi.org/10.1016/j.ecoinf.2020.101152
Received 24 April 2020; Received in revised form 3 August 2020; Accepted 13 August 2020
Available online 02 September 2020

after which the final model can be applied to free-living individuals where no direct observations or ancillary data for determining behaviors are available.

Here, we developed and evaluated the first predictive model that can be used to classify behaviors of free-living wolverines (*Gulo gulo*) using collar-mounted tri-axial accelerometers, based on visual observations of captive wolverines wearing similar collar-mounted accelerometers. Further, we used labeled accelerometer data from these captive conspecifics to create a framework by which behaviors not exhibited by the captive wolverines, and therefore unknown to the model, would be classified as "unknown," rather than incorrectly classified to the best fitting known acceleration pattern. By developing such a model, we hoped to broaden the field of possible questions that can be addressed regarding the interactions of the environment, physiology, and ecology of wolverines, and provide a framework that other researchers can employ for other species to address similar questions while explicitly addressing the problem of incorrect attribution for behaviors unknown to the model. Finally, to demonstrate the utility of our modeling approach, we applied several candidate models to a small sample of free-living wolverines and assessed temporal trends of resting, running, vigilance behavior, and behaviors associated with handling food.

2. Methods

A schematic outlining the workflow is included in Fig. 1.

2.1. Data collection

2.1.1. Captive wolverines

We collected accelerometer data and behavioral observations from three captive adult wolverines (two females and one male) at Nordens Ark, Hunnebostrand, Sweden, between March 4 and March 14, 2019. Wolverines were anaesthetized using a combination of ketamine, midazolam, and medetomidine. Collars were mounted with tri-axial accelerometers (AXY-3, 10 g, Technosmart Europe Srl., Colle Verde, Italy), GPS units (GIPSY 5, 100 g, Technosmart Europe Srl., Colle Verde, Italy), light/temperature loggers (Intigeo-C330, 3.3 g, Migrate Technology Ltd., Cambridge, United Kingdom) and timer-activated release mechanisms (TRD-L, 30 g, Lotek Wireless Inc., Newmarket, Canada), and weighed less than 3% of the animal's body mass. Accelerometers recorded at a frequency of 10 Hz. Collars were set to automatically release after approximately 10 days.

We conducted behavioral observations of these captive collared wolverines for 4–8 h per day from a platform and paths adjacent to their enclosures. We assembled an ethogram during the course of observations, creating new behavioral classes to accommodate behaviors as they were observed. We defined behaviors according to distinct motions and/or postures, which we expected to register differently in the accelerometer data (Appendix A). We opportunistically recorded behaviors exhibited by wolverines, along with associated timestamps from a watch that was synchronized with the accelerometers. Upon retrieving the accelerometers, we recorded the time each device was turned off, for use later in assessing temporal drift.

All animal handling and observation was consistent with Nordens Ark's own permitting and University of Alaska Fairbanks Institutional Animal Care and Use Committee (UAF IACUC) protocol #1373175.

2.1.2. Free-living wolverines

We collected accelerometer data from seven free-living adult wolverines (three females and four males) in the vicinity of Toolik Field Station, Alaska (68° 38′ N, 149° 36′ W) during spring and summer 2018. Captures took place between February 25 and April 18, and accelerometer data were collected between February 25 and July 27. Wolverines were captured using baited lumber box traps (modified from Lofroth et al., 2008), and anaesthetized using Telazol™ (175 mg,

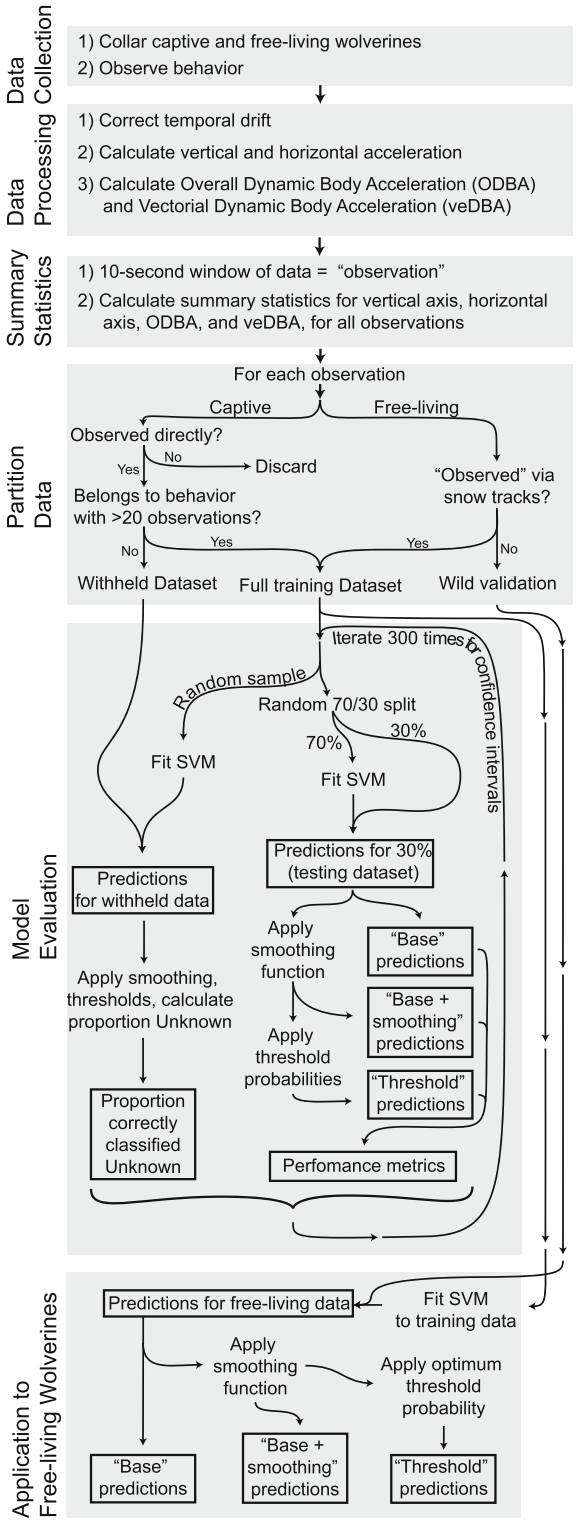

Fig. 1. Methodological workflow of the study. A combination of accelerometer data from captive and free-living wolverines was used to train, evaluate, and validate the use of Support Vector Machines (SVM) for behavior classification. The "full training" dataset refers to all accelerometer data for which we observed the animal's behavior, and which belongs to a behavior with > 20 observations. This dataset was randomly split into two (70% and 30% for training and testing, respectively) to evaluate the performance of the model, and this process was iterated 300 times to generate confidence intervals. The smoothing function reclassifies SVM predictions for each observation based on the nearest temporal neighbors, and the threshold probabilities reclassify observations as Unknown if the probability associated with the SVM prediction falls below the designated threshold.

Table 1
Summary statistics calculated for each 10 s partition of accelerometer behavior, used as predictor variables in Support Vector Machine classification models.

Summary statistic	Label/Predictor	Description
Mean	*meanH, meanV, meanODBA, meanQ*	Mean of the horizontal acceleration, vertical acceleration, ODBA, and veDBA
Max	*maxH, maxV, maxODBA, maxQ*	Maximum of the horizontal acceleration, vertical acceleration, ODBA, and veDBA
Standard deviation	*sdH, sdV, sdODBA, sdQ*	Standard deviation of the horizontal acceleration, vertical acceleration, ODBA, and veDBA
Kurtosis	*kurtH, kurtV, kurtODBA, kurtQ*	Kurtosis of the horizontal acceleration, vertical acceleration, ODBA, and veDBA
Skewness	*skewH, skewV, skewODBA, skewQ*	Skewness of the horizontal acceleration, vertical acceleration, ODBA, and veDBA
Dominant power spectrum	*dpsH, dpsV*	Maximum power spectral density of the horizontal and vertical acceleration
Frequency at the dominant power spectrum	*freqH, freqV*	Frequency at the maximum power spectral density of the horizontal and vertical acceleration

Golden et al., 2002). We used Lotek Iridium Litetrack 250 collars (~250 g, Lotek Wireless, Newmarket, Canada), to which we attached tri-axial accelerometers (AXY-3, 10 g, Technosmart Europe Srl., Colle Verde, Italy), using a combination of epoxy and steel cable ties. Accelerometers recorded at a frequency of 10 Hz. Collars weighed less than 3% of the animal's body mass, and were equipped with both mechanical release mechanisms and rot-away strips to ensure release from the animal. All capture and handling of free-living wolverines was conducted under UAF IACUC protocol #847738, and Alaska Department of Fish and Game scientific permit 18-085.

We opportunistically collected a single instance of labeled accelerometer data from a free-living collared wolverine. To do so, we followed fresh tracks in the snow at a site used < 24 h prior by the collared wolverine. We followed the tracks for approximately 1 km, during which time the individual maintained an unfaltering 3 × 3 lope characteristic of a running wolverine. Upon retrieving the accelerometer, we examined the data associated with this time period and extracted the portions having high-amplitude periodic motion, which we labeled as running. We included these data in our labeled full training dataset.

2.2. Data processing

Prior to processing data collected from captive wolverines, we first corrected for temporal drift in accelerometers by comparing the "power off" time recorded by the accelerometer with that displayed by the watch (which were synchronized at accelerometer deployment). Drift rates ranged between 0.3 and 1.5 s per day, depending on accelerometer, during the 10 days deployed. To account for this, we assumed a constant rate of drift and applied an accelerometer-specific linear correction to the timestamps associated with observations. In addition, we subtracted 1 s from the end of every behavioral observation to account for recorder error, reflecting the difference between the actual end of the activity and the moment the observer looked at their watch.

Since the accelerometers we deployed on free-living wolverines were not oriented in the same direction with respect to the animal's body, and we suspected that collars rotated around the animal's neck during deployment, we were unable to confidently delineate the surge, sway, and heave axes traditionally used in tri-axial accelerometer analysis for all individuals. To mitigate this problem, we converted the tri-axial measurements taken by the accelerometers into a vertical and horizontal component based on Mizell (2003). This required first estimating the magnitude of gravity along each axis, $g = (g_x, g_y, g_z)$, by applying a running mean over a 2 s window to the raw accelerometer data (the result is referred to as the static acceleration). We then subtracted this from the raw acceleration to estimate dynamic acceleration $d = (a_x - g_x, a_y - g_y, a_z - g_z)$ where (a_x, a_y, a_z) is the vector representing the raw acceleration data for any given time. We then computed the projection v of d on the vertical axis g using vector dot products, as

$$v = \left(\frac{d \bullet g}{g \bullet g}\right) g$$

This computation yields the vector $v = (v_x, v_y, v_z)$, which represents the vertical component of dynamic acceleration along each axis of the accelerometer. The horizontal component can then be calculated for each axis using the Pythagorean theorem, wherein

$$h = \sqrt{d^2 - v^2}$$

resulting in a horizontal value for each axis, representing the directionless magnitude of acceleration in the horizontal plane. We summed v_x, v_y, and v_z to find the total acceleraton in the vertical direction which we term "vertical acceleration" (analogous to the heave axis), and we summed h_x, h_y, and h_z as a representation of the total acceleration in the horizontal plane, which we term "horizontal acceleration" (analogous to the sum of the absolute values of the sway and surge axes). In addition, we calculated the overall dynamic body acceleration (ODBA, Gleiss et al., 2011) by summing the absolute dynamic acceleration values across the all three axes, and the vectorial dynamic body acceleration (veDBA, Gleiss et al., 2011) as

$$veDBA = \sqrt{d_x^2 + d_y^2 + d_z^2}$$

2.3. Summary statistic calculation

To generate predictor variables for behavioral classification, we partitioned the vertical and horizontal acceleration data, ODBA, and veDBA into 10 s segments, each segment termed an "observation," and calculated summary statistics for each (complete list in Table 1, distributions of summary statistics in Appendix A). We calculated dominant power spectrum (DPS) and frequency at DPS using a Fast Fourier Transform (Brigham and Morrow, 1967). We discarded any observations less than 10 s in duration, only retained behavioral classes with at least 20 observations, and termed the resulting dataset the "full training" dataset. We excluded 13 observations for which the horizontal or vertical acceleration was zero for the duration of the observation, since kurtosis and skewness could not be calculated. We assumed, and verified, that all such observations belonged to the behavioral class Rest, and employed this assumption in making predictions for the data of free-living wolverines (see Application to free-living wolverines). This resulted in nine behavioral classes ultimately included in our analysis (Table 2). All observations belonging to classes with fewer than 20 observations were termed the "withheld" dataset and were used later in evaluating performance (see Modeling).

2.4. Modeling

We used the machine learning technique Support Vector Machines (SVM) to classify behaviors from accelerometer data, implemented in R package *e1071* (Meyer et al., 2018; R Core Team, 2018). Our choice of SVM reflects this method's high predictive performance in behavior-recognition tasks (Campbell et al., 2013; Grünewälder et al., 2012; Tatler et al., 2018), and our desire to employ a probabilistic modeling framework, since probabilities associated with predictions are integral to our evaluation of unknown behaviors. SVM assign data to user-defined classes by constructing a hyperplane between binary classes. The number of observations allowed to violate the hyperplane is controlled by a user-defined cost parameter, and a margin surrounding the hyperplane is maximized. The hyperplane is chosen as that which allows

Table 2
Description, number of 10 s observations, and number of individuals represented in the full training dataset used in classification model. Observations from all three captive wolverines were included for all behaviors, to which we added observations of the behavior "Run" from a single free-living wolverine.

Intensity	Behavior	Description	*N* observations	*N* Ind (m,f)
Low	Rest	Motionless except breathing. Excludes motion during rest, *e.g.* rolling over.	299	3 (1,2)
	Scan	Survey surroundings by moving head while torso and legs remain stationary.	128	3 (1,2)
Medium	Walk	Slow, sometimes meandering, directional movement.	63	3 (1,2)
	Groom	Lick and lightly chew on feet, stomach, and groin.	141	3 (1,2)
	Eat	Chew item in mouth with head raised.	66	3 (1,2)
	Gnaw	Chew on bone, piece of wood, or frozen food by holding it in front (and sometimes hind) paws, usually using one side of the mouth only.	40	3 (1,2)
	Transfer	Pick up items such as sticks and leaves from the ground and move them with a rapid sway of the head to the side. In captivity this behavior was always associated with covering food items.	121	3 (1,2)
	Tear	Remove pieces of meat from carcass by pulling with teeth, can include short bursts of gnawing and eating.	443	3 (1,2)
High	Run	Rapid directional movement.	158	4 (1,3)

the largest separation between classes, *i.e.* the widest margin surrounding the hyperplane. Hyperplanes are defined by the observations that either fall within the separating margin, or that violate the separating hyperplane, and these observations are termed "support vectors." Hyperplanes can take nonlinear forms by applying a kernel function to the inner product of the support vectors (Aizerman et al., 1964). We chose to use a radial kernel for maximal flexibility in hyperplane definition. To generalize this binary classifier to a multiclass response, we used a "one-*versus*-one" approach, in which observations are classified for every possible pair of classes and the class most commonly selected is the predicted value. A-posteriori class probabilities can be computed by fitting a logistic distribution to the decision values of all binary classifiers and extracting class probabilities using quadratic optimization. In addition to the cost parameter, support vector machines with a radial kernel can be tuned using a gamma parameter, which weights support vectors in the definition of the hyperplane. We used 5-fold cross validation on the full training dataset and calculated accuracy (see Model evaluation) as an indication of model performance across all combinations of gamma = (0.001, 0.01, 0.1, 1, 2, 4) and cost = (0.1, 1, 10, 100, 250, 500, 750, 1000, 10,000), selecting the parameters that yielded the best performing model for all further analyses (parameters chosen per recommendations in Hsu et al., 2010). A more detailed, accessible description of SVMs can be found in James et al (2017).

Since our behavioral classes were unbalanced, we included class weights in the SVMs, calculated as

$$w_{ik} = \frac{N_k}{n_{ik}}$$

where w_{ik} is the weight of class i for the k^{th} iteration of cross validation, N_k is the total number of observations in the full training data for the k^{th} iteration of cross validation, and n_{ik} is the number of observations in class i for the k^{th} iteration of cross validation.

We were interested in building a modeling framework that explicitly incorporated unknown behaviors (*i.e.* behaviors within the ethogram of a free-living wolverine that we did not observe among those in captivity), such that these unknown behaviors would be identified as Unknown by the model. To do this, we set a threshold level for a-posteriori class probabilities associated with predicted behaviors, below which predictions were assigned to the category Unknown. To evaluate how well the model correctly classified such unknown behaviors, we fit a model to the full training dataset and made predictions for all observations in the "withheld" dataset, using these as proxies for behaviors that we didn't observe among wolverines in captivity. Since these withheld observations (observations that belonged to behavioral classes observed at low frequency among captive wolverines) were all known to belong to behaviors other than those in the full training dataset, the "perfect" model would categorize them all as Unknown, and this result is approached as the threshold probability value increases to one. However, increasing the threshold probability comes at the cost of incorrectly categorizing some known observations as Unknown, so we selected the optimum threshold probability as that which maximized both the proportion of withheld observations that were correctly classified as Unknown and the model's accuracy in classifying known observations correctly, referenced to the entire dataset (termed "full accuracy", see Model evaluation). We evaluated threshold probabilities ranging from 0 to 0.95 by increments of 0.05, and bootstrapped the procedure 300 times, resampling the full training dataset with replacement, to generate confidence intervals.

We were also interested in examining the effect of applying a smoothing function to predicted behavioral classes (Cao et al., 2012; Chimienti et al., 2016; Grünewälder et al., 2012). We therefore made predictions for the 10 s observations immediately following and preceding the observation of interest, and the predicted behavior of the observation of interest became the class that occurred most commonly within this 30 s window. If the three predicted classes within the window were all different, the observation of interest retained its original class. We performed this in tandem with applying threshold probabilities, such that predictions were subject to reclassification as Unknown according to the threshold probability before being subject to the smoothing function.

We defined the "base model" as that which was subject to neither a smoothing function nor a threshold probability, the "base + smoothing" model as that which was subject to a smoothing function but no threshold probability, and the "threshold model" as either the smoothing model or the non-smoothing model that had the highest accuracy at the optimal threshold probability.

2.5. Model evaluation

To evaluate the performance of our model, we relied on metrics derived from tallies of True Positives (T+), True Negatives (T−), False Positives (F+), and False Negatives (F−). These groups are tallied for each behavioral class in a single model independently. Specifically, for a given behavior, observations whose predicted class and true class both match the given behavior are considered T+, observations whose predicted class and true class neither match the given behavior are T−, observations whose predicted class matches the given behavior but whose true class does not are F+, and observations whose true class matches the given behavior but whose predicted class does not are F−. These terms are straightforwardly depicted using a confusion matrix (Fig. 2). We used 300 bootstrapped samples, with a new random 70/30 draw of training/testing data drawn from the full training dataset, stratified by behavior, for each iteration, to generate tallies of T+, T−, F+, and F−, and to calculate associated confidence intervals. The training/testing split mentioned here, used only during this bootstrapping process, should not be confused with the full training dataset defined above. The bootstrap method allows explicit estimation of confidence intervals for performance metrics, unlike the more traditional *k*-fold cross validation (Champagne et al., 2014).

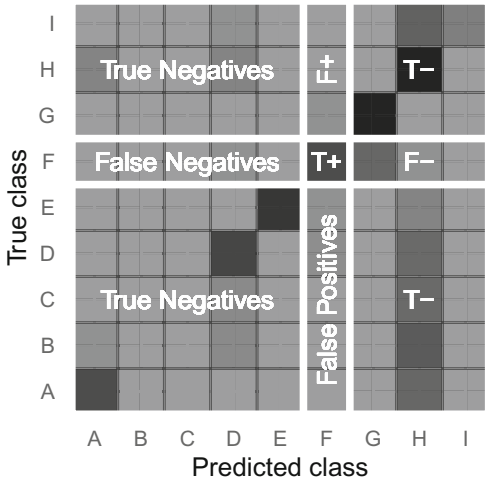

Fig. 2. Example confusion matrix for behavioral classes A-F depicting the calculation of True Positive (T+), True Negative (T−), False Positive (F+), and False Negative (F−) values. Each metric is tallied for each class independently; the figure depicts these values for behavioral class F. This figure does not represent actual data, and is included only to demonstrate how performance metrics are calculated.

Once these values are calculated for each behavior, a variety of metrics can be derived to evaluate the performance of the model. We chose to calculate accuracy, precision, and recall (sometimes called sensitivity), as these are relatively common among accelerometer-based behavioral classifiers and enable comparison across studies. Accuracy is defined as the proportion of the observations that are classified correctly, calculated as:

$$\frac{(T+) + (T-)}{(T+) + (T-) + (F+) + (F-)}$$

When calculating accuracy for the purpose of selecting a threshold probability (see Modeling), we replaced the denominator with the total number of observations (*i.e.* including those which fell below the threshold probability and were classified as Unknown), thereby providing a metric that assessed the cost of increasing the threshold probability, and termed the result "full accuracy."

Precision is defined as the proportion of predicted observations for a given behavior that actually belong to that behavior, calculated as:

$$\frac{(T+)}{(T+) + (F+)}$$

Recall is defined as the proportion of true observations for a given behavior that were predicted to belong to that behavior, calculated as:

$$\frac{(T+)}{(T+) + (F-)}$$

To assess overall model performance, each metric can be either micro-averaged, whereby the values for T+, T-, F+, and F- are summed across behaviors and the overall metric is calculated on the summed values, or macro-averaged, whereby the metric is calculated for each behavior independently and subsequently averaged (Sokolova and Lapalme, 2009). Since micro-averaging favors larger groups in an unbalanced model, we macro-averaged accuracy to reduce such bias. Since precision and recall can be undefined for a given behavior, we did not average these values and instead evaluated them at the level of individual behaviors only.

We were also interested in evaluating the relative importance of each variable in the predictive performance of the model. To do this, we iteratively left one variable at a time out of the model and calculated the decrease in accuracy from the full model. We found that model predictive accuracy was reduced by < 0.1% regardless of the variable dropped, suggesting high redundancy across variables.

2.6. Application to free-living wolverines

To illustrate the efficacy of this model, we made predictions of the behaviors exhibited by the seven free-living wolverines. To do this, we processed the data of free-living wolverines and extracted summary statistics as described above, and used all three candidate models, fit to the full training dataset (see Summary statistic calculation), to make predictions for the free-living observations. Since kurtosis and skewness for observations with horizontal or vertical acceleration values of zero could not be calculated, we did not use the models to make predictions for such observations, instead assigning them to Rest, a classification that was supported by the data of the captive wolverines (see Summary statistic calculation). We chose to retain only predictions associated with behaviors whose lower confidence level of precision was greater than 0.6 for all three models, since low precision values indicate a low probability that the prediction is correct. For each model's predictions, we calculated the proportion of each hour of the day each animal spent engaged in each behavior, averaged across a season. We defined the seasons spring and summer as between February 25–May 10 and May 11–July 27 respectively, roughly corresponding to the many ecological and physiographic changes that take place in the Arctic around May 10, including rapid snow ripening and melt (Macander et al., 2015), disappearance of river and lake ice (Arp et al., 2013), caribou and bird migration (Tape and Gustine, 2014), and grizzly bear and ground squirrel emergence (Buck and Barnes, 1999; McLoughlin et al., 2002). Since the purpose of this study was to develop a modeling framework for making such predictions, and not to make inference on the behaviors of free-living wolverines *per se*, we chose to simply visualize these predictions by plotting the mean proportions of time spent in each behavior by hour of day, averaged across individuals.

Since the purpose of this study was to develop and evaluate a modeling framework for making behavioral predictions of free-living animals, we also used the predictions made for free-living wolverines to compare broad-scale differences of predictions made by the three models.

3. Results

3.1. Model development and evaluation

Model tuning yielded optimum values of gamma = 0.01 and cost = 10. The base model had an overall predictive accuracy of 94.6% (95% CI: 93.2–95.8%) and correctly classified 326 (95% CI: 288–360) of the 433 observations in the 30% portion of the full training dataset used for testing (75.4%, 95% CI: 66.5–83.1%). Performance for individual behaviors ranged from a precision of 0 to 0.98, and recall ranged from 0 to 0.97 (Table 3). The "base + smoothing" model had an overall accuracy of 95.8% (95% CI: 94.4–96.8%), and correctly classified 349 (95% CI: 311–379) of the 433 observations in the testing dataset (80.6%, 95% CI: 71.8–87.5%). The optimum threshold probability for the non-smoothed model was 0.65, and for the smoothed model was 0.625 (Fig. 3). At the optimum threshold probability of each, the smoothed model and non-smoothed model correctly classified 55.4% (95% CI: 41.7–70.9%) and 51.2% (95% CI: 40.5–64.3%) of observations in the withheld dataset as Unknown, respectively. The full accuracy at the optimum threshold probability for the smoothed model was 57.5% (95% CI: 52.3–62.4%), which was higher than that of the non-smoothed model (53.4%, 95% CI: 49.0–57.7%). Therefore, the "threshold" model was selected as the smoothed model with a threshold probability of 0.625. The "threshold" model incorrectly classified 179 (95% CI: 157–202) observations as unknown, and had an overall accuracy of 98.3% (95% CI: 96.6–99.5%), correctly predicting 92.2% (95% CI: 88.5–95.4%) of the remaining observations (Fig. 4).

Table 3
Accuracy, precision, and recall for all behaviors for the base, base + smoothing, and threshold models. Bootstrapped 95% confidence intervals are shown in parentheses.

	Base model			Base + Smoothing			Threshold model		
Behavior	*Accuracy*	*Precision*	*Recall*	*Accuracy*	*Precision*	*Recall*	*Accuracy*	*Precision*	*Recall*
Tear	0.83 (0.80,0.86)	0.67 (0.63,0.71)	0.90 (0.84,0.95)	0.86 (0.83,0.88)	0.70 (0.66,0.74)	0.96 (0.91,0.99)	0.95 (0.92,0.98)	0.89 (0.81,0.95)	0.99 (0.96,1.00)
Groom	0.93 (0.91,0.94)	0.63 (0.52,0.76)	0.60 (0.45,0.74)	0.96 (0.94,0.97)	0.79 (0.66,0.90)	0.74 (0.58,0.86)	0.99 (0.97,1.00)	0.75 (0.00,1.00)	0.58 (0.00,1.00)
Transfer	0.96 (0.94,0.97)	0.76 (0.63,0.89)	0.69 (0.54,0.83)	0.98 (0.97,0.99)	0.90 (0.80,1.00)	0.86 (0.75,0.97)	0.99 (0.97,1.00)	0.94 (0.80,1.00)	0.90 (0.70,1.00)
Scan	0.95 (0.94,0.97)	0.78 (0.67,0.91)	0.65 (0.50,0.76)	0.97 (0.95,0.98)	0.88 (0.75,1.00)	0.71 (0.55,0.83)	0.98 (0.95,0.99)	0.94 (0.80,1.00)	0.70 (0.47,0.90)
Rest	0.96 (0.94,0.98)	0.86 (0.79,0.91)	0.97 (0.91,1.00)	0.97 (0.96,0.99)	0.88 (0.83,0.93)	0.99 (0.94,1.00)	0.98 (0.96,0.99)	0.93 (0.88,0.99)	1.00 (0.99,1.00)
Eat	0.95 (0.95,0.95)	0 (0.00,1.00)	0 (0.00,0.05)	0.95 (0.95,0.96)	1.00 (0.00,1.00)	0 (0.00,0.05)	0.98 (0.96,0.99)	–	0 (0.00,0.00)
Walk	0.97 (0.96,0.98)	0.90 (0.63,1.00)	0.47 (0.21,0.68)	0.97 (0.96,0.98)	0.90 (0.61,1.00)	0.42 (0.08,0.68)	1.00 (0.99,1.00)	1.00 (0.00,1.00)	0 (0.00,1.00)
Run	0.99 (0.98,1.00)	0.98 (0.91,1.00)	0.92 (0.84,0.98)	0.99 (0.97,1.00)	0.98 (0.93,1.00)	0.89 (0.79,0.96)	1.00 (0.99,1.00)	1.00 (0.97,1.00)	0.98 (0.92,1.00)
Gnaw	0.97 (0.97,0.97)	0 (0.00,0.00)	0 (0.00,0.00)	0.97 (0.97,0.97)	0 (0.00,0.00)	0 (0.00,0.00)	0.99 (0.98,1.00)	–	0 (0.00,0.00)

3.2. Application to free-living wolverines

At a gross level, the "base" and "base + smoothing" models yielded similar predictions for free-living wolverines, while the "threshold" model classified 32.4% of observations as Unknown (Fig. 5). Of the observations classified as Unknown by the "threshold" model, the "base" model classified 13.8% as Run, 4.3% as Scan, 8.5% as Rest, 12.0% as Transfer, and 50.8% as Tear, while the "base + smoothing" model classified 13.4% as Run, 3.7% as Scan, 8.3% as Rest, 10.4% as Transfer, and 55.4% as Tear. The remaining observations that were classified as Unknown by the "threshold" model were predicted by the other two models to belong to behavioral classes with low precision values. All three models predicted Rest as the most commonly occurring behavior at 47.3%, 48.1%, and 45.4% of all observations for the "base," "base + smoothing," and "threshold" models respectively. All three models revealed that the seven wolverines generally spent more time resting between approximately 14:00 and 00:00 local time during summer, and had peak resting times around 17:00 and 02:00 during spring (Fig. 5).

4. Discussion

Classifying behaviors from accelerometer data is an increasingly popular technique for addressing questions relating to the ecology and physiology of free-living animals. Considerable progress has been made in the field, particularly in evaluating the performance of different classification models (Nathan et al., 2012; Tatler et al., 2018) and the integration of multiple data sources, such as GPS and acoustic recorders, with acceleration to predict behavior (Shamoun-Baranes et al., 2012; Studd et al., 2019). Despite many such studies relying on predictive models built from direct observation of captive conspecifics, to our knowledge only one has explicitly addressed the problem of behaviors exhibited by free-living animals but not their captive counterparts (Rast et al., 2020), and that did so without a formal evaluation of efficacy.

Therefore, the purpose of this study was to create and evaluate a modeling framework that maximized predictive performance of behaviors from accelerometer data while simultaneously minimizing the incorrect classification of behaviors that are unknown to the model. Our

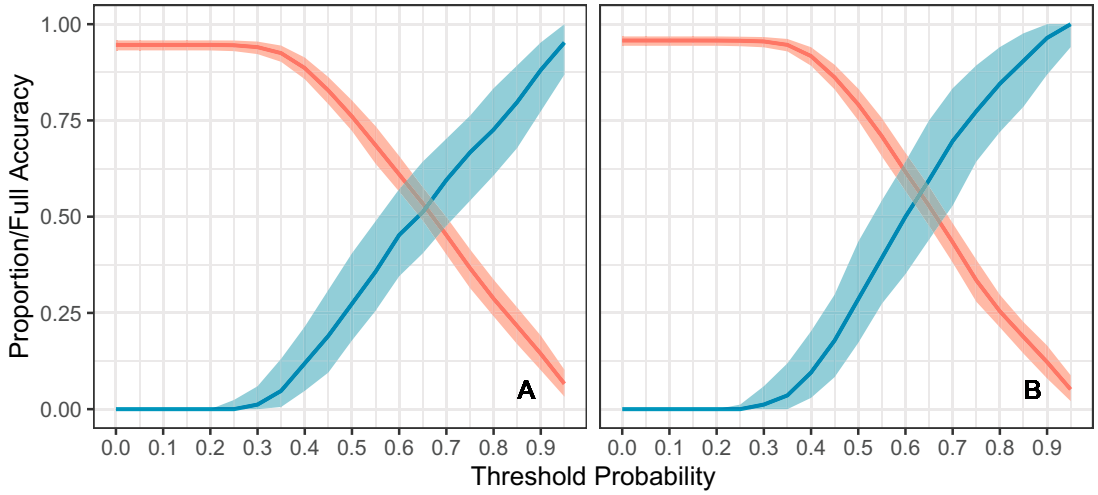

Fig. 3. Full accuracy (red) and proportion of withheld observations classified correctly as unknown (blue) across threshold probabilities, without (A) and with (B) a smoothing function. The optimum threshold probability was selected as that which maximized both values (0.625 using the smoothing function, 0.65 without the smoothing function). Confidence bands are 95% quantiles, calculated pointwise using 300 bootstrap iterations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

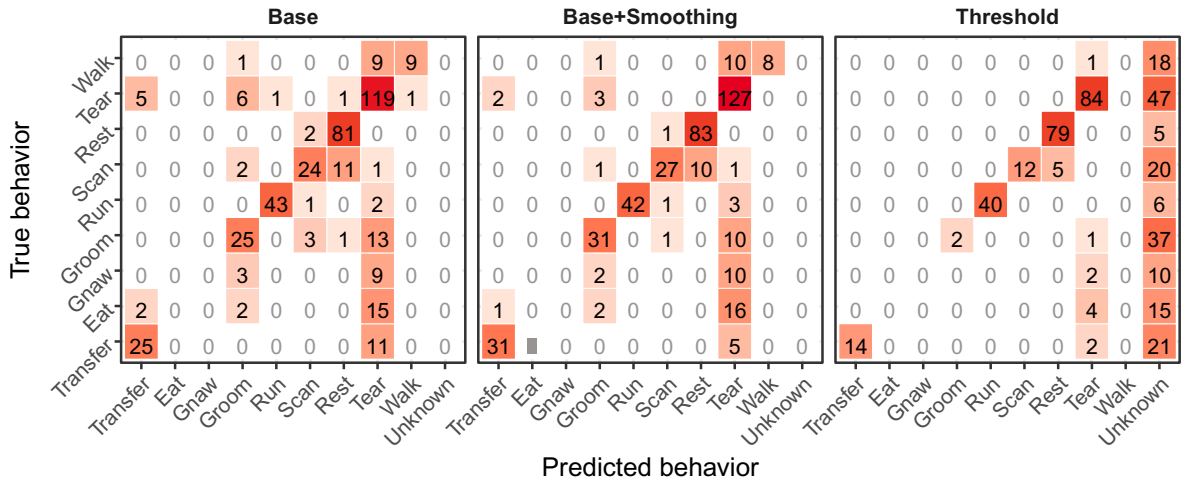

Fig. 4. Confusion matrices for the base model, base + smoothing model, and threshold model. Values are the median number of observations classified in each category across 300 bootstrap iterations. For the threshold model, the percent classified as correct in the confusion matrix differs from the median value reported in the results since the number of observations classified as Unknown varied across bootstrap iterations.

results indicate that this process can yield a high-performing model, with macro-averaged accuracy of over 98%, using a modeling framework that is conceptually straightforward and computationally efficient. Moreover, the explicit allowance for behaviors unknown to the model improves its generalizability to accelerometer data gathered on free-living animals, since it is likely that the range of behaviors exhibited by captive animals is different from those of wild animals.

Additionally, our results indicate that predictive performance can be improved by applying a temporal smoothing function to predictions (Table 3, Fig. 4), exploiting the apparently high degree of temporal correlation among behaviors exhibited by captive wolverines over a 30 s period. Previous studies have employed similar approaches (Cao et al., 2012; Chimienti et al., 2016; Grünewälder et al., 2012), although formal evaluation of the effect on model performance has been limited. A notable exception is (Cao et al., 2012), who evaluated the effect of applying such a smoothing function across a range of window lengths, and allowing such windows to vary by behavioral class. It is important to note that increasing the window length for both the smoothing function and the initial computation of summary statistics will reduce predictive performance for behaviors that typically occur at intervals shorter than the chosen window.

The practice of applying a threshold probability to behavioral predictions from accelerometer data has been employed before (Bellsolá, 2019; Bidder et al., 2014; Rast et al., 2020; Ware et al., 2015), including as an explicit means of reducing the incorrect classification of unknown behaviors (Rast et al., 2020). However, as noted above, previous studies have not formally evaluated the efficacy of this approach. By including observations from behaviors unknown to the model in our approach, we have developed a method of achieving such explicit evaluation. Our model selection process settled on 0.6 as the optimum threshold probability for maximizing model performance, but this value will undoubtedly vary by species, training dataset, and classification model used. Importantly, the model selection process developed here, whereby we evaluate the predictive performance of the model simultaneously for behaviors both known and unknown to the model across a range of threshold probabilities, can be used with any probabilistic classification model (or any classification model for which *a-priori* class probabilities can be calculated), not just SVM.

While the process of explicitly accounting for unknown behaviors improves the precision of the classifier, it does not yield a complete activity budget for a free-living animal and therefore excludes or introduces uncertainty into some biological questions, including those

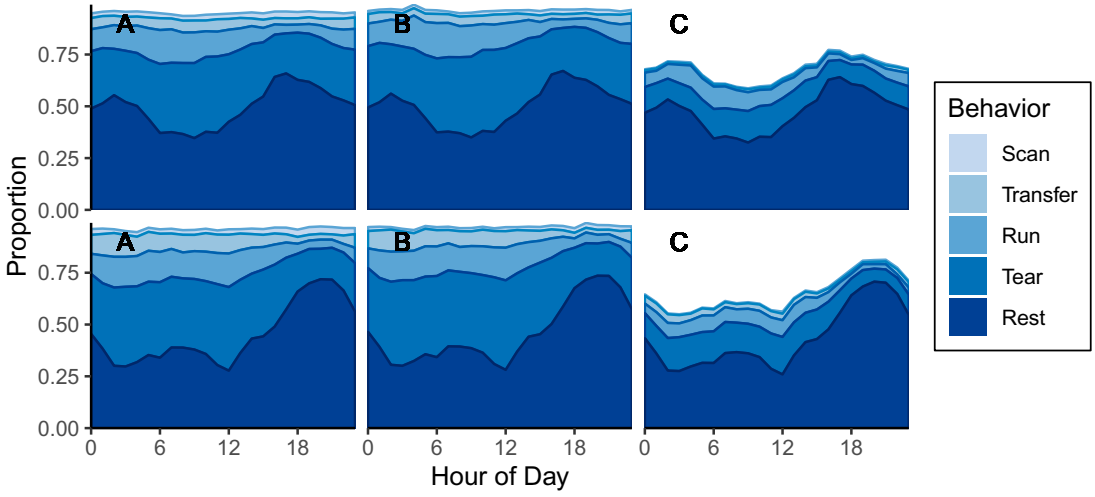

Fig. 5. Mean proportion of time per hour in each of five predicted behavioral classes for seven wolverines in Arctic Alaska during spring (top) and summer (bottom). Proportions are shown for the base model (A), base + smoothing model (B), and threshold model (C). Mean proportions were first averaged within individual across days, and then across individuals.

addressing temporal niche partitioning or rhythms of specific behaviors. We argue that this trade-off is necessary, since investigating such questions using a classifier that fails to account for such unknown behaviors would be based on the likely incorrect assumption that all behaviors are known to the model.

Although our "threshold" model had a macro-averaged accuracy of 98.3%, this metric risks overstating the performance of the model. Since model performance can be somewhat subjectively determined by whichever metric is most important for a given study, it is important to inspect the behavior-specific precision and recall values as well (Table 3). A high precision value indicates that most of the observations classified as a given behavior actually belong to that behavior, and a high recall value indicates that most of the observations that belonged to a given behavior were actually classified as that behavior. Since our interest was in making predictions for out-of-sample observations of free-living wolverines, model performance is best reflected by precision (Bidder et al., 2014). Our "threshold" model had very poor precision and recall for certain behaviors, but quite high values for others. Specifically, it failed to classify any observations, or classified only two observations, into four of the nine categories (Eat, Gnaw, Groom, and Walk), instead incorrectly classifying many of these observations as either Unknown or Tear. As a result, the model had lower precision for Tear, and low recall for Eat, Gnaw, Groom, and Walk. Each of these behaviors are medium activity (Table 2), with little to no periodicity, so this result is unsurprising. The behaviors most frequently misclassified as Tear were all associated with food handling, so for out-of-sample predictions this category could be considered a catch-all "food-handling" class. The relatively high precision values for Rest, Scan, Run, and Transfer suggest that the most reliable biological inference will be made about these behaviors.

In applying the model to data from free-living wolverines, we aimed to demonstrate its utility and provide a proof-of-concept that could be used by other researchers to evaluate the applicability of the model to their specific questions. The behavioral predictions we obtained show considerable variation by hour-of-day and season. Previous research of circadian rhythms in wolverines at high latitudes found a drop in activity around midday during spring and summer (Thiel et al., 2019), consistent with the generally higher frequency of resting that we observed during that period among the animals in our study, although the seven animals in our study also exhibited more resting around 02:00 during spring (Fig. 5). These results cannot be generalized without a larger sample.

Behavioral classification from accelerometer data is a promising area of active research, with the potential to greatly improve our understanding of the behavior of free-living animals. Coupled with other biologged data, such as location, body temperature, and heart-rate, accelerometry can shed light on the wide range of ecological and physiological processes governing behavioral decisions (Wilmers et al., 2015), including species' response to climate change (Chmura et al., 2018). However, developing accurate predictive models to extract behavior from accelerometer data remains challenging, and each model will be characterized by relative strengths and weaknesses. Formal evaluation of such characteristics, through consistently defined metrics such as accuracy, precision, and recall, is crucial to the advance of the field and comparison of different modeling approaches. Here, we have presented a novel high-performance modeling approach for classifying accelerometer data into discrete behaviors, which can be readily exploited by other researchers with accelerometer data from wolverines, or adapted to other species.

Funding

This work was supported by the M.J. Murdock Charitable Trust, Wilburforce Foundation, the University of Alaska Fairbanks (UAF) Erich Follmann Memorial Student Research Fund, a UAF College of Natural Science and Mathematics Travel Grant, Wildlife Conservation Society, 69 generous individuals *via* a crowdfunding campaign, and a National Science Foundation Graduate Research Fellowship under Grant No. 1650114.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgements

We thank the staff at Nordens Ark, in particular E. Nygren and E. Andersson, for providing crucial access to captive wolverines, logistical support for wolverine immobilizations, and accommodations for researchers. We thank the staff at Toolik Field Station, particularly J. Timm, as well as M. Kynoch, C. Haddad, and S. Andersen with Wildlife Conservation Society for logistical support in the field. We are grateful to J. McIntyre for feedback on early drafts of this manuscript and statistical advice.

References

R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Bellsolá, J.C. de A., 2019. Impala (*Aepyceros melampus*) Responses to Anthropogenic Activities - An Accelerometry Approach in the Serengeti Ecosystem. Norwegian Universty of Science and Technology.

Aizerman, M.A., Bravernam, E.M., Rozonoer, L.I., 1964. Theoretical foundations of the potential function method in pattern recognition learning. Avtom. I Telemekhaniki 25, 917–936.

Arp, C.D., Jones, B.M., Grosse, G., 2013. Recent lake ice-out phenology within and among lake districts of Alaska, USA. Limnol. Oceanogr. 58, 2013–2028. https://doi.org/10.4319/lo.2013.58.6.2013.

Battaile, B.C., Sakamoto, K.Q., Nordstrom, C.A., Rosen, D.A.S., Trites, A.W., 2015. Accelerometers identify new behaviors and show little difference in the activity budgets of lactating northern fur seals (*Callorhinus ursinus*) between breeding islands and foraging habitats in the eastern Bering Sea. PLoS One 10, 1–20. https://doi.org/10.1371/journal.pone.0118761.

Bidder, O.R., Campbell, H.A., Gómez-Laich, A., Urgé, P., Walker, J., Cai, Y., Gao, L., Quintana, F., Wilson, R.P., 2014. Love thy neighbour: automatic animal behavioural classification of acceleration data using the k-nearest neighbour algorithm. PLoS One 9, 1–7. https://doi.org/10.1371/journal.pone.0088609.

Brigham, E.O., Morrow, R.E., 1967. The fast fourier transform. IEEE Spectr. 4, 63–70. https://doi.org/10.1109/MSPEC.1967.5217220.

Buck, C.L., Barnes, B.M., 1999. Annual cycle of body composition and hibernation in free-living arctic ground squirrels. J. Mammal. 80, 430–442. https://doi.org/10.1016/j.jbankfin.2011.02.009.

Campbell, H.A., Gao, L., Bidder, O.R., Hunter, J., Franklin, C.E., 2013. Creating a behavioural classification module for acceleration data: using a captive surrogate for difficult to observe species. J. Exp. Biol. 4501–4506. https://doi.org/10.1242/jeb.089805.

Cao, H., Nguyen, M.N., Phua, C., Krishnaswamy, S., Li, X.L., 2012. An integrated framework for human activity recognition. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing, pp. 331–340. https://doi.org/10.1145/2370216.2370334.

Champagne, C., McNairn, H., Daneshfar, B., Shang, J., 2014. A bootstrap method for assessing classification accuracy and confidence for agricultural land use mapping in Canada. Int. J. Appl. Earth Obs. Geoinf. 29, 44–52. https://doi.org/10.1016/j.jag.2013.12.016.

Chimienti, M., Cornulier, T., Owen, E., Bolton, M., Davies, I.M., Travis, J.M.J., Scott, B.E., 2016. The use of an unsupervised learning approach for characterizing latent behaviors in accelerometer data. Ecol. Evol. 6, 727–741. https://doi.org/10.1002/ece3.1914.

Chmura, H.E., Glass, T.W., Williams, C.T., 2018. Biologging physiological and ecological responses to climatic variation: new tools for the climate change era. Front. Ecol. Evol. 6, 1–9. https://doi.org/10.3389/fevo.2018.00092.

Garthe, S., Benvenuti, S., Montevecchi, W.A., 2003. Temporal patterns of foraging activities of northern gannets, Morus bassanus, in the northwest Atlantic Ocean. Can. J. Zool. 81, 453–461. https://doi.org/10.1139/z03-027.

Gleiss, A.C., Wilson, R.P., Shepard, E.L.C., 2011. Making overall dynamic body acceleration work: on the theory of acceleration as a proxy for energy expenditure. Methods Ecol. Evol. 2, 23–33. https://doi.org/10.1111/j.2041-210X.2010.00057.x.

Golden, H.N., Shults, B.S., Kunkel, K.E., 2002. Immobilization of wolverines with Telazol® from a helicopter. Wildl. Soc. Bull. 30, 492–497.

Grünewälder, S., Broekhuis, F., Macdonald, D.W., Wilson, A.M., McNutt, J.W., Shawe-Taylor, J., Hailes, S., 2012. Movement activity based classification of animal behaviour with an application to data from cheetah (*Acinonyx jubatus*). PLoS One 7, 1–11. https://doi.org/10.1371/journal.pone.0049120.

Hamel, S., Côté, S.D., 2008. Trade-offs in activity budget in an alpine ungulate: contrasting lactating and nonlactating females. Anim. Behav. 75, 217–227. https://doi.

org/10.1016/j.anbehav.2007.04.028.
Hammond, T.T., Springthorpe, D., Walsh, R.E., Berg-Kirkpatrick, T., 2016. Using accelerometers to remotely and automatically characterize behavior in small animals. J. Exp. Biol. 219, 1618–1624. https://doi.org/10.1242/jeb.136135.
Hsu, C., Chang, C., Lin, C., 2010. A Practical Guide to Support Vector Classification. https://doi.org/10.1007/s11119-014-9370-9.
James, G., Daniela, W., Hastie, T., Tibshirani, R., 2017. An Introduction to Statistical Learning With Applications in R, 8th ed. Springer, New York.
Lofroth, E., Klafki, R., Krebs, J., Lewis, D., 2008. Evaluation of live-capture techniques for free-ranging wolverines. J. Wildl. Manag. 72, 1253–1261. https://doi.org/10.2193/2006-393.
Macander, M.J., Swingley, C.S., Joly, K., Raynolds, M.K., 2015. Landsat-based snow persistence map for northwest Alaska. Remote Sens. Environ. 163, 23–31. https://doi.org/10.1016/j.rse.2015.02.028.
McClune, D.W., Marks, N.J., Wilson, R.P., Houghton, J.D.R., Montgomery, I.W., McGowan, N.E., Gormley, E., Scantlebury, M., 2014. Tri-axial accelerometers quantify behaviour in the Eurasian badger (*Meles meles*): towards an automated interpretation of field data. Anim. Biotelemetry 2, 1–6. https://doi.org/10.1186/2050-3385-2-5.
McLoughlin, P.D., Cluff, H.D., Messier, F., 2002. Denning ecology of barren-ground grizzly bears in the Central Arctic. J. Mammal. 83, 188–198. https://doi.org/10.1644/1545-1542(2002)083<0188:deobgg>2.0.co;2.
Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., 2018. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071). Vienna University of Technology.
Mizell, D., 2003. Using gravity to estimate accelerometer orientation. Proc. Seventh IEEE Int. Symp. Wearable Comput. 252–253. https://doi.org/10.1515/zna-1966-0711.
Nakamura, I., Goto, Y., Sato, K., 2015. Ocean sunfish rewarm at the surface after deep excursions to forage for siphonophores. J. Anim. Ecol. 84, 590–603. https://doi.org/10.1111/1365-2656.12346.
Nathan, R., Spiegel, O., Fortmann-Roe, S., Harel, R., Wikelski, M., Getz, W.M., 2012. Using tri-axial acceleration data to identify behavioral modes of free-ranging animals: general concepts and tools illustrated for griffon vultures. J. Exp. Biol. 215, 986–996. https://doi.org/10.1242/jeb.058602.
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., Middel, K.R., Thiemann, G.W., Williams, T.M., 2017. Using tri-axial accelerometers to identify wild polar bear behaviors. Endanger. Species Res. 32, 19–33. https://doi.org/10.3354/esr00779.
Ware, J.V., Rode, K.D., Pagano, A.M., Bromaghin, J., Robbins, C.T., Erlenbach, J., Jensen, S., Cutting, A., Nicassio-hiskey, N., Hash, A., Owen, M., Jansen, H.T., 2015. Validation of mercury tip-switch and accelerometer activity sensors for identifying resting and active behavior in bears. Ursus 26, 86–96.
Rast, W., Kimmig, S.E., Giese, L., Berger, A., 2020. Machine learning goes wild: using data from captive individuals to infer wildlife behaviour. PloS one 15 (5). https://doi.org/10.1371/journal.pone.0227317.
Shamoun-Baranes, J., Bom, R., van Loon, E.E., Ens, B.J., Oosterbeek, K., Bouten, W., 2012. From sensor data to animal behaviour: an oystercatcher example. PLoS One 7, 28–30. https://doi.org/10.1371/journal.pone.0037997.
Shepard, E.L.C., Wilson, R.P., Quintana, F., Laich, A.G., Liebsch, N., Albareda, D.A., Halsey, L.G., Gleiss, A., Morgan, D.T., Myers, A.E., Newman, C., Macdonald, D.W., 2008. Identification of animal movement patterns using tri-axial accelerometry. Endanger. Species Res. 10, 47–60. https://doi.org/10.3354/esr00084.
Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Inf. Process. Manag. 45, 427–437. https://doi.org/10.1016/j.ipm.2009.03.002.
Studd, E.K., Boudreau, M.R., Majchrzak, Y.N., Menzies, A.K., Peers, M.J.L., Seguin, J.L., Lavergne, S.G., Boonstra, R., Murray, D.L., Boutin, S., Humphries, M.M., 2019. Use of acceleration and acoustics to classify behavior, generate time budgets, and evaluate responses to moonlight in free-ranging snowshoe hares. Front. Ecol. Evol. 7. https://doi.org/10.3389/fevo.2019.00154.
Switalski, T.A., 2003. Coyote foraging ecology and vigilance in response to gray wolf reintroduction in Yellowstone National Park. Can. J. Zool. 81, 985–993. https://doi.org/10.1139/z03-080.
Tape, K.D., Gustine, D.D., 2014. Capturing migration phenology of terrestrial wildlife using camera traps. Bioscience 64, 117–124. https://doi.org/10.1093/biosci/bit018.
Tatler, J., Cassey, P., Prowse, T.A.A., 2018. High accuracy at low frequency: detailed behavioural classification from accelerometer data. J. Exp. Biol. 221, 1–8. https://doi.org/10.1242/jeb.184085.
Thiel, A., Evans, A.L., Fuchs, B., Arnemo, J.M., Aronsson, M., Persson, J., 2019. Effects of reproduction and environmental factors on body temperature and activity patterns of wolverines. Front. Zool. 16. https://doi.org/10.1186/s12983-019-0319-8.
Viviant, M., Trites, A.W., Rosen, D.A.S., Monestiez, P., Guinet, C., 2010. Prey capture attempts can be detected in Steller sea lions and other marine predators using accelerometers. Polar Biol. 33, 713–719. https://doi.org/10.1007/s00300-009-0750-y.
Wang, Y., Nickel, B., Rutishauser, M., Bryce, C.M., Williams, T.M., Elkaim, G., Wilmers, C.C., 2015. Movement, resting, and attack behaviors of wild pumas are revealed by tri-axial accelerometer measurements. Mov. Ecol. 3, 1–12. https://doi.org/10.1186/s40462-015-0030-0.
Watanabe, Y.Y., Takahashi, A., 2012. Linking animal-borne video to accelerometers reveals prey capture variability. PNAS 110, 2199–2204. https://doi.org/10.1073/pnas.1216244110.
Whitney, N.M., Pratt, H.L., Pratt, T.C., Carrier, J.C., 2010. Identifying shark mating behaviour using three-dimensional acceleration loggers. Endanger. Species Res. 10, 71–82. https://doi.org/10.3354/esr00247.
Wilmers, C.C., Nickel, B., Bryce, C.M., Smith, J.A., Wheat, R.E., Yovovich, V., Hebblewhite, M., 2015. The golden age of bio-logging: how animal-borne sensors are advancing the frontiers of ecology. Ecology 96, 1741–1753. https://doi.org/10.1890/14-1401.1.
Yoda, K., Ropert-coudert, Y., 2007. Temporal changes in activity budgets of chick-rearing Adélie penguins. Mar. Biol. 151, 1951–1957. https://doi.org/10.1007/s00227-007-0631-2.