

chapter 2 - RF model

Dajun Wang

11/9/2019

1 Introduction

2 Material and methods

Random Forests (RF) is a relatively novel and powerful machine learning algorithm that has been reported to work well with complex ecological data that cannot be easily fitted with traditional methods such as generalized linear models. Through the use of tri-axial accelerometry measurements (i.e., acceleration values), RF models can also be used to predict unobservable behaviours in wild, free-ranging animals based on information collected from examining captive animals. To do this, the RF model must first be trained to recognize labelled behaviours in a ‘training’ dataset with decision trees constructed with suitable predictor variables (Table 3) used for classifying behaviours (Table 1).

2.1 Data collection

For the labelling of known behaviours in accelerometry data, I collared 11 healthy adult dogs of three different dog breeds: German shepherds ($n = 9$, six males and three females, age range: 1–8 years, mean age: 5 years old, Rottweiler ($n = 1$, male, age: 8 years old) and Golden Retriever ($n = 1$, male, age: 1 year old) with an accelerometer-equipped wildlife tracking collar (Type 1C-heavy, E-obs GmbH; Grünwald, Germany). The wildlife tracking

collar used in this chapter is also used for collaring and tracking free-roaming dogs in the subsequent chapters. All eleven dogs were well-trained individuals that could perform the selected repertoire of movement behaviours (Table 1) solely with the verbal commands of their trainer while being off-leash.

Table 1: Description of fine-scale behavioural tasks performed by kennel club-trained domestic dogs.

Behaviour	Description
Eat	Individual under studying is eating or drinking from a bowl whilst standing. No visible locomotion but the posture of the head is angled approximately 45° downwards.
Stand	Individual under study is standing on all four limbs with no visible locomotion. Resting posture of the dog’s head will vary between individuals.
Lie	Individuals under study is lying prone on the ground with all limbs extended out. For some individuals, the dog may be lying fully on its side. In all footages, there was no visible movement.
Sit	Individual under study is sitting on its haunches with no visible movement. Similar to ‘Stand’, head posture varies between individuals but body posture remains similar.
Walk	Individual under study is walking freely off-leash with no visible acceleration.
Forage	Similar to ‘Walk’ but the position of the dog’s head is angled downwards as it searches for treats on the ground.
Run	Individual under study is moving significantly faster than ‘Walk’ (e.g., trotting, running). The dog was tasked to chase after a tossed ball, and the task is considered completed once the tossed ball has been received by the dog.

For the purpose of collecting data to construct the RF models, the wildlife tracking collar was programmed to sample continuously at 10 hz, and the collar was mounted such that the x-, y-, and z- axes of the tri-axial accelerometer were parallel to the median (surge), the dorsal (sway), and the dorsal (heave) planes of the animal, respectively. The design and alignment of the collar provided the tri-axial accelerometer the capacity to measure the surge (back and forth movement on the x-axis), sway (left and right motion on the y-axis) and heave (up and down on the z-axis) motion of the animal. Data was collected in October 2016 and May 2019 in a grassy field to a) simulate the environment where the intended free-roaming dogs for study are typically found, and b) provide sufficient ground for the dogs to move continuously. The terrain selected for this component was relatively flat to avoid compromising the gravitational forces acting on the heave axis.

Table 2: Description of the coarse-scale behavioural tasks re-categorized from footages of fine-scale dog behaviours.

Behaviour	Description
Eat	Repeated as fine-scale 'Eat' behaviour.
Fast	Repeated as fine-scale 'Run' behaviour
Slow	Grouped behaviour consisting of 'Walk' and 'Forage' to represent a moving dog with slower locomotion
Stationary	Grouped behaviours consisting of 'Sit', 'Stand', and 'Lie' to represent a completely still dog.

In the collection of accelerometer measurements for fine-scale movement behaviours, the dogs were instructed to perform the 'fine-scale' behavioural tasks (Table 1) continuously for at least 2 min in a stipulated sequence except for Forage, Eat and Run as these tasks are either energetically taxing for the dogs or behaviourally inconsistent. Each dog was tasked to perform the entire repertoire of behaviours at least twice (with verbal commands from their trainer), and approximately 300 minutes of video footages were obtained. As for the

collection of ‘coarse-scale’ behaviours, footages containing fine-scale movement behaviours and their associated accelerometer measurements from the aforementioned component were re-categorized accordingly (Table 2).

Table 3: Predictor variables and their statistic labels used for predicting dog behavioural tasks in the random forest models.

Labels	Variables	Description
Mean	mean.x, mean.y, mean.z	The calculated mean of the acceleration measurements within each burst for axes x, y and z.
Min	min.x, min.y, minz	The smallest acceleration measurement within each burst for axes x, y and z.
Max	max.x, max.y, max.z	The largest acceleration measurement within each burst for axes x, y and z.
Kurtosis	kurt.x, kurt.y, kurt.z	The relative flatness of the acceleration measurements within each burst for axes x, y and z.
Skewness	skew.x, skew.y, skew.z	The relative skewness of the acceleration measurements within each burst for axes x, y and z.
Range	range.x, range.y, range.z	The calculated difference between the largest and smallest acceleration measurements of axes x, y, and z.
Standard deviation	sd.x, sd.y, sd.z	The calculated standard deviations of the acceleration measurements of axes x, y, and z.
Pitch	pitch	The calculated ratio between axes x, y, and z.
Roll	roll	The calculated angle between axes y and z.

(continued)

Labels	Variables	Description
Overall Dynamic Body Acceleration	ODBA	The sum of the dynamic acceleration values of axes x, y, and z.

2.2 Data Management

The collected accelerometry data was viewed with an acceleration viewer (<http://www.movebank.org>, version 33) and the time-stamps of the collected accelerometer measurements were synchronized and manually labelled to the time-stamps of the video footages of each behavioural task. The labelled acceleration measurements were binned into two-second windows (or ‘bursts’) which accomodates at least two full strides of movement-based dog behaviours (i.e., walking, foraging) while reducing the influence of un-intentional behavioural transitions.

For the construction of RF models, a raw acceleration (RAW) dataset was derived by applying a series of summary statistics (Table 3) onto each labelled burst of raw acceleration measurements from all three axes. In addition, the periodic properties of the acceleration signals in each labelled burst of accelerometer measurement allowed me to apply a fast Fourier transformation in order to determine the frequency of each particular movement, thus deriving a fast-Fourier transformed (FFT) dataset. For both RAW and FFT dataset, the Overall Dynamic Body Acceleration (ODBA) was also included by summing the absolute dynamic acceleration of each burst after removing the static acceleration component (running mean of 2s; Gleiss, Wilson, and Shepard (2011)). In total, 24 predictor variables were calculated from each burst of (transformed) data and these variables have been used in other movement-based studies (Wang et al. 2015; Nathan et al. 2012)

2.3 The construction of RF models

The `randomForest` package (Liaw and Wiener 2002) in R was used to prepare and construct the various models used for classifying dog behaviours. In the construction of all RF models, the labelled datasets were randomly subsampled at a 80% proportion to form a training dataset for preparing the RF model while ensuring that each specific behaviour were sufficiently represented where possible. In addition, the default number of trees ($n = 500$) were used in the RF model construction as the Out-of-Bag (OOB) error estimates, a method of examining classification error rates, were found to be consistently stable (Figure 2) across all models.

To maximise the RF model’s classification capabilities, I cross validated the RF models by comparing the classification accuracies of RF models constructed ($n = 24$) with a range of `mTry` values mirroring the number of described predictor variables (Table 3) used. The `mTry` value represents the number of predictor variables used for making a classification (Liaw and Wiener (2002); i.e., splitting a ‘tree node’), and the use of an optimum `mTry` value contributes significantly to the RF model’s classification accuracy.

For the classification of fine-scale and coarse-scale behaviours, two different RF models were constructed for each type of classification based on the type of dataset used (i.e., raw acceleration (RAW) or fast Fourier transformation (FFT)). In the comparison of classification accuracies within each scale, I used accuracy, precision and recall to measure the behaviour classification performance of each model. Accuracy is the overall measure of the model’s classification capacity and is defined as the proportion of correctly classified behaviour. For individual behaviours, precision is defined as the proportion of positive classifications from the actual behaviours (e.g., no. of positive / no. of actual) while recall (also deemed as sensitivity) refers to the proportion of model predictions that were positive (e.g., no. of positive / no. of wrong) (Bidder et al. 2014).

Accuracy was used as the overall performance metric due to its simplicity and the fact that it takes into account all classification outcomes (Bidder et al. 2014). To evaluate

classification performance of each individual behaviour, precision was used as the main performance metric as it is most applicable to biological inferences which generally rely on true positive classifications, as was the case in this study (Bidder et al. 2014). Recall was included as recommended by Bidder et al. (2014) as it allows one to examine the model’s capacity to produce positive classifications and compares between classification methods.

2.4 Accelerometer threshold accelerometer-informed GPS sampling

The wildlife tracking collar is equipped with the capacity to sample GPS relocations dynamically based on the accelerometer measurements collected in real-time. With the use of a selected accelerometer threshold (ACT), researchers are able to selectively intensify the sampling of GPS re-locations only when the animal is moving (hence higher variance in accelerometer readings). This feature not only reduces the energy and memory consumption during the research period, it also allows the researcher to identify movement paths or patterns that are ecologically interesting (e.g., predation or harrassement events).

For this function to work realistically, I used the above constructed RF model to identify the variances of the predicted movement behaviours and selected a variance value (Table 9) for the axis that best represents movement or motion in a dog (i.e., Z axis; the acceleration in the heave axis typified by the rolling gait in dogs). Following which, I ground truth the selected variance by walking four dogs (Mongrel breeds, 4 females, age range 2-7 years old) with the wildlife tracking collar equipped. All dogs were walked continuously for approximately ten minutes, and with the studied dog resting (with the collar on) only in the beginning and end of each walk. The dogs were rested in a non-motion movement behaviour (e.g., lying, sitting or standing) to simulate the resting behaviour of free-roaming dogs in the wild. The GPS and ACC dataset were then retrieved from the collar, and the occurrence of quick- and long-burst gps relocations were plotted sequentially against the variance of the accelerometer measurements collected from all three axes (Figure 3).

3 Results

For the purpose of building the training dataset of the RF model, I collected more than 400,000 raw acceleration measurements (per axis) and approximately 400 minutes of video footage from 11 kennel-trained domestic dogs. From the collected video footages, seven different dog behaviours were identified which included sitting, standing, lying, eating, walking, foraging, and running (Table 1). After identifying and cleaning the collected acceleration measurements, approximately 3000 bursts of labelled acceleration data comprising of nearly 1,260,000 transformed acceleration measurements (from the three axes) were used to construct and train the RF model. The RF models constructed with the four types of datasets produced reliable classification performances as Out-of-Bag error rates remained relatively stable (Figure 2).

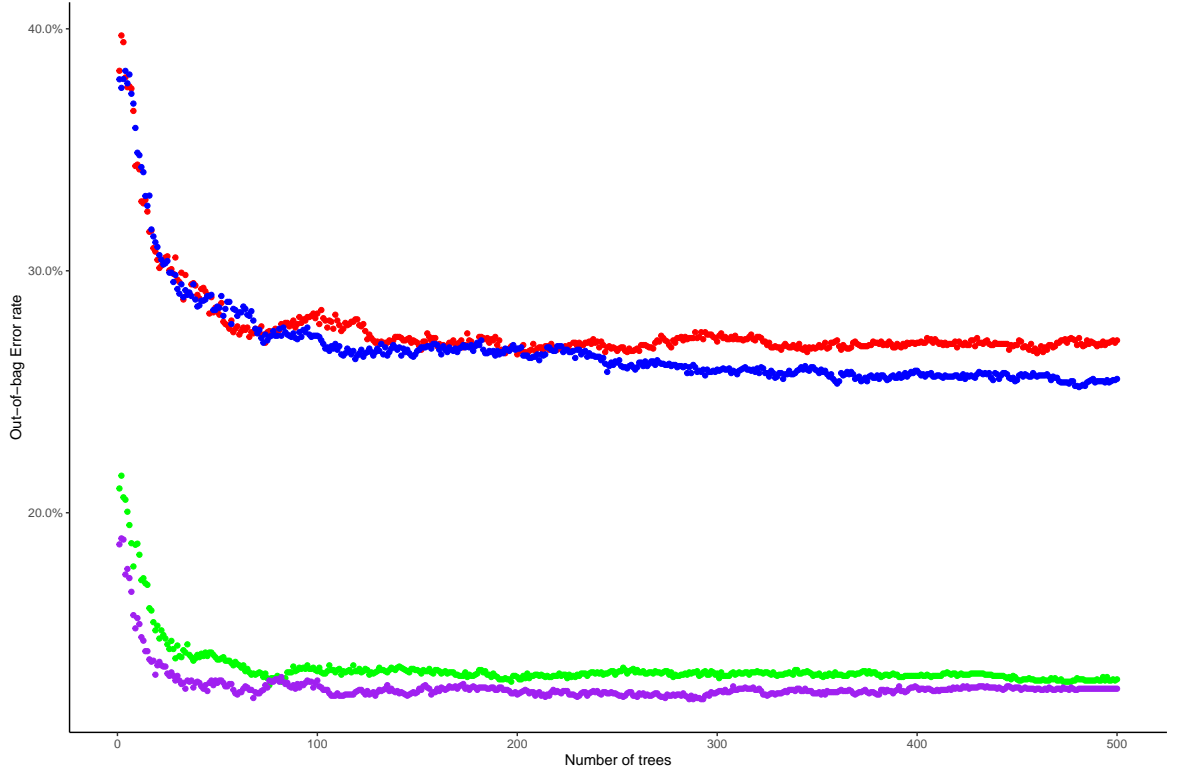


Figure 2: The Out-Of-Bag error estimates of the Random Forest model plotted against the number of trees (n) used in making predictions. The red and blue dots represents the fine-scale Random Forest models constructed with the Fast-fourier transformation (FFT) and

raw acceleration measurement (RAW) datasets respectively.

Table 4: Confusion matrix of actual (rows) vs. predicted (columns) of fine-scale dog behaviours derived from the Random Forest model constructed with the Fast-fourier transformation (FFT) dataset.

Behaviour	Eat	Forage	Lie	Run	Sit	Stand	Walk	Precision	Recall
Eat	39	5	13	0	2	3	4	59.09%	81.25%
Forage	1	65	11	4	4	1	40	51.59%	63.11%
Lie	7	7	297	1	70	27	25	68.43%	66.15%
Run	0	2	0	350	0	1	55	85.78%	84.75%
Sit	1	7	82	0	241	13	12	67.70%	71.30%
Stand	0	3	41	4	17	96	20	53.04%	67.13%
Walk	0	14	5	54	4	2	422	84.23%	73.01%

Table 5: Confusion matrix of actual (rows) vs. predicted (columns) of fine-scale behaviours derived from the Random Forest model constructed with the raw acceleration measurement (RAW) dataset.

Behaviour	Eat	Forage	Lie	Run	Sit	Stand	Walk	Precision	Recall
Eat	33	6	16	0	3	1	7	50.00%	76.74%
Forage	1	59	12	7	2	1	44	46.83%	60.82%
Lie	5	9	309	2	61	23	25	71.20%	69.75%
Run	0	3	2	353	0	0	50	86.52%	86.52%
Sit	3	8	70	1	246	15	13	69.10%	74.55%
Stand	1	3	29	1	14	109	24	60.22%	70.78%
Walk	0	9	5	44	4	5	434	86.63%	72.70%

- Comparing between fine-scale models (Table 4, Table 5), the RAW dataset produced an overall classification accuracy of 74.47% while the FFT dataset produced an overall classification accuracy of 72.88%.
- Looking specifically at the precision of fine-scale model’s classification, both models

were found to classify movement-based behaviours (e.g., ‘Run’, ‘Walk’ and ‘Forage’) more positively than non-movement based behaviours (e.g., ‘Lie’, ‘Sit’, and ‘Stand’).

- For both models, ‘Eat’ was classified most poorly in comparison to the other behaviours, with the FFT dataset performing comparatively better.
- In addition, it appears that the mis-classification of behaviours tend to occur interchangeably. E.g., ‘Sit’ as ‘Lie’, ; and ‘Lie’ as ‘Sit’ while the mis-classification for ‘Stand’ tends to be distributed evenly between ‘Lie’, ‘Sit’ and ‘Walk’.
- Looking at models’ sensitivity (or recall ability) to predict individual behaviours, both models performed relatively well with most accuracies averaging above 65%. This suggests that both models are able to perform relatively well at behaviour predictions.

Table 6: A confusion matrix of actual (rows) vs. predicted (columns) of coarse-scale behaviours derived from the Random Forest model constructed with the raw acceleration measurement (RAW) dataset.

Behaviour	Eat	Fast	Slow	Stationary	Precision	Recall
Eat	20	0	11	35	30.30%	76.92%
Fast	0	346	59	3	84.80%	86.72%
Slow	2	47	540	39	85.99%	80.84%
Stationary	4	6	58	904	93.00%	92.15%

Table 7: A confusion matrix of actual (rows) vs. predicted (columns) of coarse-scale behaviours derived from the Random Forest model constructed with the Fast-fourier transformation (FFT) dataset.

Behaviour	Eat	Fast	Slow	Stationary	Precision	Recall
Eat	22	0	11	33	33.33%	78.57%
Fast	0	340	65	3	83.33%	83.95%
Slow	2	56	537	33	85.51%	80.27%
Stationary	4	9	56	903	92.90%	92.90%

- Comparing between coarse-scale models (Table 6, Table 7), both types of datasets produced comparable prediction accuracies. I.e., RAW 87.27% vs FFT 86.89%.
- Similar to the fine-scale behaviour predictions, both coarse-scale models produce non-discernable differences in precision (i.e., correct behaviour classification) and recall ability (i.e., positive behaviour prediction) between them.

Table 8: A confusion matrix of actual (rows) vs. predicted (columns) of grouped fine-scale behaviours derived from the Random Forest model constructed with the Fast-fourier transformation (FFT) dataset.

Behaviour	Eat	Fast	Slow	Stationary	Precision	Recall
Eat	39	0	9	18	59.09%	81.25%
Fast	0	350	57	1	85.78%	84.75%
Slow	1	58	541	27	86.28%	79.44%
Stationary	8	5	74	884	91.04%	95.05%

- Unlike the fine-scale behaviour models, the precision of coarse-scale models were noticeably higher, except for the ‘Eat’ behaviour, of which the fine-scale model were more precised
- In addition, the grouped fine-scale behaviour model (grouped behaviour model; Table 8) produced an classification accuracy of 87.55% which was comparable to classification from both coarse-scale models.
- However, the grouped behaviour model performed significantly better due to the higher

classification accuracy for the ‘Eat’ behaviour, which can be useful for fine-scale behaviour examination.

Table 9: The mean, minimum and maximum variance values of accelerometer measurements for the movement behaviours of dogs.

Activity	Behaviour	Mean			Median		
		X	Y	Z	X	Y	Z
Resting	Lie	4382.160	5236.047	3097.256	994.4877	693.0654	554.8569
	Eat	7497.864	4759.058	4506.089	876.4385	726.9538	573.3985
	Sit	5966.968	3639.992	7489.651	1363.1085	1878.6846	1209.4341
	Stand	10536.696	15618.678	7963.085	3127.2800	2393.4560	2586.0185
Movement	Forage	26128.293	23022.594	22260.306	16095.5015	17731.9938	9886.4615
	Walk	56863.595	34096.077	42351.569	43938.8615	23650.1862	23709.2738
	Run	335325.959	152629.015	326456.067	223783.0185	121704.4369	273266.6838

Using the variances calculated from the labelled acceleration measurements (Table 9), I sampled a range of variance values between ‘Stand’ and ‘Forage’ across the three axes to identify a variance threshold value (VTV) that could consistently initiate the collection of short-bursts GPS re-locations in the wildlife tracking collar. To do this, I used the `rle()` function from base R to examine and count the frequency of which a feasible VTV (from the range of feasible variance values) is met or exceeded by the variance of movement-based labelled acceleration measurements across the three axes. The VTV that determines the highest frequency of met or exceeded variance values was then used to program the wildlife tracking collar for the dog walking experiment.

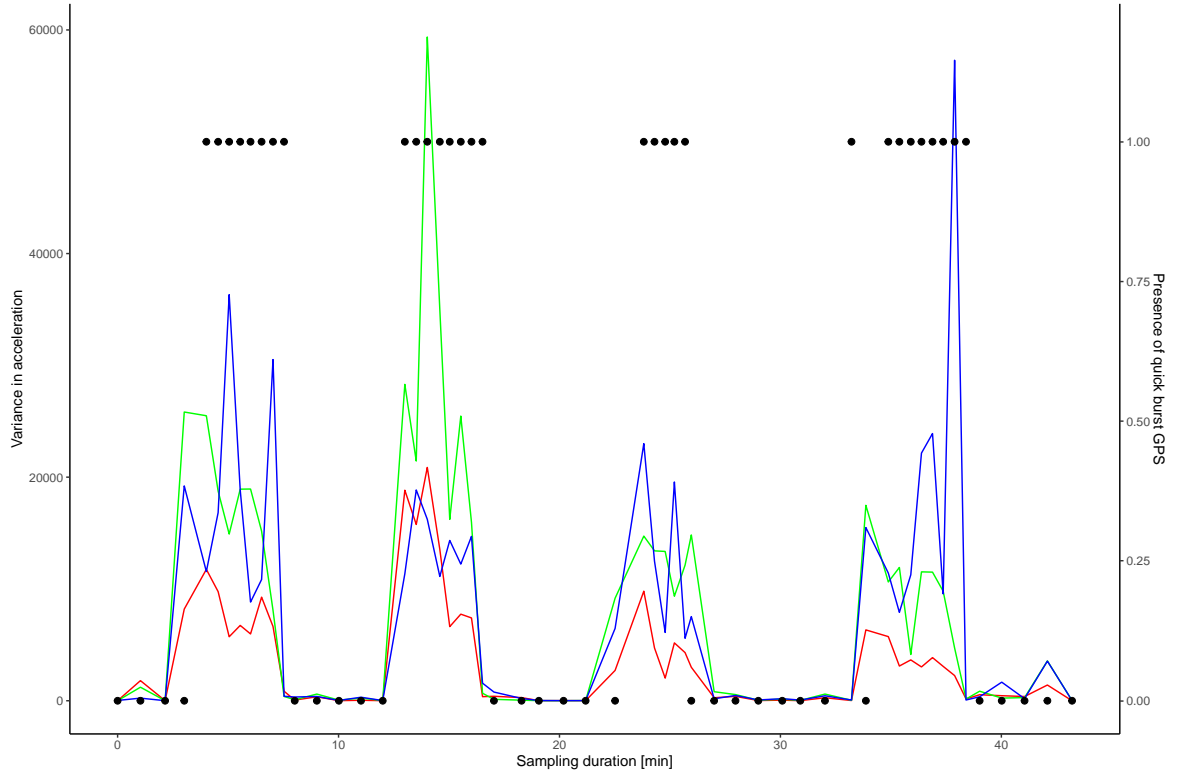


Figure 3: The sampling of GPS re-locations based with an accelerometer threshold of 10,000 in the Z axis. The red, green and blue lines represents variance in acceleration measurements collected from the X, Y and Z axis respectively.

With the VTV identified and programmed into the collar, Figure 3 found that the collection short-burst GPS re-locations was most consistent when a VTV of 10,000 is programmed onto the Z-axis. For example, the collection of short-term GPS re-location is only triggered if the variance of collected acceleration measurements from the Z axis exceeds 10,000 repeatedly. That being said, movement is typically associated with acceleration in the surge axis (X-axis; red line) yet the heave axis (Z axis) appears to be more sensitive to the up-and-down acceleration forces exerted by rolling gait of a walking dog. Seemingly, having access to the ground truthed information on the acceleration forces of a moving dog allows one to program a suitable VTV to examine biologically and ecologically interesting movement patterns in the subsequent chapters.

4 Discussion

- The fine-scale models have the lowest classification rates for ‘Forage’ and ‘Eat’ behaviours.
- As discussed previously, it is likely that the RF model was unable to accurately predict ‘Forage’ and ‘Eat’ accurately as these behaviours bear strong postural and movement resemblance to ‘Walk’ and ‘Lie’ (e.g., and, respectively).
- E.g., the ‘Eat’ behaviour represents the dog in-study consuming or drinking from a bowl and the these dogs often assume a lying position in the duration of this behavioural task. Likewise, the ‘Forage’ behaviour can be difficult for the RF model to differentiate from the ‘Walk’ behaviour as the behavioural task for ‘Forage’ was represented by the dog in-study walking with its head angled downwards (Figure 4).

5 Conclusion

6 References

Supplementary codes

Bidder, Owen R., Hamish A. Campbell, Agustina Gómez-Laich, Patricia Urgé, James Walker, Yuzhi Cai, Lianli Gao, Flavio Quintana, and Rory P. Wilson. 2014. “Love Thy Neighbour: Automatic Animal Behavioural Classification of Acceleration Data Using the K-Nearest Neighbour Algorithm.” Edited by Allan V. Kalueff. *PLoS ONE* 9 (2): e88609. <https://doi.org/10.1371/journal.pone.0088609>.

Gleiss, Adrian C., Rory P. Wilson, and Emily L. C. Shepard. 2011. “Making Overall Dynamic Body Acceleration Work: On the Theory of Acceleration as a Proxy for Energy Expenditure.” *Methods in Ecology and Evolution* 2 (1): 23–33. <https://doi.org/10.1111/j.2041-210X.2010.00057.x>.

Liaw, Andy, and Matthew Wiener. 2002. “Classification and Regression by randomForest.” *R News* 2 (3): 6.

Nathan, R., O. Spiegel, S. Fortmann-Roe, R. Harel, M. Wikelski, and W. M. Getz. 2012. “Using Tri-Axial Acceleration Data to Identify Behavioral Modes of Free-Ranging Animals: General Concepts and Tools Illustrated for Griffon Vultures.” *Journal of Experimental Biology* 215 (6): 986–96. <https://doi.org/10.1242/jeb.058602>.

Wang, Yiwei, Barry Nickel, Matthew Rutishauser, Caleb M Bryce, Terrie M Williams, Gabriel Elkaim, and Christopher C Wilmers. 2015. “Movement, Resting, and Attack Behaviors of Wild Pumas Are Revealed by Tri-Axial Accelerometer Measurements.” *Movement Ecology* 3 (1): 2. <https://doi.org/10.1186/s40462-015-0030-0>.