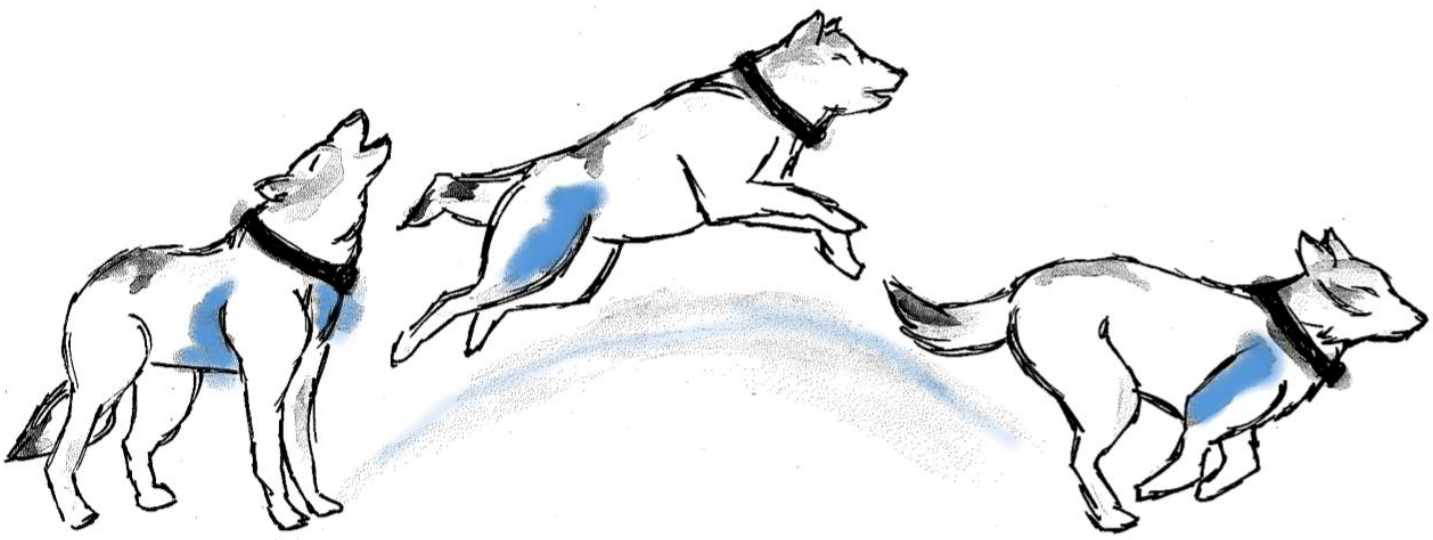
UNIVERSITÉ DE LYON

UNIVERSITÉ JEAN MONNET SAINT-ÉTIENNE

Master Ecology-Ethology

Connecting acceleration data to specific behaviors in captive wolves (*Canis lupus*)



Léa Bouet

Master thesis 2019

Inland Norway University of Applied Sciences

Abstract

Background:

Currently the barriers between animals and men tend to decrease due to fragmentation, degradation or loss of natural habitats. Therefore, gathering information about animal movements and behaviors is crucial. Indeed, knowledge about changes in animals' home range and behaviors in responses to those landscape changes gives highly valuable information for developing both management and conservation measures for species and populations.

Animal-attached accelerometers provide valuable high-resolution information for identifying behaviors by providing measures of animals' body motion, posture and proxies for energy expenditure. Interestingly, this technique has been applied successfully to large carnivores including wild dogs, polar bears and pumas, but has not yet been calibrated for wolves.

Description:

Here we present a pilot study aiming to develop and validate a method which can automatically identify common and specific wolf behaviors from acceleration data. Acceleration data were sampled from captive wolves at 32 Hz in the three dimensions. In parallel, surveillance cameras were used to record their behaviors during three periods of 24h for each group of wolves. In this report, we have analyzed the behavior of three wolves from one enclosure over a 24h period.
We used a supervised machine learning algorithm, called Random Forest (using the"h2o" R package), to classify accelerometer data into behavioral classes. We selected 8 consistent and ecologically meaningful behaviors for our models: walking, trotting, galloping, stationary, howling, chewing, digging and smelling.

Conclusion:

Random Forest models showed a strong ability to discriminate between the eight selected wolf behaviors. By testing both individual wolves and grouped models, we achieved a log loss (logarithmic loss metric) with a range very close to 0 (< 0.003), meaning models are correctly assigning behaviors with a probability near 100%.
Then, with the grouped model, that combined the data sets from the three wolves, we demonstrated that we were able to classify perfectly five behaviors: "digging", "galloping", "howling", "smelling" and "trotting".
Furthermore, the Overall Dynamic Body Acceleration (odba) appeared to be the principal variable that contributed the most to building our decision trees. Then, pitch and the amplitude variables followed in varying order of importance, depending on the model.

Key-words: Accelerometry, Animal behavior, Wolves, *Canis lupus*, Machine learning, Random Forest.

Acknowledgments

I would especially like to thank Alina Lynn Evans and Ane Eriksen, who responded favorably to my request for a spontaneous internship in their team and gave me all their confidence. Working at Inland Norway University of Applied Sciences and being involved in field work on other projects has been a real opportunity.

Besides my internship supervisor, I would like to thank the rest of my thesis committee: Oliver Devineau for his help with the R script and statistics along with Master students Erik Versluijs and Tine S. Johansen for their stimulating discussions and help.

I would also like to thank Marianna Chimienti from Aarhus University (Denmark) for collaborating on this project. It allowed me to learn how to analyze acceleration data. Many thanks for her pedagogy and availability during my two weeks visiting her team at Aarhus University.

Moreover, this project would not have been possible without the collaboration of the Wildlife Science Center in Minnesota that allowed us to use their wolves for this study and were responsible for all the immobilizations and captures. They also performed the filming after the initial camera set-up. And, I would like to mention specially Tim Laske who helped with the camera set-up with power bank and solar panels and Kristoffer Nordli who supervised the collaring of the wolves.

Last but not least, I would like to thank my parents who supported me for this project.

Table of Contents

Introduction

Presently, boundaries between animals and human are decreasing due to fragmentation, degradation and loss of natural habitats. Animals may not be able to adapt their behavior appropriately to these rapid changes and they may be faced with isolation from one another and from their food sources. Animal's home ranges are affected by these parameters, such as carnivore species which range over very large areas and often come close to human. This can lead to direct conflict of interest between humans and wildlife (Mattison et al., 2013). Therefore, the understanding and the management of animal movements and behaviors in natural landscapes is becoming increasingly important, especially for developing both management and conservation measures of species and populations (Doherty and Driscoll, 2018).

For carnivorous species, such as wolves or bears, conservation policies depend on the sociopolitical contexts as much as on the biological landscape. For instance, one of the biggest problems in wolf management is the perceptions and feelings of people. Most of the time wolves symbolize danger and threat, and the issues are well known: livestock depredation, competition for huntable game, mutilated dogs and the fear of attacks on people. So, while there are conflicts between wolves and humans, there are also conflicts between people, involving groups with diverging interests on that topic (Skogen and Krange, 2003).

A better understanding of the behaviors of wild wolves could potentially reduce people's fear. In this way, future strategies for developing conservation policies of carnivorous animals must be situation-specific and knowledge based (Treves and Karanth, 2003).

In Scandinavia, GPS collars have been used on wolves for over 15 years, providing valuable information on topics including predation (Sand et al., 2005, San et al, 2008 and Zimmermann et al, 2015), activity patterns (Eriksen et al., 2011), competitions with other species (Wikenros et al., 2010), habitat selection and space use (Sanz-Pérez et al., 2018 and Milleret et al., 2019), distribution and abundance (Chapron et al., 2014) and even causes of their mortality (Liberg et al., 2011). All these elements are based on GPS movement data.

Accelerometer devices can record acceleration at a high temporal resolution of up to 300 Hz (Wilson et al., 2013) and in three dimensions. Accelerometers provide measures of animal body motion, posture and can also ebe used as a proxy for nergy expenditure (Halsey et al., 2008). Indeed, one of the main objectives of accelerometer measurement is to use patterns of accelerometer waveforms to deduce specific behaviors through animal movement and body posture. The second main objective is to use the variation in accelerometer waveform measurements to correlate with energy expenditure (Brown et al., 2013). Therefore, accelerometry is a potentially powerful tool, but the potentially high resolution is limited by storage and battery capacities.

Acceleration data represents highly valuable information for identifying animal behaviors (Yoda et al., 1999; Wilson, 2008). Two different approaches can be used to characterize behaviors from accelerometry data. Firstly, the unsupervised learning approach, based on machine learning, allows for the classification of data that has not been labelled and calibrated with observed behaviors (Chimienti et al., 2016). Secondly, the supervised learning approach, based on the manual labelling of the behavioral database using observed behaviors, allows for creating a trained algorithm to determine the acceleration pattern for each behavior of the

species of interest (Resheff et al., 2014). The unsupervised learning approach is however limited by the number and type of behavioral groups, yet allows detection of unknown behaviors that would not be detected with a supervised approach

Some studies have used biologging technology such as triaxial acceleration to investigate and reveal a complex picture of the behavior and ecology of animals in their natural habitat, out of sight of human observers (Rutz and Hays, 2009; Graf et al., 2015). Interestingly, this technique was successfully applied to large carnivores like wild dogs (English, 2018), polar bears (Pagano et al., 2017) and pumas (Wang et al., 2015) but so far, it has not been reported for wolves.

However, having a model that can be used to identify behavioral patterns in free- ranging Scandinavian wolves will be extremely valuable for conservation and management issues by revealing the detailed responses of wolves to their surrounding landscapes. Indeed, this may allow future application of the method to classify behaviors of wild wolves equipped with accelerometers, adding high-resolution behavioral information to the GPS location data, which would give information on how wild wolves are responding to anthropogenic landscape features.

In this study, we aimed to develop and validate a method for analyzing an accelerometer dataset that is able to automatically identify common or specific behaviors from the collar of captive wolves of which the behaviors were observed using triaxial accelerometer loggers at 32Hz at the Wildlife Science Center in Minnesota. Using surveillance cameras, behaviors were recorded continuously during bouts of 24 hours. Then, a supervised machine learning algorithm was used to classify accelerometer data into behavioral classes. We selected the Random Forest algorithm as is performing the best (Nathan et al., 2012).

Material and Methods

Location of the data collection

Data collection took place between October 2018 and January 2019 at the Wildlife Science Center in Minnesota (U.S.A, N45°23’51, E93°1’3.72).

Enclosure setup

The captive wolves (*Canis lupus*) studied were regrouped in four enclosures (Appendix 1) comprising a total of five to eight wolves in each.

In the enclosures, water was available *ad libitum* and wolves received carcasses from road kills for food. All enclosures contained shelters. The first, second and fourth enclosure did not contain endemic vegetation, but they were enriched with a concrete pipe. In contrast, the third enclosure was closer to natural condition containing endemic vegetation and both standing and downed trees.

Study subjects

Twelve captive wolves (*Canis lupus*) were studied, three in each of the four enclosures (Table 1). To be able to tell apart the different wolves in the camera footage, specific parts of their body were dyed with black hair coloring during the immobilization. In each enclosure, one wolf was dyed on both sides of its chest and ribcage, another on both hips, and the last one on both shoulders.

The wolves were immobilized using low doses of ketamine and xylazine to immobilize them (Kreeger et al.1986) and equipped with Vectronics GPS collars. During the time of collar deployment, measurements (height, forelimb, hindlimb, and body length) were also taken, following the same protocol as for wild wolves in Scandinavia (Appendix 2). These measurements would be used for calculating energy expenditure in later studies.

The three focal wolves from the first enclosure were orphaned siblings from a litter of wild wolves. The three siblings were taken to the center and raised by a pair of foster parents. When the data collection started, they were 1.5 years old and still sharing the enclosure with their foster parents.

Table 1. Groups of grey wolves (N=12), *Canis lupus*, housed in the Wildlife Science Center in Minnesota (U.S.A).

Subject	Sex	Age (years)	Description	Enclosure
Keneli	♂	1.5	38.5 kg, dye both sides of chest/ribcage, wild-born, wolf-raised	1
Page	♀	1.5	34 kg, dye both hips, wild-born, wolf-raised	1
Chelsea	♀	1.5	32 kg, dye both shoulders, wild-born, wolf-raised	1
Odin	♂	1.5	47.6 kg, dye both sides of chest/ribcage	2
Siff	♀	1.5	34.3 kg, dye both hips	2
Loki	♂	1.5	45.5 kg, dye both shoulders	2
Kameron	♂	1.5	47 kg, dye both sides of chest/ribcage	3
Dwight	♂	1.5	50 kg, dye both hips	3
Mara	♀	1.5	34 kg, dye both shoulders	3
Laski	♂	1.5	54.5 kg, dye both sides of chest/ribcage	4
Lexi	♀	1.5	40.8 kg, dye both hips	4
Parker	♀	1.5	31.8 kg, dye both shoulders	4

Camera setup and filming

In order to link acceleration data to observed behaviors for each captive wolf, four surveillance cameras (Cromorc wireless 1.3 MP security camera system) were installed at each corner of the enclosure. They were able to record during the night using infra-red LEDs, but unable to record sounds.

As only three GPS collars were available for this study (see section *Accelerometer devices*), data were not collected from all enclosures at the same time. Three wolves from the first enclosure were equipped with the GPS collars in October 2018, then three wolves from the second enclosure in November 2018, three from the fourth enclosure in November and December 2018 and lastly three wolves from the third enclosure in January 2019.

Cameras were recording for 24h periods, two or three periods for each set of wolves. The first enclosure was filmed the 16/10/2018, 19/10/2018 and 30/10/2018, the second the 05/11/2018 and 13/11/2018, the third the 08/01/2019 and 10/01/2019 and the fourth one on the 18/11/2018, 25/11/2018 and 07/12/2018. Video data files of a 24h period was composed of 24 files of 1-hour duration for each camera.

Building the ethogram

An ethogram was built (Appendix 3) based on observed behaviors over 4 hours of video footage (30 min every 3 hours through a 24-hour period) for each enclosure except for enclosure number 4 because the data was not available at this time. In total, the ethogram was made from 12 hours of video observations. Because this behavioral data bank can be used for a number of studies and scientific questions, we chose to record as many relevant behaviors as possible, which could be grouped or down sampled later if necessary. The ethogram included 38 behaviors, divided in two groups. The first group was composed of 25 behaviors relevant for this accelerometry study, and the second group was composed of 13 additional behaviors relevant for another master thesis in progress about connecting heart rate to specific behaviors and interactions between wolves and humans or dogs.

Analyses of the videos

Behaviors were registered and categorized using the program BORIS© (Behavioral Observation Research Interactive Software, v.7.4.1) which can run all four cameras simultaneously. This program can record behaviors in two different ways: as a “point event” or a “state event”. Point event behaviors are recorded as behaviors without duration while state event behaviors have a duration with a start and stop time recorded. We recorded state event behaviors when they lasted at least 5 seconds to pick up clear, unambiguous behaviors. For common and long state event behaviors such as walking, trotting and resting, they had to last at least 10 seconds to be recorded. Only clear and distinct behaviors were picked up and only when we were fully confident that we could identify the wolf correctly.

Because the four cameras did not start precisely at the same time, we used the software Movie Maker© (v 16.4.3528.0331) to cut the 1-hour videos files of each camera in order to run them simultaneously in BORIS. We ran this process for all 24-hour periods for the first enclosure and on one 24-hour period for the three other enclosures. We had to cut 576 one-hour files (144 for each camera), and it took at least 15 minutes of processing for each 1-hour file, therefore requiring around 144 hours for cutting videos.

For this study we wanted to analyze one 24-hour period for each enclosure to have data on 12 wolves and different types of enclosures. However, due to power cuts and wolves chewing the power cables, the cameras reset to incorrect times. The exact time is necessary in order to match the videos to the acceleration data. Therefore we choose to only focus on the video data with correct timing (all periods of the first enclosure and the first period of the second enclosure).

We started by analyzing the second 24-hour period of the first enclosure (19/10 to 20/10). After more than 20 minutes the cameras sometimes desynchronized again. We decided not to analyze behaviors after desynchronization to conserve an exact behaviors timing of the observed behaviors. So, we only analyzed 21-hours based on the second 24-hour period of the first enclosure for the three following wolves, Keneli, Page and Chelsea. Because analyzing the videos for this period took 50 hours, we did not have time to analyze other periods.

Accelerometer devices

Threes Vectronics GPS collars (one VERTEX Plus Iridium V 2.1 collar and two VERTEX Plus GSM V 2.1 collars, Vectronic Aerospace Inc) were used for the study.

All the Vertex Plus GPS collars have triaxial acceleration sensors which are customizable from 2 to 32 Hz. The higher the frequency, the more precise the identification of specific behaviors. This however produces a very large amount of data thus becoming a challenge for storage and data processing.

For this study, acceleration data were sampled at a frequency of 32 Hz (32 data point/rows per axies per second) and sensitivity of 4 G in the three dimensions (surge (x axis), sway (y axis) and heave (z axis)) (Fig.1). Acceleration data can later be down sampled at 16, 8 or 4 Hz.

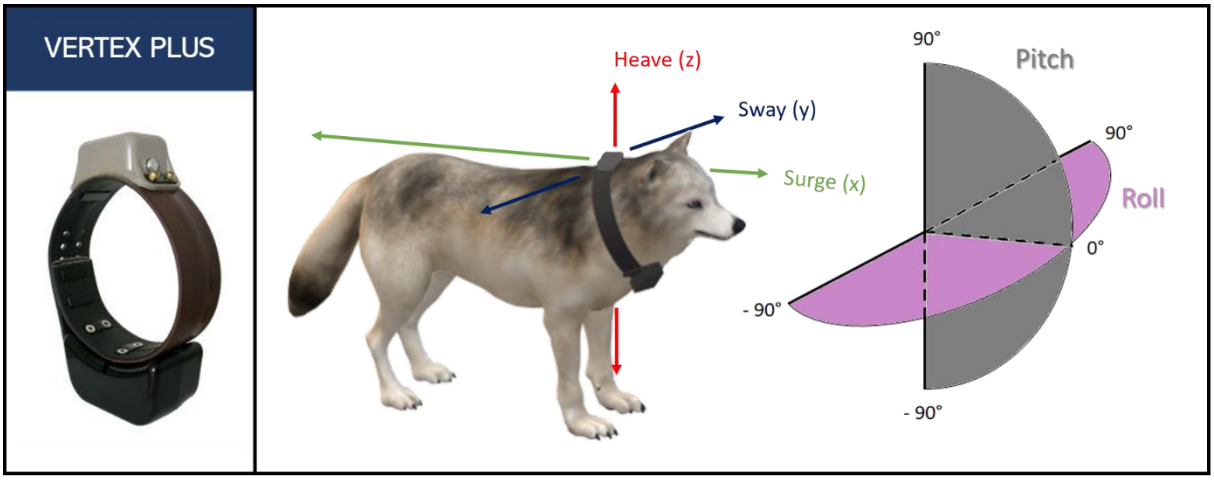


Figure 1. Schematic representation of the orientation of a triaxial accelerometer on a wolf *C.lupus*. Arrows indicate the three-dimensional movements (surge, sway and heave) recorded by the accelerometer.

Data preparation and variable extraction

Triaxial acceleration data were save as BINV2 files. In order to see the files and extract them in CSV files, we use the following software : Acceleration Data Viewer V1.0.2 Setup (Windows user interface application) from https://www.vectronic-aerospace.com/wildlife-monitoring/downloads/#Software.

By checking if the devices had recorded the acceleration data correctly and if they had problem with time stamps, we found out that the GPS collars were rebooting each day for a period of around 10 seconds. We corrected the data and filled time gaps by NA's (no data).

Then, we extracted new variables based on the raw data from the axes (Table 2). Static acceleration is calculated as a running mean of 1 second of the raw data. It's includes gravity and measure the incline of the accelerometer for each axis. This variable measured along three axes allowed for the calculation of the angle of the neck of the instrumented wolf: the body pitch (1) and roll (2) (Wilson et al., 2008). The pitch is the vertical position of the neck across a 180° angle and the roll is the lateral body orientation across a 180° angle (Fig.1). The dynamic acceleration (3) gives information on movement efforts. We used a running mean of 1 second to extract dynamic acceleration. The amplitude of the surge, sway and heave (4) were calculated with 4 running means: 5 sec, 10 sec, 20 sec and 30 sec; to highlight the consistency of behaviors over different time windows (Appendix 4). The variance of the pitch (5) was also measured with the same four running means. This variable allows to highlight the big changes in the posture of the neck. The overall dynamic body acceleration (ODBA) (6) was extracted to quantify overall effort along all axes and can be a proxy for energy expenditure.

$$pitch = \tan^{-1}\left(\frac{S_{surge}^2}{\sqrt{S_{sway}^2 + S_{heave}^2}}\right) * \frac{180}{\boldsymbol{\pi}} \quad (1)$$

$$roll = \tan^{-1}\left(\frac{S_{sway}^2}{\sqrt{S_{surge}^2 + S_{heave}^2}}\right) * \frac{180}{\boldsymbol{\pi}} \quad (2)$$

$$D_{surge} = x - S_{surge} \qquad D_{sway} = x - S_{sway} \qquad D_{heave} = x - S_{heave} \quad (3)$$

$$A_{surge} = SE(x[\tfrac{(i-v)}{(i+v)}]) \qquad A_{sway} = SE(y[\tfrac{(i-v)}{(i+v)}]) \qquad A_{heave} = SE(z[\tfrac{(i-v)}{(i+v)}]) \quad (4)$$

$$V_{pitch} = var(pitch[\tfrac{(i-v)}{(i+v)}]) \quad (5)$$

$$odba = |D_{surge}| + |D_{sway}| + |D_{heave}| \quad (6)$$

*Standard error= SE, **variance= var, ***number of seconds= v

Table 2. List of the variables calculated from the accelerometer data.

Variables	Labels	Definition	Unit
X axis Y axis Z axis	$surge$ $sway$ $heave$	Acceleration recorded from the accelerometer	g
Static acceleration	S_{surge} S_{sway} S_{heave}	Running mean of X, Y and Z axis	g
Dynamic acceleration	D_{surge} D_{sway} D_{heave}	Measure of effort	m/sec²
Amplitude	A_{surge} A_{sway} A_{heave}	Standard error of the signal	
Pitch	$pitch$	Vertical orientation of the accelerometer	degrees (°)
Pitch variance	V_{pitch}	Variance of the pitch	
Roll	$roll$	Lateral orientation of the accelerometer	degrees (°)
Overall Dynamic Body Acceleration	$odba$	Measure of the general effort across the three axes	m/sec²

Matching acceleration and behavior data

Since the camera clocks were set at Central Daylight Time (CDT) and GPS collars were in Coordinated Universal Time (UTC), we had to transform CDT time into UTC time by adding 5 hours.

We joined the acceleration and behavior data tables using the foverlaps function (‘data.table’ package) in R (with type = within and nomatch = NULL). It was essential to confirm that the timestamps of both the camera system and the accelerometer were the same in order to prevent incorrect matching. For this reason, for every hour we looked at a part of the video footage when each wolf was resting and then moving its head. Then using the pitch variable, we checked the exact time of this event in the acceleration data.

Using this method, we detected, a delay of 20 seconds between the camera system and the accelerometers. We could see in the camera footage that the wolf always moved its head 20 seconds earlier than in the pitch data (Appendix 5). This delay was consistent for all the 3 wolves and the whole 24-hour period, and was therefore corrected.

After correcting for the delay, we joined the two data tables again. Then, we examined the number of rows for each behavior before and after the data matching to verify how efficient the joining was (Appendix 6). It appeared that we lost more than 95% of behavioral data when we recorded them as “point event” but did not lose any behavioral data when they were recorded as “state event”. This was due to the fact that the "point event" was too precise a time to match exactly with the 32 Hz measurement of the accelerometer data.

Data analyses

R statistical program [R Core Team, v3.5.3] was used for the statistical computing (Appendix 11), and we did all the modeling using a computer with 128 Gb of RAM. We chose a powerful supervised machine learning approach, Random Forest, that can perform both regression and classification tasks. We first ran the Random Forest models using the 'random Forest' package. However, it was unable to handle our huge amount of data. Instead, we used the 'h2o' package that can process large datasets. It is a Java Virtual Machine optimized for doing "in memory" processing. Both of these Random Forest algorithms performs parallel learning by creating a forest with a given number of decision trees driven from a subset of the data.

For creating the models, we chose to select consistent and ecologically meaningful behaviors. Doing that, we had to group two behaviors, resting and freezing, into a new class called stationary. Indeed, the accelerometry signatures of those two behaviors were indistinguishable, with amplitudes and odba close to 0 as well as a varying pitch. Consequently, we selected 8 behaviors for our models: walking, trotting, galloping, stationary, howling, chewing, digging and smelling.

Before running the Random Forest algorithm, the dataset with our behaviors of interest, was split into two different parts: the training and the testing dataset. The training set is used to train the models and the testing set to estimate the performance of the model. Here we randomly selected 80% of the data for the training and 20% for the testing. This split is a commonly used division (Konstantinos, 2018 and Elsayad, 2010).

The efficiency of a Random Forest model depends on the number of trees and on the variables included. In general, the more trees in the forest the more robust the prediction is, yet at a certain number of trees the precision reaches a plateau and stops increasing. For our models, the plateau was reached at 1000 trees. To select variables of interest for the algorithm, biologically meaningful accelerometer variables should be considered (Table 2). Then, models with different sets of variables can be launched in order to pick variables that help the model the most. For our study n= 1000 trees and 7 variables were selected: pitch, odba, roll as well as pitch variance, surge amplitude, sway amplitude and heave amplitude calculated with a running mean of 30 seconds. The variables were centered and scaled before their implementation in the model.

After those preliminary steps, the final Random Forest models can be created (Fig.2). To begin, the algorithm will create a bootstrapped dataset randomly selected from the original training dataset. Then, using this bootstrap it builds n = 1000 decision trees and the data will be classified into the different behaviors given. To do that, at each node it will pick the best variable to split the decision into two daughter nodes. Then, the algorithm will combine all the decision trees and select the classification having the most votes from the trees in the forest. At the end, the Random Forest algorithm gives its final decision ranking.

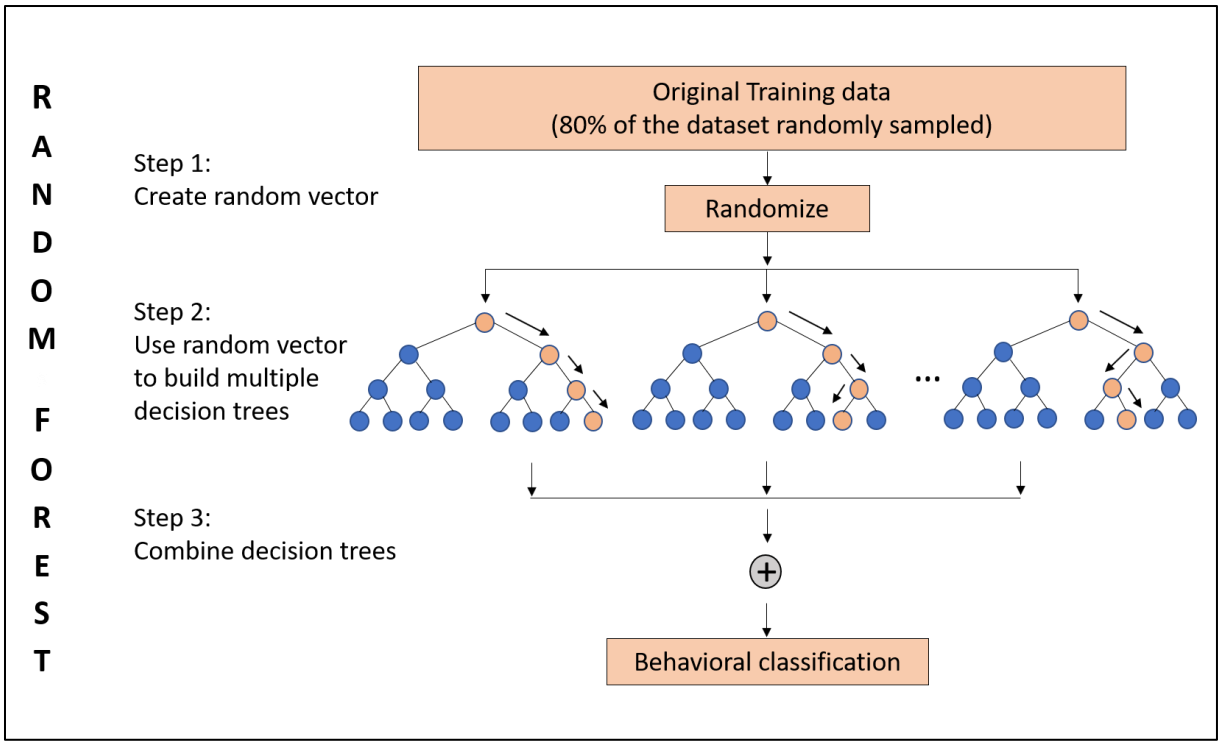


Figure 2. Schematic representation of the Random Forest algorithm.

After creating a good classification, the models need to be tested. For that, we applied the models on the test dataset composed of the remaining 20% of data.

To evaluate the efficiency of the prediction of Random Forest on the test dataset, we identified the true positive (TP), true negative (TN), false positive (FP) and false negative (FN). Using formulas from Bidder et al. (2014), we calculated the accuracy, precision and recall. 'Accuracy' (7) is defined as a measure of the overall proportion of correctly assigned data points, 'Precision' (8) is defined as the proportion of positive classifications that were correct and 'Recall' (9) is the proportion of data belonging to a behavioral category that were classified correctly as positive.

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \tag{7}$$

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

The log loss (logarithmic loss metric) was also extracted from the h2o model's results. It is a range that evaluates how close a model's predicted values are to the actual target value. Its range is between 0 and 1, with 0 meaning that the model is correctly assigning behaviors with a probability of 100% and 1 meaning that the model makes correct predictions assigning behaviors with a probability of 0%. Log loss is a classification loss function that quantifies the uncertainty of a classifier by penalizing false classification.

In total, we created four different models to test the classification of our eight behaviors of interest. One model for each individual wolf and another model with all three wolves combined.

Ethics statement

This protocol was reviewed and approved by the Institutional Animal Care and Use committee at the Wildlife Science Center in Minnesota.

Results

Determination of the wolf behaviors from triaxial accelerometry data

Using seven variables calculated from the acceleration data (pitch, roll, pitch variance, odba, surge amplitude, sway amplitude and heave amplitude), the Random Forest model algorithms were able to identify variation in acceleration signatures of all the eight selected wolf behaviors (walking, trotting, galloping, stationary, howling, chewing, digging and smelling; see Fig.3 and Appendix 7). These behaviors were extracted from video footage of the three captive wolves, named Keneli, Page and Chelsea, from one enclosure over a 24h period (see Mat & Meth and Table 1).

The distribution of the most important variables for the behavioral classifications are shown in Fig.3, and the remaining variables are shown in Appendix 7. The galloping behavior showed the highest variation in the surge (Appendix 7a), sway (Appendix 7b) and heave (Fig.3c) amplitudes as well as in the odba (Fig.3a). The second highest variation in those variables was for trotting. By contrast, stationary showed the lowest variation in those same variables. These variables represent differences in the intensity of behaviors. Since walking, smelling, digging, howling and chewing all have medium intensity, it was not possible to discriminate between them by using only the amplitude and odba variables, but required other parameters. For instance, howling was characterized by the highest positive angle of the pitch (Fig.3b) and smelling by the highest negative angle. Walking had similar acceleration signature as smelling, digging and chewing except for the pitch. Indeed, when the wolves were walking their head was most of the time at a 0° angle, while for the three other behaviors, wolves had their head mostly at a negative angle. Two behaviors, digging and chewing, seemed to overlap a lot. It makes sense ecologically, because both behaviors were associated with medium-intensity body movements and head down. The roll variable (Appendix 7d) showed individual variation of

lateral body movement between the three wolves. Individual variation was also found for chewing, digging and smelling in pitch variance (Appendix 7c) and in the amplitude variables. The wolf named Page showed the highest pitch variance for chewing, while Chelsea had the highest pitch variance for digging and smelling.

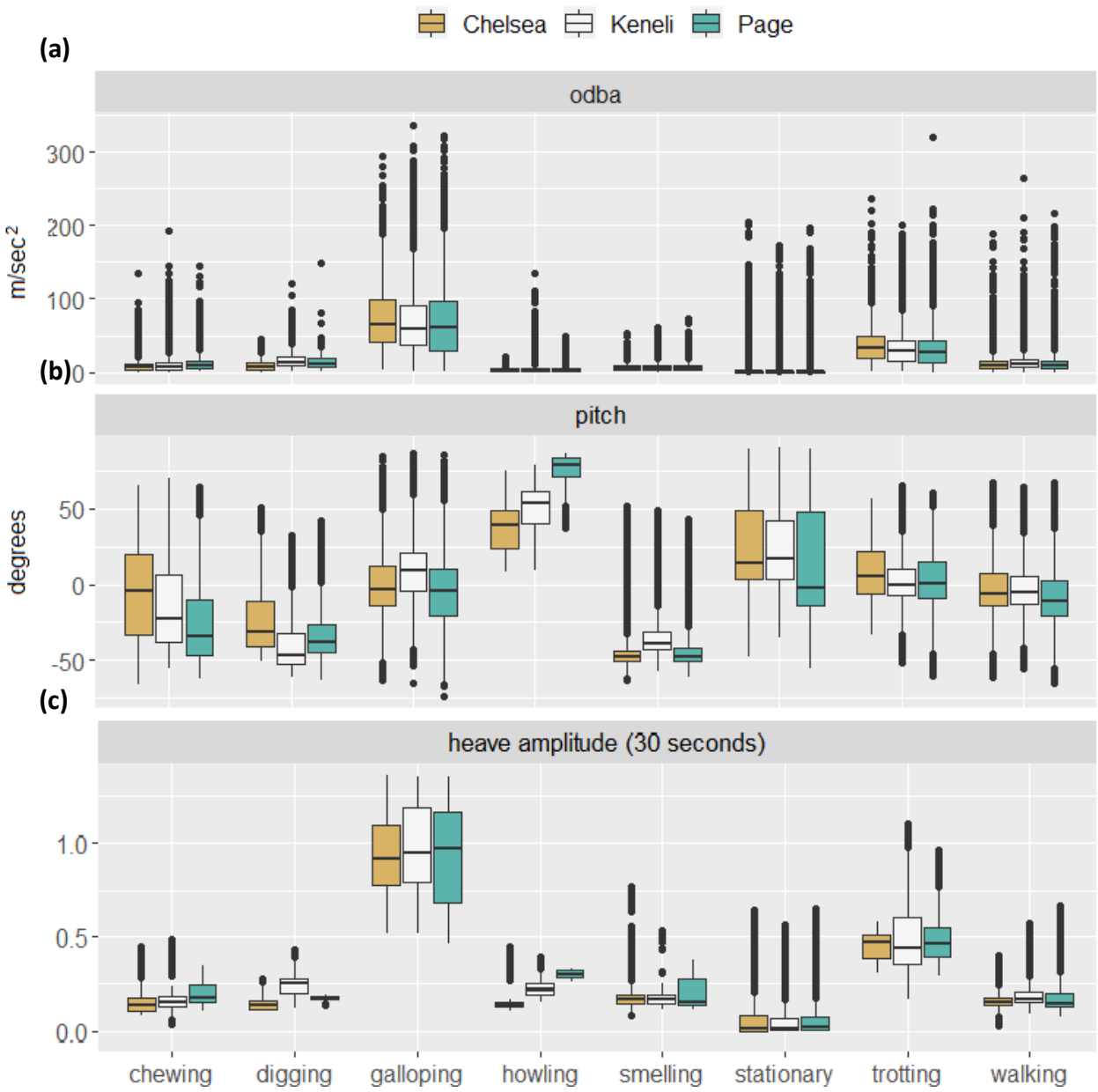


Figure 3. Acceleration signature for different behaviors of three captive wolves.
Example of **(a)** overall dynamic body acceleration (odba) , **(b)** pitch(vertical neck orientation) and **(c)** heave (dorso-ventral) amplitude with a running mean of 30 seconds signals for 8 behaviors. Orange, white and green colors were attributed to each wolf, Chelsea, Keneli and Page, respectively.

Classification of the behaviors of individual wolves

In this section, we show tables with results from Keneli as examples, whereas the same tables for Chelsea and Page can be found in Appendices.

For Keneli, our model with training data consisting of 80% of the original data set (Table 3) had a log loss of 0.00356 (Table 4), meaning that the model correctly classified behaviors with a very high probability.

Chelsea and Page (Appendix 8) had even better training models with a log loss of 0.00328 and 0.00284, respectively (Table 4).

For each wolf, most of the training misclassification were observed into the stationary or walking classes. This could be explained by the fact that when a wolf his smelling, chewing or howling it is also stationary. Moreover, if a wolf is walking too slowly, his acceleration range overlap more with the stationary behavior range.

Seeing how our models learned and misclassified some behavior in the test data sets, such as walking and stationary, we can already know that we will have future confusion with these behavioral classes.

Table 3. Confusion matrix, for the training data set, of observed (rows) and predicted (columns) behaviors from the wolf Keneli as categorized by the Random Forest model.

	Chewing	Digging	Galloping	Howling	Smelling	Stationary	Trotting	Walking
Chewing	**21972**	0	0	0	0	43	0	63
Digging	1	**3825**	0	0	0	0	0	0
Galloping	0	0	**5121**	0	0	0	0	0
Howling	0	0	0	**2469**	0	0	0	0
Smelling	0	0	0	0	**3202**	0	0	25
Stationary	1	0	0	0	0	**1119141**	0	29
Trotting	0	0	0	0	0	0	**5990**	12
Walking	2	0	0	0	0	47	1	**44568**

Table 4. Log loss (logarithmic loss metric) ranges measured on each confusion matrix for our four models.

Model	Log loss (training confusion matrix)	Log loss (test confusion matrix)
Keneli	0.00356	0.00105
Page	0.00284	0.00108
Chelsea	0.00328	0.00052
Three wolves grouped	0.00699	0.00258

*A perfect model would have a log loss of 0 and the log loss values would increase towards 1 as more of the predicted behaviors diverge from the observed behaviors.

When applying each model on the corresponding test data set, we obtained even lower log loss values. Chelsea (Appendix 9) got the best value (log loss = 0.00052) followed by Keneli (Table 5) and Page (Appendix 9) with a log loss of 0.00105 and 0.00108 respectively. These results clearly indicate that our algorithms learned very well with the training data sets.

Indeed, “galloping” was perfectly predicted (accuracy=1, precision=1 and recall=1) for all three wolves. “Smelling” was perfectly predicted for Keneli and Page’s models, as well as “trotting” for Keneli and Chelsea’s models, and “chewing” for Chelsea’s model (Table 6).

Test confusion matrices showed almost no misclassifications except with the class "walking" (Table 5). Indeed, this behavior showed the lower rate of precision for all three wolves (Table 6, Appendix 10), meaning that it had a smaller proportion of positive classifications that were correct. As an example, six of Keneli’s records of chewing were misclassified as walking. In addition, for each individual model on the test data sets, “walking” had also the lower rate of recall (Table 6, Appendix 10), indicating that for some observed records of walking, the model had predicted another behavioral class, such as “stationary”.

For every wolf, odba and pitch were the two principal variables that explained the most of the behaviors, followed by the amplitude variables, with the order varying between the individuals (Table 6, Appendix 10).

Table 5. Confusion matrix, on the test data set, of observed (rows) and predicted (columns) behaviors from the wolf Keneli as categorized by the Random Forest model.

	Chewing	Digging	Galloping	Howling	Smelling	Stationary	Trotting	Walking
Chewing	**4488**	0	0	0	0	0	0	6
Digging	0	**0**	0	0	0	0	0	0
Galloping	0	0	**4036**	0	0	0	0	0
Howling	0	0	0	**0**	0	0	0	0
Smelling	0	0	0	0	**570**	0	0	0
Stationary	0	0	0	0	0	**284535**	0	0
Trotting	0	0	0	0	0	0	**1007**	0
Walking	0	0	0	0	0	5	0	**6983**

Table 6. Performance and variables importance of Keneli Random Forest model.

Behaviors	Accuracy	Precision	Recall
Chewing	0.99998	1	0.99866
Digging	NA	NA	NA
Galloping	1	1	1
Howling	NA	NA	NA
Smelling	1	1	1
Stationary	0.99998	0.99998	1
Trotting	1	1	1
Walking	0.99996	0.99914	0.99928

Variable	Relative importance	Scaled importance	Percentage (%)
odba	28128168.0	1.00000	23.22
pitch	23991454.0	0.85293	19.80
heave amplitude	22961282.0	0.81631	18.95
surge amplitude	18806410.0	0.66859	15/52
sway amplitude	11685016.0	0.41542	9.65
pitch variance	10302194.0	0.36626	8.50
roll	5272154.0	0.18743	4.35

*Performance for each behavior (left table) and variables importance (right table) of the model.

Since some behaviors were rare, such as howling and digging, we had fewer data points for these and almost all of them went into the training set. Because of that, we did not have enough data for these behaviors in the test data sets. But we could expect at least the same high level of accuracy for them in with the test data sets as with the training data sets.

Individually, each model was remarkably precise at predicting the behavior of the individual wolves. Nevertheless, to build a model that could be applied on wild wolves, we needed to create a last model that included the individual variation of the behaviors by combining the three data sets from the three wolves. By doing that, we also increased the amount of data and should be able to obtain predictions for the less common behaviors.

Classification of behaviors of all wolves combined

We randomly selected 80% of the combined data from the three wolves for the training and left 20% for the test. This last model had a larger amount of data for each of the eight behaviors and included individual variation between the three wolves.

This model had the higher log loss range (Table 4) for the training results (log loss = 0.00699). Indeed, in the corresponding confusion matrix (Table 7) we observed more misclassifications, mainly for the stationary and the walking prediction classes.

Table 7. Confusion matrix, of the training data set, of observed (rows) and predicted (columns) behaviors of the three wolves combined, as categorized by the Random Forest model.

	Chewing	Digging	Galloping	Howling	Smelling	Stationary	Trotting	Walking
Chewing	**48716**	0	0	0	0	446	0	855
Digging	2	**6020**	0	0	0	1	0	28
Galloping	0	0	**10230**	0	0	0	0	0
Howling	0	0	0	**4579**	0	24	0	0
Smelling	5	0	0	0	**8598**	3	0	140
Stationary	62	0	0	0	0	**3167243**	7	460
Trotting	0	0	0	0	0	1	**33536**	23
Walking	6	0	0	0	2	387	40	**122349**

Then, applying the model on the test data set, we acquired a log loss of 0.00258 (Table 4). Even if this is a higher value compared to the individual models, we managed to make an algorithm that learned to classify behaviors from different wolves at a high resolution.

Moreover, adding more data to the model, we were now able to obtain predictions from the confusion matrix (Table 8) for all eight behaviors, including the less common ones. Looking at our model performances (Table 9), we were able to perfectly classify five behaviors; “digging”, “galloping”, “howling”, “smelling” and “trotting”, with accuracy, precision and recall = 1.

Here as well, walking showed the lowest precision rate (Table 9) with 90% of the misclassifications, mostly due to 135 actual chewing behaviors that were predicted as walking behaviors. This also explains why chewing had the lower recall rate (Table 9).

Nevertheless, the accuracy, precision and recall scores for “chewing”, “stationary” and “walking” were still remarkably high, > 0.98, 0.99994 and 0.991 respectively.

Table 8. Confusion matrix, of the test data set, of observed (rows) and predicted (columns) behaviors of the three wolves combined, as categorized by the Random Forest model.

	Chewing	Digging	Galloping	Howling	Smelling	Stationary	Trotting	Walking
Chewing	**8787**	0	0	0	0	12	0	135
Digging	0	**3743**	0	0	0	0	0	0
Galloping	0	0	**4789**	0	0	0	0	0
Howling	0	0	0	**1698**	0	0	0	0
Smelling	0	0	0	0	**803**	0	0	0
Stationary	0	0	0	0	0	**811247**	0	26
Trotting	0	0	0	0	0	0	**1216**	0
Walking	1	0	0	0	0	4	0	**18481**

Table 9. Performance and variables importance of the combined Random Forest model for the three wolves.

Behaviors	Accuracy	Precision	Recall
Chewing	0.99983	0.99989	0.98355
Digging	1	1	1
Galloping	1	1	1
Howling	1	1	1
Smelling	1	1	1
Stationary	0.99995	0.99998	0.99996
Trotting	1	1	1
Walking	0.99980	0.99136	0.99973

Variable	Relative importance	Scaled importance	Percentage (%)
odba	78642416.0	1.00000	25.23
surge amplitude	54563032.0	0.69381	17.50
heave amplitude	52540080.0	0.66809	16.85
pitch	41742704.0	0.53079	13.39
sway amplitude	35974072.0	0.45744	11.54
pitch variance	32338780.0	0.41121	10.37
roll	15944087.0	0.20274	5.11

*Performance for each behavior (left table) and variables importance (right table) of the model.

Also, for the combined model, the odba was the principal variable that contributed the most to building the decision trees. Surge and heave amplitude were second and third in the variables' importance hierarchy. This result changed from the individual models, where the pitch appeared in second place.

Generally, with this last model that pooled the data from the three captive wolves, we were able to automatically classify eight behaviors, with 100% accuracy for five of them and > 99.9% for the remaining ones.

Discussion

Our aim in this study was to use accelerometer measurement and video footage recorded from captive wolves to develop and validate a method to automatically identify specific wolf behaviors. This method can then be used to classify behaviors in wild wolves.

Our results show that eight different behaviors from captive wolves could indeed be successfully predicted automatically by using a Random Forest algorithm, with a log loss between 0.00052 and 0.00258 for our four models (Table 4).

Video recording

Using four surveillance cameras to document behaviors was a practical method to calibrate accelerometers on captive wolves, and infra-red LEDs made it possible to film at night. However, individual animal tags (dyed fur) and body movements were not easy to see with the infra-red light, sometimes making it impossible to identify the wolf or its behavior. So, during night-time, we recorded fewer behaviors to avoid wolf and behaviors mis-identifications.

We also had a decrease in our behavior data collection due to cameras clocks resetting (wolves chewing cables and batteries running out of power during very cold nights). A better installation of the system for controlling the camera is advised to avoid that problem.

Our method for recording behaviors of short duration could be improved. We showed that recording point behaviors as a single point in time was not a good method (see Appendix 6). Indeed, the timestamp was too precise to match exactly with the accelerometer data points, even at 32 Hz. This percentage of lost data can only increase if we lower the acceleration sampling frequency. Hence, in the future, point behaviors will be recorded as short periods of time rather than single points.

Linking behavior with acceleration variables

There is no common solution for extracting specific behaviors from acceleration data, and several different approaches have been used previously (Nathan et al., 2012). This is the same for the selection of variables to build models. For instance, dynamic acceleration and pitch are commonly used (Shepard et al., 2008), but other studies have chosen to only use the raw data (x, y and z axis) (Sha et al., 2017). The choice of variables should be relevant to the study species, the placement of the accelerometer on the body of the animal, and to the research question. Other non-accelerometric variables can also be added to increase model performance, such as conductivity (Pagano et al., 2017), body temperature (Hetem et al., 2019) or depth (Chimienti et al., 2016).

In our study, the overall dynamic body acceleration (odba) appeared to be the most important predictor for our models. Then, other variables that contributed the most to building the decision tree variated between pitch and amplitude variables, depending on the model.

We also need to keep in mind that what we are able to register as separate behaviors from cameras footage is slightly different from what acceleration variables can identify. Therefore, being more general and grouping behaviors with similar accelerometry signature is important.

Limitations in prediction of behaviors

One of the benefits of a supervised learning approach is that behavioral classes are exactly defined, thus making their interpretation relatively straightforward. However, it should be remembered that applying those prediction models on free-ranging animals could be challenging, especially if looking at very specific and inconsistent behaviors.

The first limitation in identifying behaviors with accelerometry is when animals are performing multiple behaviors at the same time, such as walking and smelling. Because of these overlapping signals, models will have difficulties classifying them into one behavioral category. When recording behaviors from the video footage, we tried as much as possible to avoid recording those multiple behaviors. But when the models will be applied on other accelerometry data, such as data from wild wolves, they will encounter them, which may significantly lower the accuracy and precision of the predictions.

Moreover, acceleration signatures from specific behaviors can be influenced by external factors. For example, the terrain can alter posture and body movement. Indeed, we can expect a different accelerometer signal when a wolf is walking in a hilly area than when it is walking in a flat area. Yet, in our study, the ground in the enclosure was flat, so terrain variations were not included in our models.

In addition, at the time of data collection, the Wildlife Science Center in Minnesota had at most only a few centimeters of snow, so wolf movements were not affected. However, in Scandinavia, deep snow can affect the locomotion of wild wolves for several months of the year. These parameters would have to be tested to accurately transfer the model prediction from captive to wild wolves. For instance, an unsupervised model can be used to categorize those locomotion behaviors, which can then be compared with the signatures found from the supervised models on captive wolves, to highlight acceleration variations in the deep snow.

In this study, we have analyzed eight behaviors, but shown in the ethogram (Appendix 2), there are many more behaviors performed by the wolves. All of these could overlap with our eight selected behaviors classes. Indeed, we can expect that shaking may be classified as chewing or digging and jumping as galloping, since they may have similar accelerometric signatures. In other words, if we applied our models on wild data, they could predict howling behavior, but the wolf could have just stared at the stars or having any other behavior with a stationary position and the head up.

Since the wolves were captive, we could not record hunting events. So, using our models alone, we cannot identify this ecologically interesting behavior. Nevertheless, we can expect that a hunting event will be associated with a set of individually identifiable behaviors (such as: “smelling” following by “galloping”, “jumping” then “chewing” and possibly “digging”).

Our models were based on a limited number of individuals (n = 3). We may have missed different individual behavioral variation due to this small number of wolves studied.
Given that, our model accuracy can be slightly different when applying on data from wild wolves, and interpretation of behaviors will be more challenging.

Finally, computer memory is a limiting factor for large-scale accelerometry analyses. Indeed, in order to run these models created with a large amount of data, we needed a computer with a high memory power. This limitation will also occur during the future processing and predictions with even larger data sets from wild wolves. It is advisable to either purchase a powerful computer or to rely on high computing clusters developed by universities to be able to handle and analyze accelerometer data.

Perspectives

Because of time limitations, it was not possible to analyze the remaining part of the acceleration and video datasets. However, we plan to apply the R code and analysis developed in this study to the remaining part of the dataset so to build more robust models. Doing that, we will add more wolves and enclosures variation in the models.

Here we used acceleration data recorded at 32 Hz. The high frequency allowed us to identify behaviors with a very high accuracy, but it was also producing a large amount of data, thus becoming a challenge for data processing. A continuation of this study should be to down-sample our acceleration data to 16, 8, 4 and 2 Hz and compare how robustly we can classify behaviors at these reduced sample rates. It will then be possible to choose the frequency that optimizes our data handling while still having a good classification accuracy.

Animal ecology is affected by energy costs, and because metabolism is closely connected to mechanical work, accelerometer data have the potential to be a proxy providing detailed information on energy expenditure. Indeed, studies have suggested that the overall dynamic body acceleration can be used to estimate the behavior-specific rate of energy expenditure of free-living animals (Wilson et al., 2006; Elliott et al., 201). Moreover, as doubly labelled water and heart rate recording have been applied on our wolves, both methods known to provide information about energy costs (Schoeller et al., 1986 and Ceesay et al., 1989), a comparison of those results can be made in order to get an estimate for the metabolic costs of different behaviors in wolves.

Actually, cluster analyses of telemetry data are currently being used to identify likely predation event in wild wolves (Sand et al.,2005). In order to reliably identify either kill sites or resting sites, clusters of GPS position have to be examined in the field. Yet, combining telemetry and accelerometry data would make it possible to separate between these two types of GPS clusters by looking at the acceleration signatures in the period of arriving at the cluster (expected to include galloping and possibly jumping for predation events) and while the wolf is at the cluster (expected to include e.g. chewing at kill sites, and a larger proportion of stationary behavior at resting sites). This difference could be validated by combining acceleration data with traditional field studies.

Moreover, unsupervised machine learning can also be used on data from wild wolves to see which behavioral categories can be identified and a comparison of both supervised and unsupervised method can be done (Leos-Barajas et al., 2017).

Finally, our grouped model can be used to identify these eight behaviors in free-ranging wolves, e.g. during experimental approaches by humans (ongoing master's study by Erik Versluijs) and when wolves move close to anthropogenic structures. The purpose will be to combine animal location and behaviors using GPS and accelerometer data from collars. Such information will be extremely valuable for conservation and management issues by revealing the detailed responses of wolves to their surrounding landscapes.

The behavioral data bank extracted from the video footage will also be used in an ongoing master's study by Tine S. Johansen, relating social behavioral interaction and heart rate, as well as connecting behaviors and heart rate to the presence of humans or dogs.

Conclusion

Accelerometry data are a powerful tool for studying animal behavior remotely. By using this approach on captive wolves, we were able to categorize a set of 8 different behaviors: walking, trotting, galloping, stationary, howling, chewing, digging and smelling.

Using Random Forest models, we validated the use of triaxial accelerometer data to identify locomotion patterns and more complex behaviors of captive wolves with remarkably high accuracy, obtaining individual and group models a log loss (logarithmic loss metric) with a range very close to 0 (< 0.003) when tested on our data. Our log loss ranges indicated that our models are correctly assigning behaviors with a probability near 100%.

With a grouped model that combined the data from the three wolves, we demonstrated that we were able to classify five behaviors perfectly: “digging”, “galloping”, “howling”, “smelling” and “trotting”.

Moreover, for the individual and the grouped models, the overall dynamic body acceleration appeared to be the principal variable that contributed the most to building our decision trees. Then, pitch and the amplitude variables followed in varying order of importance, depending on the model.

Finally, studies using accelerometers for identifying such a detailed suite of behaviors with this level of accuracy are rare, and this study is one of few using acceleration data to classify specific behaviors in large carnivores (Rekvik, 2015; English, 2018; Pagano et al., 2017 and Wang et al., 2015).

References

Ceesay et al. (1989) The use of heart rate monitoring in the estimation of energy expenditure : a validation study using indirect whole-body calorimetry. British Journal of Nutrition 61:175-186.

Chapron et al. (2014) Recovery of large carnivores in Europe's modern human-dominated landscapes. Science 346(6216):1517–1519.

Chimienti et al. (2016) The use of an unsupervised learning approach for characterizing latent behaviors in accelerometer data. Ecology and Evolution 6(3):727-741

Bidder et al. (2014) Love thy neighbour: automatic animal behavioural classification of acceleration data using the k-nearest neighbour algorithm. PLoS ONE 9(2): e88609.

Brown et al. (2013) Observing the unwatchable through acceleration logging of animal behavior. Animal Biotelemetry 1:1-16.

Doherty and Driscoll (2018) Coupling movement and landscape ecology for animal conservation in production landscapes. The Royal Society 285(1870).

Elliott et al. (2013) Accelerometry predicts daily energy expenditure in a bird with high activity levels. The Royal Society 9(1).

Elsayad (2010) Classification of ECG arrhythmia using learning vector quantization neural networks. IEEE :139-144.

English (2018) Behavior, energetics and movement through biologging in the canidae. Swansea University 1-78.

Eriksen et al (2011) Activity patterns of predator and prey: a simultaneous study of GPS-collared wolves and moose. Animal behavior 81(2):423-431.

Graf et al. (2015) The Use of Acceleration to Code for Animal Behaviours; A Case Study in Free-Ranging Eurasian Beavers Castor fiber. PLoS ONE 10(8): e0136751.

Halsey et al. (2008) Acceleration versus heart rate for estimating energy expenditure and speed during locomotion in animals: Tests with an easy model species, Homo sapiens. Zoology 111(3):231-241.

Hetem et al. (2019) Body temperature, activity patterns, and hunting in freeliving cheetah: biologging reveals new insights. Integrative Zoology 14(1):30-47.

Kreeger et al. (1986) Xylazine hydrochloride–ketamine hydrochloride immobilization of wolves and its antagonism by tolazoline hydrochloride. Journal of Wildlife Diseases 22(3):397-402.

Konstantinos (2018) Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture 145:311-318.

Leos-Barajas et al. (2017) Analysis of animal accelerometer data using hidden Markov model. Methods in Ecology and Evolution 8:161-173.

Liberg et al. (2011) Illegal killing of wolves in Scandinavia 1998–2011: variation in space and time. A report to WWF, Sweden 1-24.

Mattisson et al. (2013) Home range size variation in a recovering wolf population: evaluating the effect of environmental, demographic, and social factors. Oecologia 173(3):813-825.

Milleret et al. (2019) Testing the influcence of habitat experienced during the natal phase on habitat selection later in life in Scandinavian wolves. Scientific Reports 9:6526.

Nathan et al. (2012) Using tri-axial acceleration data to identify behavioural modes of free-ranging animals: general concepts and tools illustrated for griffon vultures. The Journal of Experimental Biology 215:986-996.

Pagano et al. (2017) Using tri-axial accelerometers to identify wild polar bear behaviors. Endangered species research 32:19-33.

Resheff,et al (2014) AcceleRater: a web application for supervised learning of behavioral modes from acceleration measurements. Movement Ecology 2:27.

Rekvik (2015) Identification of five brown bear (Ursus arctos) behaviors using-tri-axial accelerometers. Telemark University College 1-33.

Rutz and Hays (2009) New frontiers in biologging science. Royal Society 5(3):289-292.

Sand et al. (2005) Using GPS technology and GIS cluster analyses to estimate kill rates in wolf-ungulate ecosystems. Wildlife Society Bulletin 33(3):914-925.

Sand et al. (2008) Summer kill rates and predation pattern in a wolf-moose system: can we rely on winter estimates? Oecologia 156(1):53-64.

Sanz-Pérez et al. (2018) No place like home ? A test of the natal habitat-biased dispersal hypothesis in Scandinavian wolves. The Royal Society 5(12).

Schoeller et al. (1986) Energy expenditure by doubly labeled water: validation in humans and proposed calculation. The American Physiological Society 250(5):823-830

Shal et al. (2017) Estimating activity of Japanese macaques (Macaca fuscata) using accelerometers. American Journal of Primatology 79(10):e22694.

Skogen and Krange (2003) A Wolf at the Gate: The Anti-Carnivore Alliance and the Symbolic Construction of Community. Sociologia Ruralis 43(3):309-325.

Treves and Karanth (2003) Human-Carnivore Conflict and Perspectives on Carnivore Management Worldwide. Conservation Biology 17(6):1491-1499.

Wang et al. (2015) Movement, resting, and attack behaviors of wild pumas are revealed by tri-axial accelerometer measurements. Movement Ecology 3:2.

Wikenros et al. (2010)Competition between recolonizing wolves and resident lynx in Sweden. NRC Research Press 88(3):271-279.

Wilson et al. (2006) Moving towards acceleration for estimates of activity-specific metabolic rate in free-living animals: the case of the cormorant. Journal of Animal Ecology 75:1081-1090.

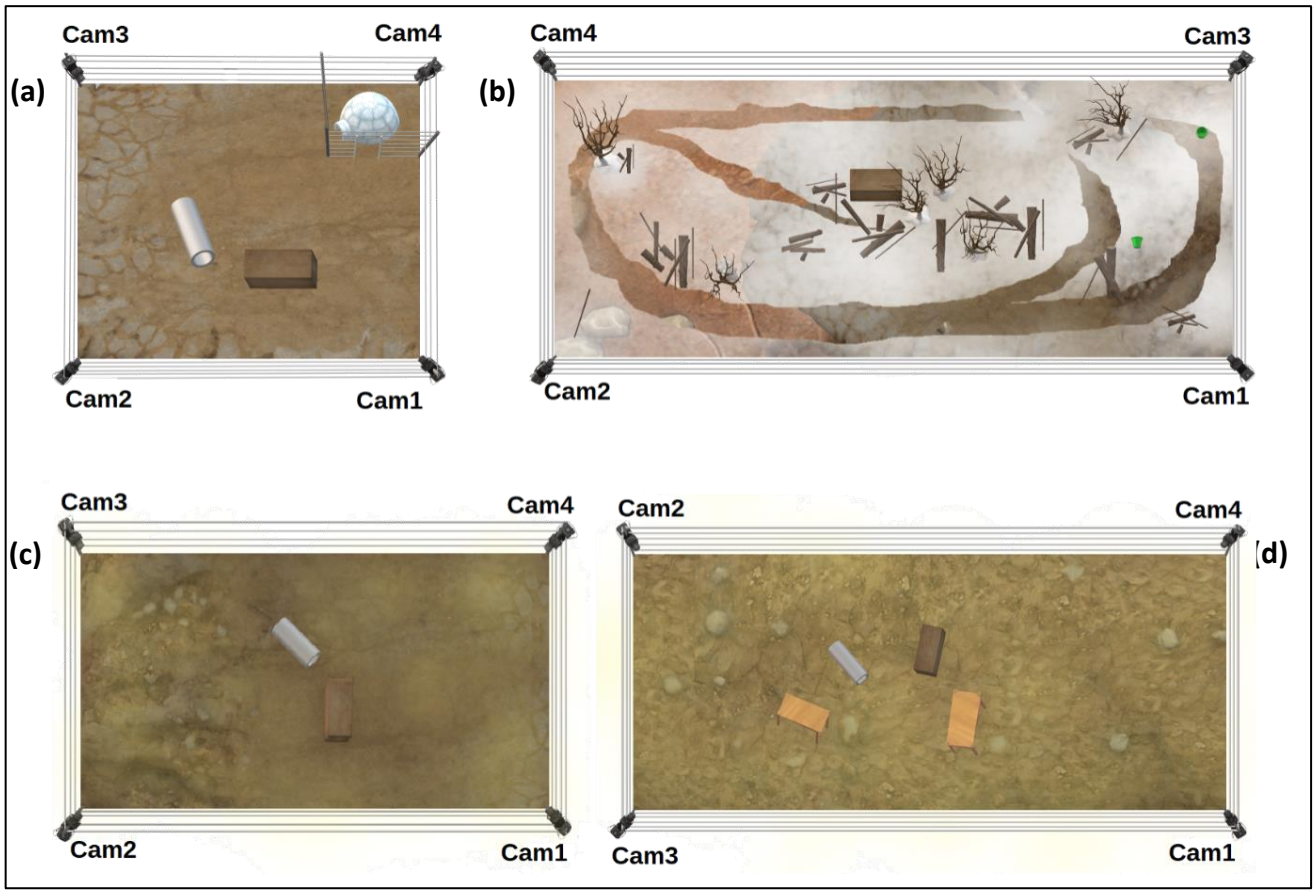
Wilson et al. (2008) Prying into the intimate details of animal lives: use of a daily diary on animals. Endangered species research 4:123-137.

Wilson et al. (2013) Locomotion dynamics of hunting in wild cheetahs. Nature 498:185-189.

Yoda et al. (1999) Precise monitoring of porpoising behaviour of adélie penguins determined using acceleration data loggers. The Journal of Experimental Biology 202:3121-3126.

Zimmermann et al. (2015) Predator-dependent functional response in wolves: from food limitation to surplus killing. Journal of Animal Ecology 84:102-112.

Appendix



Appendix 1. Schematic representation of enclosure and behavior recording device.

For each enclosure, 4 cameras (Cromorc wireless 1.3 MP security camera system) called here Cam1, Cam2, Cam3 or Cam4 were hung to record the behavior of wolves, **(a)** enclosure 1, 5 wolves, **(b)** enclosure 3, 7 wolves, **(c)** enclosure 2, 8 wolves and (**d**) enclosure 4, 7 wolves.

SKANDULV 2018-01-04 Ane and Kristoffer

IMMOBILISATION – CAPTURE PROTOCOL

WOLF | YYYY MM DD 2018 - 10 - 15 | SHEET

OLD COLLAR REMOVED ○YES ✓NO ○IMP removed

NEW TRANSMITTER ✓COLLAR ○IMPLANT. ○EAR MARK

CAPTURE TEAM

ID # 558 Keneli | SEX M | ✓Pack – # wolves: 5 ○ Pair ○Single

BORN YEAR 2017 | AGE 1.5 yr | ENCLOSURE/TERRITORY Temp Nap

TYPE | SERIAL#. 20423

FREQ | MAGNET REM. ✓YES

VHF TESTED AFTER DEPLOYMENT ○YES ○NO

TIME	
10:40	START "HUNT"
	LAYS DOWN

FUNCTION (active/passive/mort.)

LONGEVITY (months)

COLLAR LENGTH (cm) 59 cm Width: 4 cm

COLLAR MATERIAL

COLLAR COLOR

MICRO CHIP ○ NEW ○ READ

CHIP PLACEMENT

MICRO CHIP #

TIME	DRUG	DOSE	ADMIN

HAIR SAMPLE (number)

TISSUE SAMPLE (ear, number)

BLOOD SAMPLE (serum, number)

BLOOD (edta, number)

FECES (rectum, number)

EAR SWAB

RECTUM SWAB

TIME		TIME	TEMP
	STANDS UP		
	STEADY		
	LEAVES		

WEIGHT (kg) 85 lb

HEAD CIRCUMFERENCE (cm)

EAR LENGHT

TIME	HR	TIME	RESP

NECK CIRCUMFERENCE (cm) 46,5 cm

CHEST CIRCUMFERENCE (cm)

BODY LENGTH (cm) 91 cm

TAIL LENGHT (cm)

HEAD LENGTH (mm) 31 cm

RIGHT ELBOW-HAND (mm) 29+19 cm

L 29+19 cm

RIGHT KNEE-HEEL 32+26 cm

L 31+26 cm

COMMENTS Dye: Both sides of chest/ribcage

LARGEST NIPPLE (HxW, mm)

NIPPLES WITH MILK ○ YES ○ NO

VULVA SWOLLEN ○ YES ○ NO

VET EVALUATION:

TEETH, AGE: | DENTAL FORMULA:

COMMENTS

R TESTICLE (LxW, mm)

L TESTICLE (LxW, mm)

TANDSTATUS OBSERVERAD FRAMMIFRÅN

DENTAL STATUS, OBSERVED FROM THE FRONT

MARK INJURIES WITH DARK COLOR

TANDSKADOR SVÄRTAS MARKERA MJÖLKTÄNDER

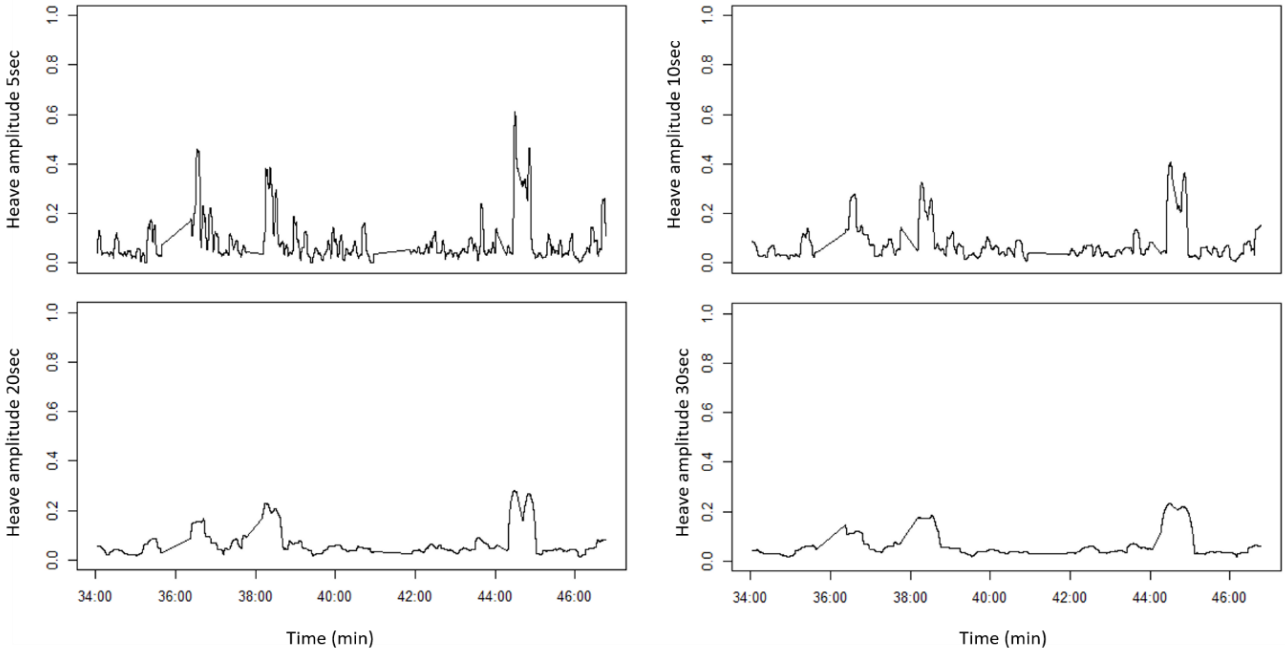
Appendix 2. Wolf measurement during immobilization protocol.

Appendix 3. Ethogram.

	Behavior	Record types	Description
GROUP 1	Walking	STATE EVENT	One leg lifted at a time with the other on the ground
	Trotting	STATE EVENT	Animal is moving with diagonally positioned front and back legs raised at the same time
	Galloping	STATE EVENT	Full-on running (all four legs off the ground)
	Chewing	STATE EVENT	Action of eating a food or chewing branches
	Resting	STATE EVENT	In a resting position/state on the ground (siting or sleeping)
	Grooming themselves	STATE EVENT	Licking or scratching his fur or body part
	Digging	STATE EVENT	Animal is using his front paws for digging in soil
	Scratching	STATE EVENT	Scratching the surface of a vertical structure with one paw
	Smelling	STATE EVENT	Animal inhales with his nose air, ground, objects or someone
	Stretching	STATE EVENT	Paws are straight ahead, rump is raised high, and tail is raised or back is straight. Then the position is reversed so that the fore legs and head are raised.
	Howling	STATE EVENT	Standing, sitting or lying with head titled back, ears flat back on head and making an O with mouth
	Climbing	STATE EVENT	Action of getting up on the hind legs while front paws are up against something
	Freezing	STATE EVENT	Freezing or paused position, animal suddenly becoming rigid or motionless
	Rolling	STATE EVENT	Body contact with the ground, vigorous rolling and wriggling movement of whole body
	Bowing	STATE EVENT	Action of stretching its front legs out in front, leaning down on its elbows. Playful bow
	Standing	STATE EVENT	Animal standing on four legs
	Lay down	POINT EVENT	Action of lying down. Animal goes from a standing posture to a lying posture on the ground
	Get up	POINT EVENT	Action of standing up. Animal goes from a lying posture on the ground to a standing posture (relax)
	Shaking	POINT EVENT	Quick sudden movement of the head or the body
	Jumping	POINT EVENT	Body and legs stop touching the ground
	Turning left	POINT EVENT	Animal making a turn of 90° or more in left direction
	Turning right	POINT EVENT	Animal making a turn of 90° or more in right direction
	Head raise	POINT EVENT	Head lifted over the shoulder (in combination with other behaviors)
	Head down	POINT EVENT	Head towards ground, under the level of the shoulder (in combination with other behaviors)
	Sitting down	POINT EVENT	Animal goes from a standing or lying posture to a sitting posture on the ground. The forelegs are stretched straight under the body.
	Chasing	STATE EVENT	Running after another animal, performing pursuit event
	Grooming other	STATE EVENT	Licking or scratching fur or body part of another individual
	Other interaction	STATE EVENT	Body contact or standing close with another individual
	Pack behavior	STATE EVENT	Running in pack

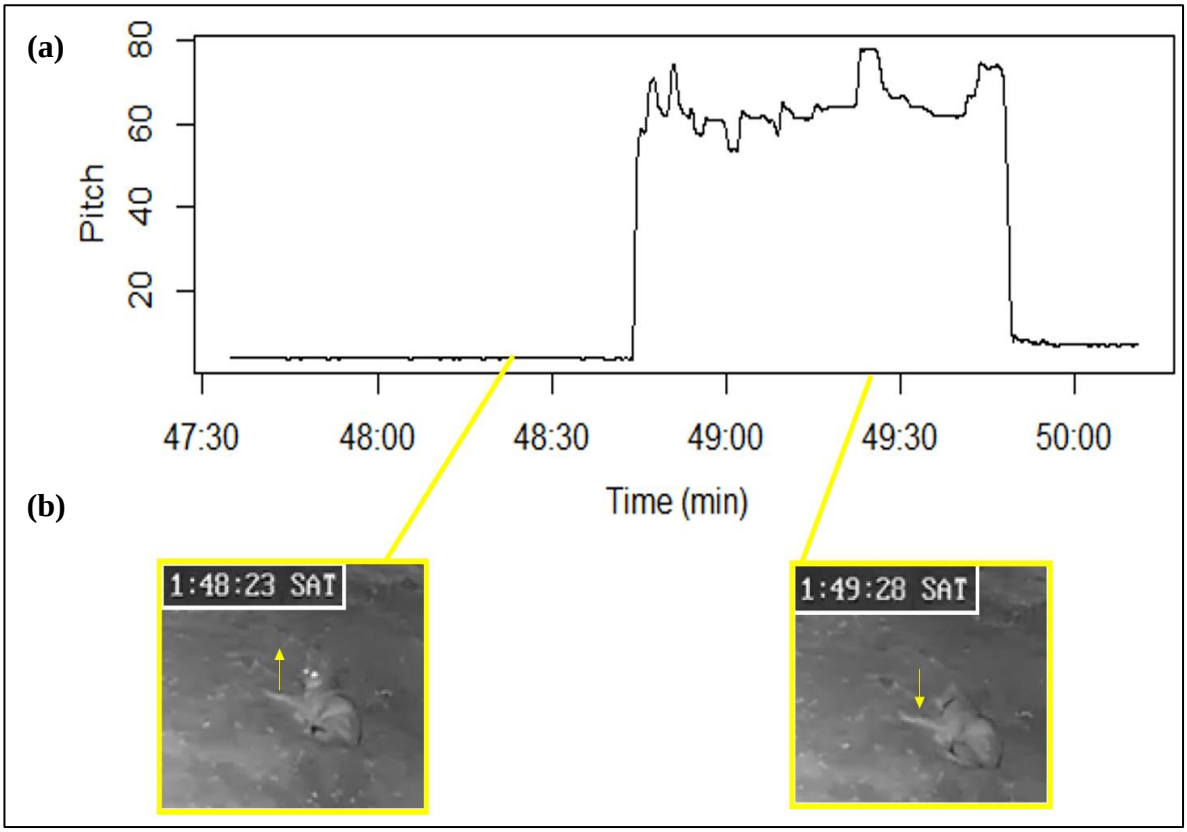
GROUP 2	Mounting	STATE EVENT	Standing of hind legs while placing forelimb on back of a mate
	Crouching	STATE EVENT	Animal with a crouching posture with curled down rump and tail tucked and/or wagging, ears lowered
	Pacing	STATE EVENT	Rapid movement, changes directions often, for 3 or more repetition
	Marking	STATE EVENT	Kicking dirt after urinating or defecation
	Urinating	STATE EVENT	Animal is urinating episodically by lifting a hind leg while semi-squatting. Defecation are also included.
	Tail wag	STATE EVENT	Consistent movement of tail from side to side
	Exposing belly	STATE EVENT	Rolling over onto the back and presenting his belly
	Awareness	POINT EVENT	Look and ears oriented towards the outside of the fence
	Startled	POINT EVENT	Sudden movement, then alert

*Group 1 is composed by behaviors for this study and accelerometry data, and group 2 by behaviors for Tine S. Johansen‘s study.



Appendix 4. Heave amplitude with 4 running means thought time.

Heave amplitude over a running of 5 seconds (top left), 10 seconds (right left), 20 seconds (bottom left) and 30 seconds (bottom right) through time (min). The consistency of heave amplitude movements was highlighted over different time windows.



Appendix 5. Illustration of time delay between camera and accelerometer after the first data tables join.
(a) Evolution of the pitch variable through time for Keneli. **(b)** Screen captures from camera footage of Keneli head movements. Yellow arrow represents head movement of the wolf and time clock is written as "Hour : Minute : Second" format.

(a)

Record type	Behavior	Number of rows before matching	Number of rows after matching	% of loss
STATE EVENT	Walking	55773	55773	0
	Trotting	7503	7503	0
	Galloping	6401	6401	0
	Chewing	27598	27598	0
	Stationary	1398964	1398964	0
	Grooming themselves	796	796	0
	Digging	4783	4783	0
	Scratching	0	0	0
	Smelling	4034	4034	0
	Stretching	1563	1563	0
	Howling	3086	3086	0
	Climbing	545	545	0
	Rolling	0	0	0
	Bowing	104	104	0
POINT EVENT	Lay down	115	0	100
	Get up	101	3	97
	Shaking	34	1	97,1
	Jumping	10	1	90
	Turning left	79	0	100
	Turning right	57	2	96,5
	Head raise	211	9	95,7
	Head down	159	6	96,2
	Standing	222	5	97,7
	Sitting down	19	1	94,7

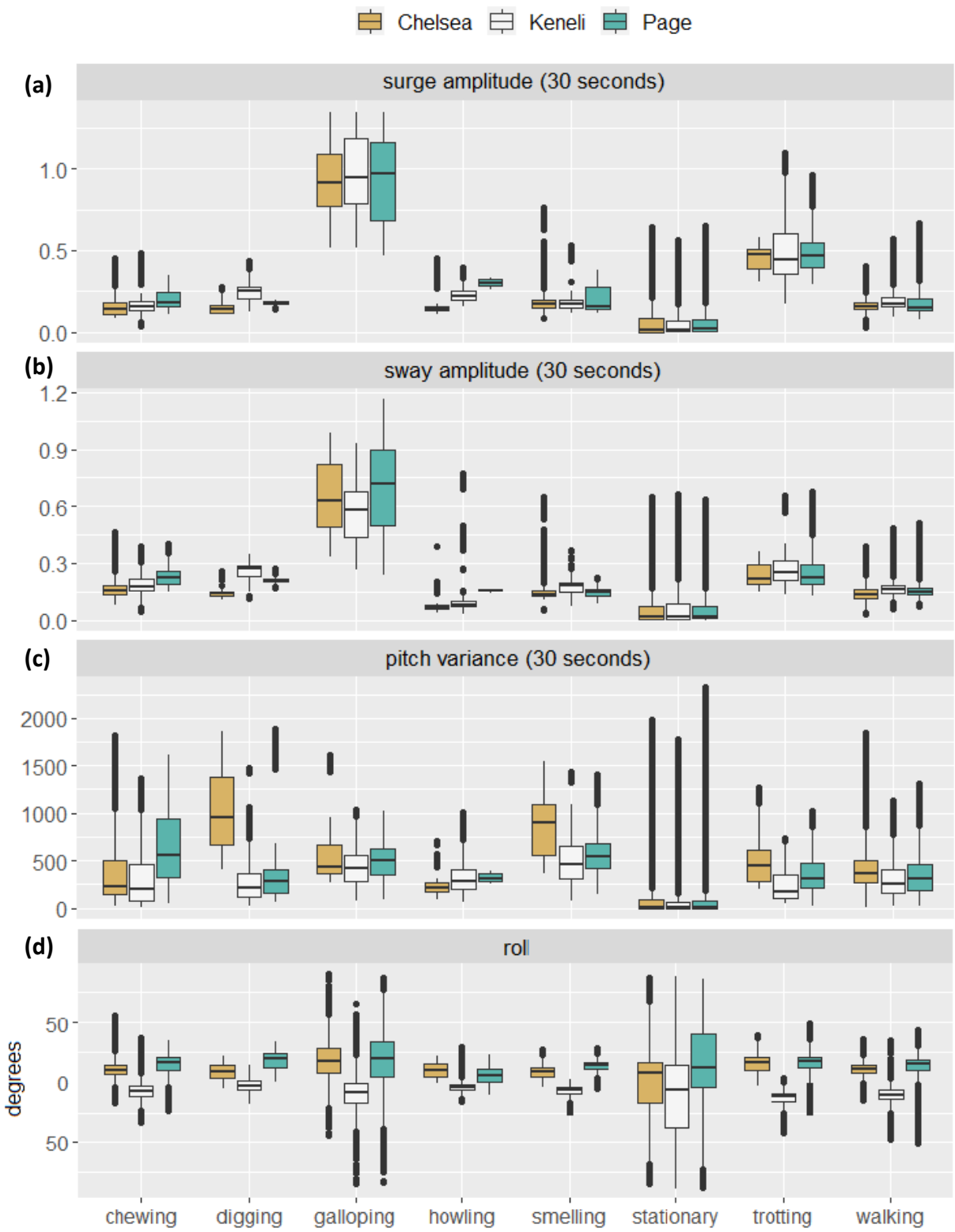
(b)

Record type	Behavior	Number of rows before matching	Number of rows after matching	% of loss
STATE EVENT	Walking	30149	30149	0
	Trotting	3142	3142	0
	Galloping	2528	2528	0
	Chewing	22842	22842	0
	Stationary	1443984	1443984	0
	Grooming themselves	395	395	0
	Digging	1756	1756	0
	Scratching	0	0	0
	Smelling	3229	3229	0
	Stretching	805	805	0
	Howling	2052	2052	0
	Climbing	108	108	0
	Rolling	0	0	0
	Bowing	0	0	0
POINT EVENT	Lay down	189	9	95,2
	Get up	157	13	91,7
	Shaking	49	2	95,9
	Jumping	37	1	97,2
	Turning left	40	2	95
	Turning right	27	2	92,5
	Head raise	309	6	98,3
	Head down	186	0	100
	Standing	275	10	96,3
	Sitting down	27	3	88,8

(c)

Record type	Behavior	Number of rows before matching	Number of rows after matching	% of loss
STATE EVENT	Walking	67558	67558	0
	Trotting	31305	31305	0
	Galloping	3859	3859	0
	Chewing	12081	12081	0
	Stationary	1116767	1116767	0
	Grooming themselves	439	439	0
	Digging	1025	1025	0
	Scratching	0	0	0
	Smelling	3670	3670	0
	Stretching	528	528	0
	Howling	616	616	0
	Climbing	286	286	0
	Rolling	0	0	0
	Bowing	0	0	0
POINT EVENT	Lay down	108	3	97,2
	Get up	89	3	96,6
	Shaking	43	1	97,6
	Jumping	27	0	100
	Turning left	71	3	95,8
	Turning right	63	3	95,2
	Head raise	338	16	95,3
	Head down	170	7	95,9
	Standing	268	12	95,5
	Sitting down	5	0	100

Appendix 6. Summary tables of rows numbers before and after matching behaviors and accelerations data tables.
(a) Keneli**,** with a mean of 96.5% of lost for POINT EVENT behaviors**, (b)** Chelsea, with a mean of 95,1% of lost for POINT EVENT behaviors and **(c)** Page, with a mean of 96.9% of lost for POINT EVENT behaviors.



Appendix 7. Acceleration signature for different behaviors of three captive wolves. Example of **(a)** surge amplitude, **(b)** sway amplitude **(c)** pitch variance, all three with a running mean of 30 seconds and (d) roll signals for 8 behaviors. Colors were attributed for each wolf.

Appendix 8. Confusion matrix, of the training data set, of observed (rows) and predicted (columns) behaviors of Page and Chelsea, as categorized by the Random Forest model.

	Chewing	Digging	Galloping	Howling	Smelling	Stationary	Trotting	Walking
Chewing	**9660**	0	0	0	0	2	0	3
Digging	0	**820**	0	0	0	0	0	0
Galloping	0	0	**3087**	0	0	0	0	0
Howling	0	0	0	**493**	0	0	0	0
Smelling	0	0	0	0	**2915**	0	0	21
Stationary	1	0	0	0	0	**893407**	1	5
Trotting	0	0	0	0	0	0	**25042**	2
Walking	0	0	0	0	0	34	34	**53978**

	Chewing	Digging	Galloping	Howling	Smelling	Stationary	Trotting	Walking
Chewing	**18150**	0	0	0	0	119	0	5
Digging	1	**1404**	0	0	0	0	0	0
Galloping	0	0	**2022**	0	0	0	0	0
Howling	0	0	0	**1630**	0	12	0	0
Smelling	1	0	0	0	**2582**	0	0	0
Stationary	14	0	0	0	0	**1155168**	1	4
Trotting	0	0	0	0	0	1	**2513**	40
Walking	2	0	0	0	1	53	0	**24063**

*Page's confusion matrix (up) with a log loss=0.00284 and Chelsea's confusion matrix (bottom) with a log loss=0.00328.

Appendix 9. Confusion matrix, of the test data set, of observed (rows) and predicted (columns) behaviors of Page and Chelsea, as categorized by the Random Forest model.

	Chewing	Digging	Galloping	Howling	Smelling	Stationary	Trotting	Walking
Chewing	**0**	0	0	0	0	0	0	0
Digging	0	**0**	0	0	0	0	0	0
Galloping	0	0	**1978**	0	0	0	0	0
Howling	0	0	0	**0**	0	0	0	0
Smelling	0	0	0	0	**960**	0	0	0
Stationary	0	0	0	0	0	**227722**	0	0
Trotting	0	0	0	0	0	0	**7460**	0
Walking	0	0	0	0	0	0	1	**9255**

	Chewing	Digging	Galloping	Howling	Smelling	Stationary	Trotting	Walking
Chewing	**256**	0	0	0	0	0	0	0
Digging	0	**0**	0	0	0	0	0	0
Galloping	0	0	**1452**	0	0	0	0	0
Howling	0	0	0	**0**	0	0	0	0
Smelling	0	0	0	0	**1245**	0	0	0
Stationary	0	0	0	0	0	**296677**	0	0
Trotting	0	0	0	0	0	0	**249**	0
Walking	0	0	0	0	1	2	0	**2054**

*Page's confusion matrix (up) with a log loss=0.00108 and Chelsea's confusion matrix (bottom) with a log loss=0.00052.

Appendix 10. Performance and variables importance of Page and Chelsea Random Forest models

Behaviors	Accuracy	Precision	Recall
Chewing	NA	NA	NA
Digging	NA	NA	NA
Galloping	1	1	1
Howling	NA	NA	NA
Smelling	1	1	1
Stationary	1	1	1
Trotting	0.99999	0.99987	1
Walking	0.99999	1	0.99989

Variable	Relative importance	Scaled importance	Percentage (%)
odba	34171156.0	1.00000	25.89
pitch	29155758.0	0.85323	22.09
heave amplitude	19698368.0	0.57646	14.93
surge amplitude	16229028.0	0.47493	12.30
sway amplitude	13954411.0	0.40837	10.57
pitch variance	12704131.0	0.37178	9.63
roll	6058475.5	0.17729	4.59

Behaviors	Accuracy	Precision	Recall
Chewing	1	1	1
Digging	NA	NA	NA
Galloping	1	1	1
Howling	NA	NA	NA
Smelling	0.99999	0.99919	1
Stationary	0.99999	0.99999	1
Trotting	1	1	1
Walking	0.99999	0.99854	1

Variable	Relative importance	Scaled importance	Percentage (%)
odba	15741707.0	1.00000	22.45
pitch	13695290.0	0.87000	19.53
surge amplitude	11962735.0	0.75994	17.06
heave amplitude	9142653.0	0.58079	13.04
sway amplitude	8083349.5	0.51349	11.53
pitch variance	7406748.5	0.47052	10.56
roll	4081784.5	0.25929	5.82

*Performance for each behavior (top left) and variables importance (top right) of the Page's model and performance for each behavior (bottom left) and variables importance (bottom right) of the Chelsea's model.

Appendix 11. Script R used for this study.

*See attached R file called "Script_study-Acc_Beh_Lea_Bouet".