A probabilistic framework for behavioral identification from animal-borne accelerometers

Jane E. Dentinger[1,2†], Luca Börger[3], Mark D. Holton[3], Ruholla Jafari-Marandi[4,5], Durham A. Norman[1,2], Brian K. Smith[5], Seth F. Oppenheimer[6], Bronson K. Strickland[1], Rory P. Wilson[3], Garrett M. Street[1,2]

[1]Department of Wildlife, Fisheries, and Aquaculture, Mississippi State University, Mississippi State, MS, 39762, USA

[2]Quantiative Ecology and Spatial Technologies Laboratory, Mississippi State University, Mississippi State, MS, 39762, USA

[3]Department of Biosciences, College of Science, Swansea University, Swansea, SA2 8PP, United Kingdom

[4]Industrial and Manufacturing Engineering Department, California Polytechnic State University, San Louis Obispo, CA, 93407, USA

[5]Department of Industrial and Systems Engineering, Mississippi State University, Mississippi State, MS, 39762, USA

[6]Department of Mathematics and Statistics, Mississippi State University, Mississippi State, MS, 39762, USA

[†]Corresponding author (jd2381@msstate.edu)

Article type: Standard Paper

Word count: 5958

Number of figures: 6

Number of tables 2

ABSTRACT

1. Many studies of animal distributions use habitat and climactic variables to explain patterns of observed space use. However, without behavioral information, we can only speculate as to why and how these characteristics are important to species persistence.
2. Animal-borne accelerometer and magnetometer data loggers can be used to detect behaviors and when coupled with telemetry improve our understanding of animal space use and habitat requirements. However, these loggers collect tremendous quantities of data requiring automated machine learning techniques to identify patterns in the data. Supervised machine learning requires a set of training signals with known behaviors to train the model to identify the unique signal characteristics associated with each behavior. In contrast, unsupervised approaches aggregate unlabeled signals into groups based purely on signal similarity but, without additional information, do not identify specific behaviors.
3. In this paper, we propose a probabilistic framework for interpreting uncertainty in machine learning techniques—the probability profile—and demonstrate how to *post hoc* identify behaviors within signal groups. We assess model performance using a matrix-based measure of dissimilarity. We used a Random Forest (RF) and a clustered self-organizing map (CSOM) for comparison and demonstrate the use of a behavioral profile for each using a data set of high-frequency accelerometer and magnetometer data collected from 7 captive wild pigs (*Sus scrofa*) moving in a 1 ha outdoor enclosure.
4. We found that the RF had more discrimination than the CSOM which had fewer clusters associated with high probabilities of a single behavior (>50%). The leave-*p*-out cross validation statistic of the probability matrix ($\overline{L^1}$) indicated that there was an average

maximum dissimilarity of 20% and 65% between the training and test data sets for the RF and CSOM methods, respectively.

5. Using a probability profile to describe groups predicted from machine learning allows the variation and error inherent in behavioral prediction to be incorporated directly into the model to better reflect the nuances of behavior derived from accelerometer and/or magnetometer signals. We discuss the data requirements of this framework, demonstrate its application to field data, highlight critical assumptions and caveats, and examine how it may be used to generate new ecological inference.

Keywords: accelerometers, behavior, machine learning, *k*-means clustering, SOM, Random Forest

1. INTRODUCTION

A central aim of ecology is to understand the environmental drivers of fitness and the spatial distribution of individuals. Resources (e.g., energy, water, nutrients) are required for animal survival, growth, and reproduction, thus resource availability is an important predictor of animal fitness (Brown *et al.*, 2004; Matthiopoulos *et al.*, 2015). How organisms seek and acquire resources can also be influenced by predation risk (DeCesare *et al*., 2014), thermoregulation (Street *et al*., 2016), social interactions (Cozzi *et al.*, 2018), and landscape structure (Beyer *et al.*, 2016). Therefore, animal spatial distributions are the result of interacting and competing immediate and long-term needs (Nathan *et al*., 2008). Species-distribution models use knowledge of habitat and behavioral requirements to explain observed patterns of use (Boyce & McDonald, 1999; Guisan & Thuiller, 2005; Elith & Leathwick, 2009); however, individuals may exhibit variation in space use across habitat types or utilize a single landscape for multiple purposes (Street *et al*., 2016; Mitchell *et al.,* 2020). Quantifying this variation is an important step in understanding animal distributions (Johnson, 1980; Van Horne, 1982; Nathan *et al*., 2008; Matthiopoulos *et al*., 2015).

Early studies used direct observation to study animal space use, yet many animals are difficult to observe *in situ* resulting in low sample sizes, human-observer bias and/or detection bias (Altmann, 1974; Johnson, 1980; Aebischer, Robertson & Kenward, 1993; Burghardt *et al.*, 2012). Telemetry and GPS reduce these issues by allowing animals to be relocated from a distance (Cochran & Rexford, 1963; Rodgers *et al*., 1996). Many approaches exist to associate animal spatial locations with habitat characteristics including compositional analysis (Aebischer, Robertson & Kenward, 1993), Resource- and Step-Selection Functions (Boyce & McDonald, 1999; Fortin *et al*., 2005), and Poisson point-process models (Warton & Shepherd, 2010).

However, these approaches are often coarse in scale and thus divorced from the decision-making processes made by animals on a per-step basis. Further, these models are correlative, and the behavioral drivers of animal presence must be inferred or hypothesized from model outputs. Characteristics of movement tracks such as step length and turning angle may be used to identify coarse behavioral states such as "encamped" and "exploratory" behavioral states, thereby improving the accuracy of distribution models (Morales *et al.*, 2004; Auger-Méthé *et al*., 2016; Abrahms *et al*., 2017), but these states are also correlative. For example, a small step length and broad turning angle are often associated with an encamped state (Morales *et al.*, 2004), but numerous animal behaviors may occur within that state so long as they are characterized by similar steps and angles (e.g., fighting, resting, area-restricted foraging, etc.). These states may be better thought of as "movement modes", and without additional information such modes cannot explicitly identify the behaviors performed in a particular state.

Animal-borne accelerometer and magnetometer dataloggers measure the change in speed, direction, and orientation of an animal in the longitudinal, lateral, and dorso-ventral (X-, Y-, and Z-) axes allowing for more detailed behavioral identification (Shepard *et al.*, 2008; Wilson, Shepard, & Liebsch, 2008; Brown *et al.*, 2013; Abrahms *et al*., 2017; Williams *et al.,* 2020). In such a data set, a behavior such as resting may be sensed as a lateral orientation with little to no acceleration versus walking as an upright orientation with oscillations corresponding to each step (Wilson *et al.*, 2018; Chakravarty *et al.,* 2019). However, for many behaviors high sampling frequencies are required for detailed resolution of signals. While these signals can be distinguished visually, loggers often produce tremendous quantities of data requiring automated techniques such as decision trees, Random Forests (RFs), and Artificial Neural Networks (ANN; Yoda *et al.*, 2001; Halsey *et al.*, 2009; Nathan *et al*., 2012; Wang & Xu, 2015) to identify the

behaviors associated with a given signal. Much accelerometer research has applied machine learning techniques in a supervised manner which requires *a priori* labeling of training signals with known behaviors so that the trained model can classify similar signals throughout the full data (Nathan *et al.*, 2012; Leos-Barajas *et al.*, 2017; Chakravarty *et al.*, 2020). These approaches predict the behavior associated with a signal based on its similarity to the training signal but are limited to pre-defined behaviors for signal identification (Nathan *et al*., 2012; Wang & Xu, 2015). By contrast, unsupervised approaches group the data signals based purely on similarity, but biological interpretation of the signal groups identified requires additional *post hoc* attention (Sakamoto *et al.*, 2009; Chimienti *et al.*, 2016). This has often been accomplished by using auxiliary information such as depth or altitude to determine the behavior or behaviors within the groups (Sakamoto *et al.*, 2009).

This is further complicated by the highly subjective nature of behavior and the rarity of standardized ethograms (Stanton *et al.,* 2015). Behaviors can be characterized on multiple scales and depending on the definition, a single behavior may be characterized by multiple signals which may occur in different groups or be so similar in their movement characteristics as to be indistinguishable. For example, behaviors such as grazing and browsing have different head postures and may produce distinct signals, though both may be classified by the analyst as foraging. Conversely, a single signal may be shared across many behaviors; sniffing and rooting may share a head-down and meandering trajectory but have different implications for the animal (one is searching, and the other is consuming). Traditionally, a signal group (cluster or class) would be defined as its majority behavior and other behaviors that produce the same signal determine the error rate (e.g., Nathan *et al.*, 2012). Although we may see variation in a signal group due to model error, this can be assessed through model validation; still, a well validated

model that assigns a group its majority behavior ignores much of the potential for biomechanical similarity among behavioral signals and unrealistically simplifies behavior. The aggregated error may reveal interesting behavior-specific trends such as signal commonalities between behaviors or unique signals from a single behavior. Understanding the probability of observing one of many possible behaviors (rather than a single behavior) given a data signal can lead to new tools and insights in theoretical and applied ecology. Here we propose a framework to investigate variation in behavioral classification and clustering by explicitly characterizing the probability of observing target behaviors given a data signal. We demonstrate the construction of stochastic matrices for unique data signals (groups) and develop a cross-validation procedure with matrix-based validation metrics to evaluate model fit and the consistency of predicted behavioral profiles. Our framework can be applied using either supervised and unsupervised approaches to identify behaviors from biologger data, and we demonstrate the application and interpretation of outputs using both approaches applied to empirical data.

2. METHODS

2.1 *Workflow Overview*

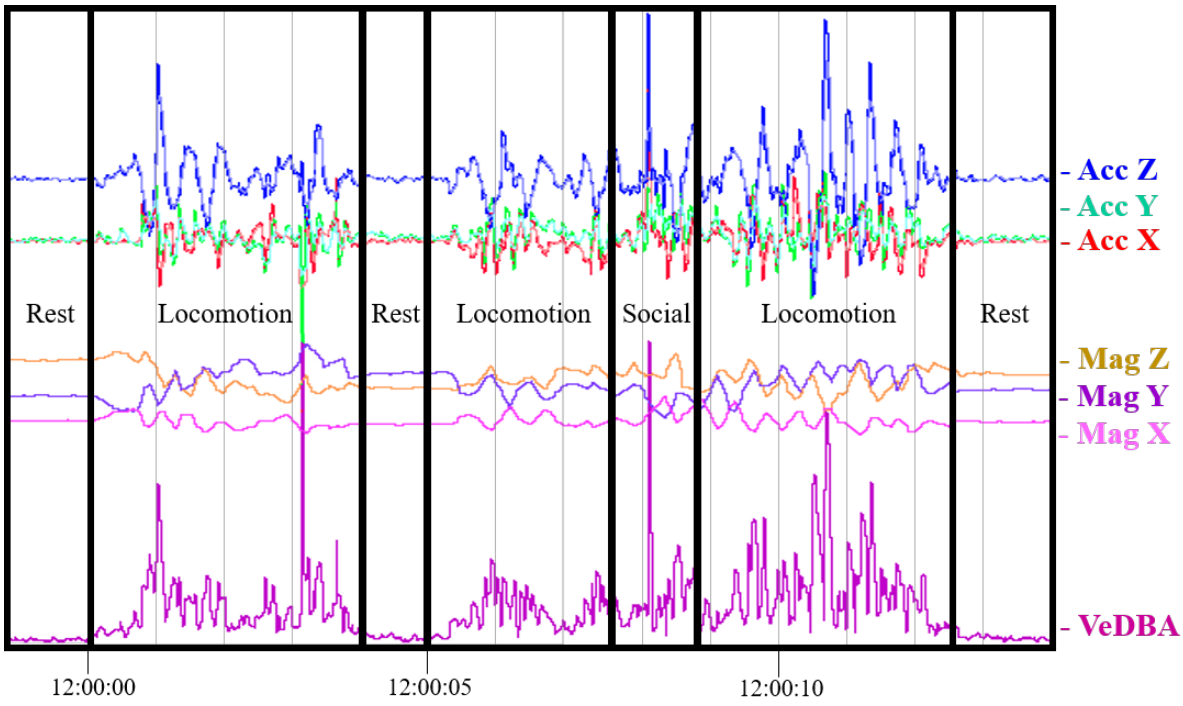
Our proposed approach uses two data sources: 1) a time-series of accelerometer and magnetometer signals and 2) temporally explicit behavioral observations collected simultaneously. Using these two data streams our workflow is:

(0) Combine accelerometer and magnetometer data with behavioral observations.

(1) Process the data to reduce noise and compress the data over a time interval.

(2) Perform a machine learning method to group the signals.

149 (3) Construct a matrix of stochastic vectors for the behavioral probabilities given data
150 signals.
151 (4) Evaluate the predictive accuracy of the matrix given new data.
152 There is a large body of literature on the deployment and use of accelerometers and
153 magnetometers (Wilson, Shepard & Liebsch, 2008; Brown *et al.*, 2013; Williams *et al.*, 2020) and
154 on classifying animal behaviors from direct observation (Yoda *et al.*, 2001; Shepard *et al.*, 2008;
155 Nathan *et al.*, 2012; Wilson *et al.*, 2020; Chakravarty *et al.*, 2020). As such we do not discuss these
156 here and presume that readers have already collected the relevant data (Fig. 1).
157



158 Figure 1. A time series of accelerometer and magnetometer signals annotated with observed
159 behaviors.

160 *2.2 Data Preprocessing*

161 There are many ways to preprocess accelerometer/magnetometer (hereafter, AM) data for
162 analysis, but these often involve calculating additional metrics, smoothing and dimensionality

reduction. Animal-borne accelerometers measure both the dynamic acceleration due to animal movement and the static pull of Earth's gravitational field. Thus, many separate the raw acceleration into the static and dynamic components of acceleration (Shepard *et al.*, 2008). The dynamic acceleration can then be used to derive proxies for energy expenditure, such as VeDBA, the vectorial sum of dynamic body acceleration, or ODBA, Overall Dynamic Body Acceleration (Gleiss *et al.,* 2011; Qasem *et al.,* 2012, Wilson *et al.*, 2020). But regardless of the input channels chosen by the analyst, accelerometer and magnetometer measurements are often noisy and require smoothing of data channels to isolate the signals from the noise (Viviani, Grön & Spitzer, 2005). Given an appropriate smoothing procedure, smoothed signals may be evenly resampled and subsequently subjected to dimensionality reduction (Fig. 2; Appendix VI: 5). Many analysts choose a fixed time-period segmentation (e.g., 1s) and summarize the data over this period using running averages and standard deviations (Shepard et al., 2008). Alternatively, principal components analysis (PCA) can be used to compress the data but retain the critical characteristics of the data (Górecki & Krzyśko, 2012; Appendix VI: 7). The processed data are then paired to their respective direct behavioral observations by their shared dates and times to construct a time-series of dimensionally reduced accelerometer and/or magnetometer signals and their associated behaviors.

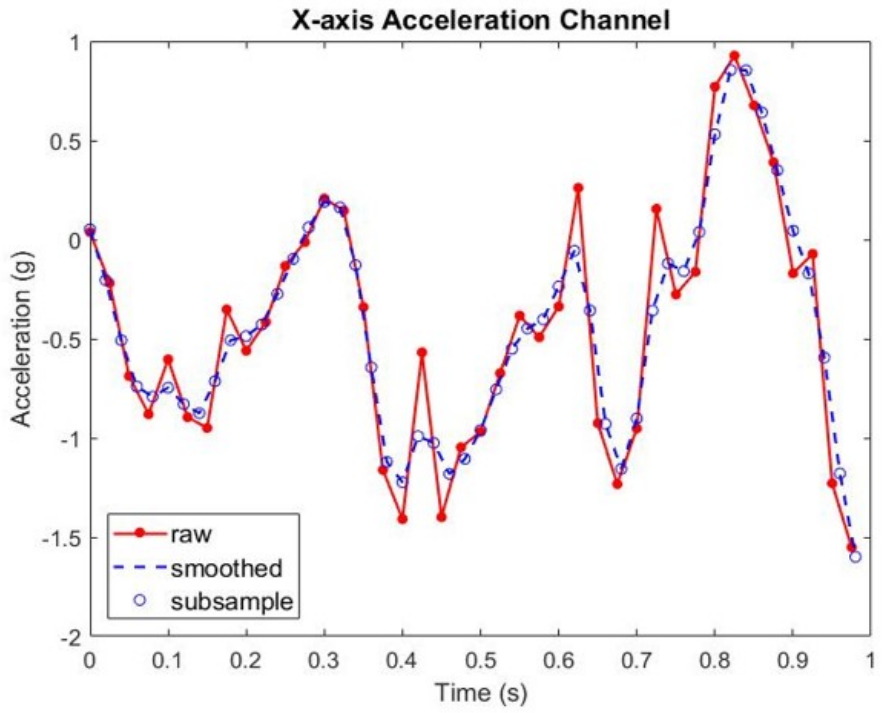


Figure 2. Preprocessing may include segmenting the data into equal time intervals and smoothing across the intervals to reduce noisy data.

2.3 Machine Learning for Classification or Clustering

We next must group the signals based on common signal structures and characterize them in terms of the behaviors observed for each signal type. This may be performed using either supervised or unsupervised machine learning algorithms, and the analyst may use whatever procedure is most appropriate for their data and hypotheses. Here we demonstrate behavioral classification using a common supervised algorithm (Random Forests; RFs), and signal clustering using a novel application of an uncommon unsupervised algorithm (Clustered Self-Organizing Maps; CSOMs). We chose these to demonstrate the robustness of the framework to the nuances and complexity of the selected procedure. Because RFs are increasingly common in ecology, we

do not discuss details of RF-based analysis and instead direct interested readers to previous literature on these subjects (see Cutler *et al.*, 2007; Nathan *et al.*, 2012; Ladds *et al.*, 2017).

SOMs are a type of unsupervised artificial neural network (ANN) and are comprised of two neural layers and their weighted connections (Chon, 2011; Kohonen, 2013; Appendix VI:8). Each input neuron, associated with an explanatory feature in the data, is exhaustively connected to all output neurons (cells on the map; Appendix I: Figure S1). Each cell represents data that share similar characteristics, and neighboring cells are generally more closely related than to more distant cells (Kohonen, 2013; Appendix I: Figure S2). SOMs learn by the exposure of data points to a randomly weighted (initialized) network. Each exposure for a total of N_{epoch} rounds results in a winning neuron (the weighted neuron that best matches the datapoint). During the initial covering phase, both the winning neurons and their neighbors are updated each round according to the initial covering (C, the number of times neighbors will be updated) and the initial neighborhood radius (r_{init}, which controls how many neighboring neurons are updated with the winning neuron). During this phase, the neighborhood radius will decrease incrementally from r_{init} to 1 to minimize overfitting (where $r = 1$ is only the winning neuron and $r = 2$ is the winning neuron and its immediate neighbors; Jafari-Marandi & Keramati, 2014). After the initial covering phase, only the winning neuron will be updated for the remaining rounds. Given an index for an updated neuron (i), a winning neuron (c), an exposure number (t), an epoch number (T), a model associated with neuron i at t ($m_i(t)$), and a data point being exposed to the network ($x(t)$), the model at time t+1 is:

$$m_i(t+1) = m_i(t) + h_i(T, c, r) \times [x(t) - m_i(t)]. \quad \text{Eqn. 1}$$

The sub-function $h_i(T,c,r)$ controls the degree of change in the iterated model based on the proximity of neuron i to neuron c (the winning neuron), the epoch T, and the current neighborhood radius r.

The number of distinct signal types (cells on the map) is determined by the analyst. Determination of the optimal number of cells in the SOM for a given data set is the subject of much debate in the literature (Vesanto & Alhoniemi, 2000; Chon, 2011; Spanakis & Weiss, 2016). While there is no hard-and-fast rule, it is recommended to run SOMs with multiple grid sizes (Shalaginov & Franke, 2015). Regardless of grid size chosen, an objective approach is needed following SOM creation to identify the number of unique signal types that occur in the SOM. Because adjacent cells of the SOM are more similar than are more distant cells (Kohonen, 2013), we use *k*-means clustering to group similar cells into clusters (Vesanto & Alhoniemi, 2000; Jafari-Marandi *et al.*, 2017; Appendix VI: 9). This approach produces a new CSOM wherein each cell is assigned a cluster based on its similarity to its neighbors and the number of clusters (k) to be identified (Fig. 3). Typically, *k*-means clustering requires k to be defined beforehand (Vesanto & Alhoniemi, 2000), but by using a silhouette analysis (Rousseeuw, 1987) conducted on a range of k we can eliminate subjectivity in cluster parameterization. Given k total clusters, the silhouette for a given cell i in cluster l is:

$$s_{i,l} = [min(b_{i,j}) - a_{i,l}] / max[a_{i,l}, min(b_{i,j})], \quad \text{Eqn. 2}$$

where $a_{i,l}$ is the average distance between cell i and all other cells in l, and $b_{i,j}$ is the average distance between cell i and all cells within cluster j (j = 1, 2, …, k, excluding l). The average silhouette ($\overline{s_l}$) indicates within cluster similarity with values approaching 1 indicating appropriate and -1 inappropriate clustering (Rousseeuw 1987). Because *k*-means clustering is sensitive to the starting observations at initialization (Vesanto & Alhoniemi, 2000), the datapoints must be randomized and the clustering repeated N_{rep} times with $\overline{s_l}$ calculated for each cluster and each repetition. The value of k that maximizes the average $\overline{s_l}$ across all N_{rep} trials (k_{best}) is the optimal number of clusters (Appendix I: Figure S3).

After fitting the SOM and identifying the optimal number of signal types, we produce a clustering scheme whereby data in a SOM cell are assigned to the *l*-th cluster, and observations in *l* are as similar as possible. To determine the behaviors associated with each cluster and assess the discriminatory power of the scheme, we pair the observation time of the behaviors to the time of each clustered signal to identify the behaviors that occur in each cluster.

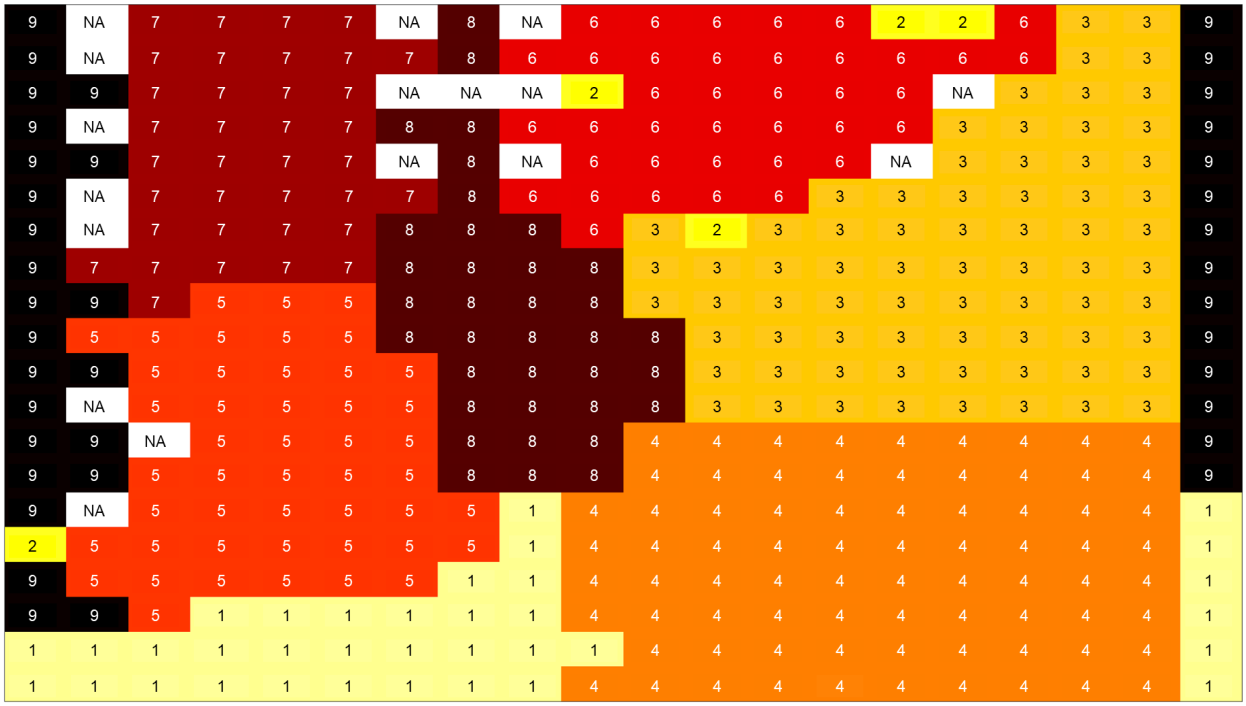


Figure 3: A clustered self-organizing map (CSOM) of wild pig (*Sus scrofa*) AM data with k=9 clusters. Clusters are groups of related cells in the self-organizing map.

2.4 Behavioral Probability Matrix

We summarize the results of the classification and clustering procedures by constructing a probability profile describing the probability of observing each of the potential behaviors given that the data signal was categorized into a signal group (cluster or class). For either the RF or

CSOM (or both), we may construct a $n \times k$ matrix **D** where each entry is the probability of observing the *i*-th behavior in the *j*-th cluster/class. We then use **D**, the probability profile, as our classifier to maintain the signal-specific behavioral variation. This matrix improves upon previous attempts at remote identification of animal behavior as it explicitly characterizes the probabilities of each behavior given a signal type (Nathan *et al.*, 2012). To derive **D**, we construct a $n \times k$ contingency table from the behaviorally explicit data (matrix **A**) describing the proportion of each behavior-by-group combination relative to the entire data set, i.e. P(behavior and group). The summation of each column in **A** produces a $1 \times k$ vector $\vec{\boldsymbol{B}}$ that describes the probability of observing a signal group independent of the behaviors, i.e. P(group), and the summation of each row in **A** produces a $n \times 1$ vector $\vec{\boldsymbol{C}}$ that describes the probability of observing a behavior in the data set independent of the groups, i.e. P(behavior). We may calculate **D** by dividing each entry in column *j* of **A** by its associated column sum or value in $\vec{\boldsymbol{B}}$ (Fig. 4; Appendix VI:10). The resulting matrix **D** represents the probability that a signal is associated with each behavior given that it was assigned to a specific group, i.e. P(behavior|group) (Tables 1 & 2). To extend prediction to data for which there are no corresponding direct behavioral observations, (i.e. data whose true behaviors are unknown), we expose the new data to the model to obtain a class or cluster assignment. Assuming the training data is representative of new data, we can use **D** from the original data to approximate the unknown profile $\mathbf{D}^{+}$.

Matrix A (Proportional matrix)	**Group 1**	**Group 2**	**…**	**Group k**	**Vector C** (Behavior prop.)
Behavior 1	0.20	0.01	…	0.02	0.25
Behavior 2	0.02	0.05	…	0.06	0.15
…	…	…	…	…	…
Behavior n	0.05	0.33	…	0.01	0.45
Vector B (Group prop.)	0.30	0.40	…	0.10	1.00

Matrix D (Probability profile)	**Group 1**	**Group 2**	**…**	**Group k**
Behavior 1	**0.67**	0.03	…	0.20
Behavior 2	0.07	0.13	…	**0.60**
…	…	…	…	…
Behavior n	0.17	**0.83**	…	0.10
Sum	1	1	1	1

Figure 4: An example Matrix **D** (the probability profile) calculated by dividing matrix **A** (the proportional confusion matrix or matching matrix) by vector **B** (its corresponding column sum).

2.5 Model Validation

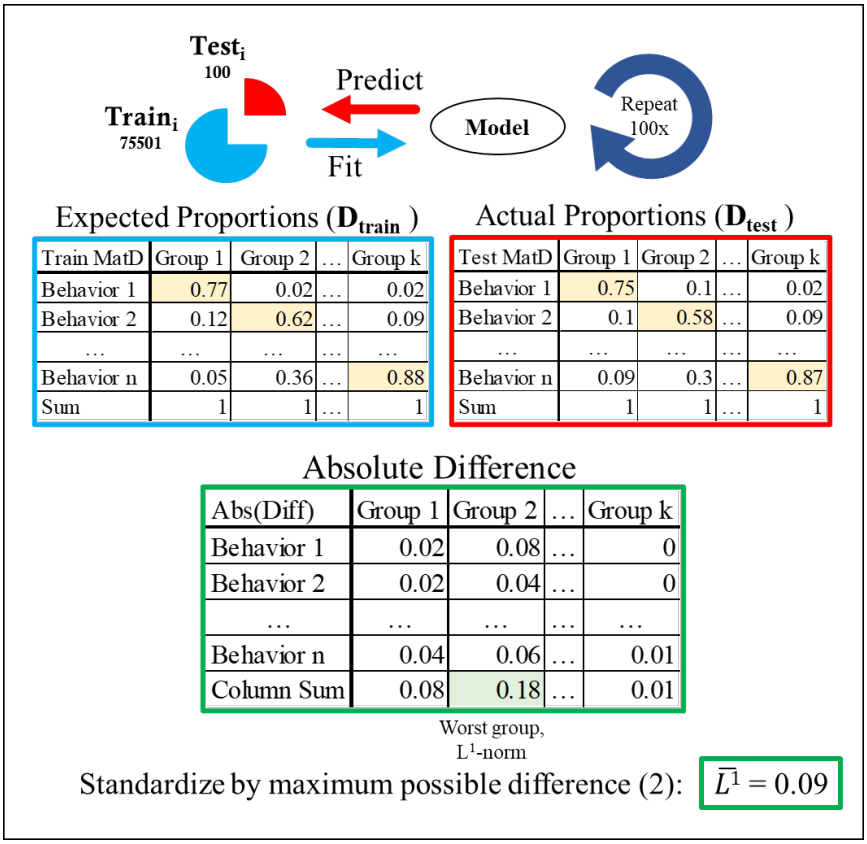
To test model performance and assess whether **D** from the original data approximates the unknown profile $\mathbf{D}^{+}$, we must perform model validation. Validation is conducted using a resampling exercise whereby data with corresponding direct behavioral observations (hereafter termed the known behavioral data set) are randomly subsetted into test (holdout) and training (total excluding holdout) data sets (James *et al.* 2013, Ch. 5). For example, the test subset may contain $n_{test} = 100$ random observations from the known behavior data set, and the remaining

283 (training) data may be used to create a validation model. For the RF, we construct a classification
284 matrix $\mathbf{D}_{\mathbf{train}}$ using a RF algorithm applied to the training data, and the test data are subsequently
285 exposed to the same RF to classify the holdout signals and construct the matrix $\mathbf{D}_{\mathbf{test}}$ (Appendix
286 VI: 14). For the SOM, we construct a validation SOM (vSOM) using the training data and the
287 number of clusters k_{best} as determined by the general SOM (all known behavior) to maintain
288 consistency across the iterations. The matrix $\mathbf{D}_{\mathbf{train}}$ is constructed from vSOM, and the test data is
289 then exposed to the vSOM to obtain cluster assignments and construct $\mathbf{D}_{\mathbf{test}}$ (Appendix VI: 11 &
290 12). If the clustering or classification system is effective, then on average $\mathbf{D}_{\mathbf{train}}$ and $\mathbf{D}_{\mathbf{test}}$ should
291 be similar. To evaluate this, we calculate the L^1-norm (L^1), a measure of matrix dissimilarity,
292 between $\mathbf{D}_{\mathbf{train}}$ and $\mathbf{D}_{\mathbf{test}}$ (Paolini *et al.*, 2018). Given n observed behaviors (i.e. the number of
293 rows in $\mathbf{D}$) and K clusters (i.e. the number of columns in $\mathbf{D}$), L^1 is the maximum across the K
294 clusters:

295 $$L^1 = max\left(\sum_{i=1}^{n}\left|\boldsymbol{D}_{i,k}^{test} - \boldsymbol{D}_{i,k}^{train}\right|\right), k = \{1, 2, \dots, K\}. \quad \text{Eqn. 3}$$

296 The theoretical interval of L^1 as calculated here is [0,2], representing the cases where there is
297 complete similarity (L^1=0) and at least one column of complete dissimilarity (L^1=2) between the
298 test and training datasets (i.e., the behavior predicted and the true behavior are different). Thus,
299 we divide L^1 by 2 to produce $\overline{L^1}$ such that 0 represents perfect similarity and 1 perfect
300 dissimilarity between test and training data sets. If the test data are representative of the training
301 data and if the algorithm accurately predicts new data, then over many iterations of this
302 procedure using M randomly selected test/training sets we expect $\overline{L^1}$ to approach 0 (Fig. 5). Note
303 that L^1 is inherently a conservative metric as it characterizes performance of $\mathbf{D}$ only by its worst
304 performing column. Alternatively, rather than the maximum one could use a standardized
305 arithmetic mean for a metric that may better describe overall performance:

306 $$\overline{L^{avg}} = \sum_{k=1}^{K}\left(\sum_{i=1}^{n}\left|\boldsymbol{D}_{i,k}^{test} - \boldsymbol{D}_{i,k}^{train}\right|\right)\Big/_{2K}.$$ Eqn. 4



Train MatD	Group 1	Group 2	…	Group k
Behavior 1	0.77	0.02	…	0.02
Behavior 2	0.12	0.62	…	0.09
…	…	…	…	…
Behavior n	0.05	0.36	…	0.88
Sum	1	1	…	1

Test MatD	Group 1	Group 2	…	Group k
Behavior 1	0.75	0.1	…	0.02
Behavior 2	0.1	0.58	…	0.09
…	…	…	…	…
Behavior n	0.09	0.3	…	0.87
Sum	1	1	…	1

Abs(Diff)	Group 1	Group 2	…	Group k
Behavior 1	0.02	0.08	…	0
Behavior 2	0.02	0.04	…	0
…	…	…	…	…
Behavior n	0.04	0.06	…	0.01
Column Sum	0.08	0.18	…	0.01

307

308 Figure 5: Validation is performed by randomly subsetting training and test data sets such that the
309 training set is used to estimate a behavioral classification model. The probability profile $\mathbf{D}_{train}$ is
310 the expected behavioral proportions given that group *k* is observed. The test data is then exposed
311 to the model to create the true probability profile of the test set $\mathbf{D}_{test}$. We evaluate how well $\mathbf{D}_{train}$
312 approximates $\mathbf{D}_{test}$ by comparison using the standardized L^1-norm. By performing *M* iterations,
313 we generate a distribution of standardized L^1-norms from which the average L^1-norm may be
314 calculated.

315

316 *2.6 Application to Field Data: Wild Pigs*

317 We used an AM data set collected on wild pigs (*Sus scrofa*) in a semi-natural enclosure to
318 demonstrate the application of this framework to field data. Biologging data was collected from
319 wild pigs translocated to a 1-hectare enclosure in Starkville, MS. The enclosure was equipped
320 with 6 motion-activated infrared cameras synced to the internet to maintain accurate time. Seven
321 wild pigs (3 males and 4 females) were captured and fitted with Daily Diary (DD) accelerometer
322 and magnetometer data loggers (WildBytes Inc., Swansea, UK) sampling at 40 and 13 Hz,
323 respectively. We include these data for demonstration purposes only and emphasize that further
324 processing may be necessary for biological inference (e.g. by assessing different tri-axial rotation
325 rules; see Shepard *et al.*, 2008; Wilson, Shepard & Liebsh, 2008; Brown *et al.*, 2013).
326 Individual pigs were identified from videos and the time, behavior exhibited, and duration
327 of behavior were recorded. Behaviors were classified into 5 broad categories in order of
328 importance: 1) consumption, 2) resting, 3) environmental interaction, 4) locomotion, and 5) social
329 interaction (Appendix II). We used a ranking system to ensure that we detected behaviors even if
330 they commonly occurred with other behaviors (i.e. head down and consuming food continuously
331 while walking would be classified as consumption). Each observation was "trimmed" (i.e. the first
332 and last second removed), resulting in a 1-second minimum for all observations. If a single
333 behavior was interrupted by a brief bout of a different behavior (<2 s), whereafter the original
334 behavior was resumed, the whole bout was classified as the first behavior. For example, a time
335 series consisting of walking (5 s), freeze and scan for predators (1 s), and resume walking (5 s)
336 would be classified as locomotion (11 s). Behaviors were paired with DD data by using the video
337 timestamp and the DD's internal clock (synched prior to deployment).
338 Each second of DD data (7 channels: Acc X, Acc Y, Acc Z, Mag X, Mag Y, Mag Z &
339 VeDBA at ~40 Hz/channel) was smoothed using a cubic polynomial spline ($\lambda = 1$) and resampled

evenly from each curve (7 channels x 50 samples/s = 350 measurements/s). A PCA was performed on the samples for each of the 7 data channels, and the first 3 principal component vectors (PCs; 95-99% of the variation) per channel were selected for subsequent analysis (7 channels x 3 PCAs: 21 PCAs). A 20 x 20 SOM was parameterized with $r_{init} = 3$, $C = 30$, and $N_{epoch} = 200$, and k-means clustering was performed using $k = 3, 4, \ldots, 25$ ($N_{rep} = 3000$; see Appendix IV for additional analyses using alternative grid sizes). The RF was fit with 500 trees.

8,418 videos of 7 captive wild pigs were recorded and classified manually, resulting in 75,601 seconds (21 hours, ~3 hrs/pig) of observed behavior from 5 behavioral categories. Only biologging data for which there was a corresponding behavioral observation were used in the analysis. The number of observed behaviors in each group were tallied and matrices **A** and **D** and vectors $\vec{\boldsymbol{B}}$ and $\vec{\boldsymbol{C}}$ were constructed. Data cleaning, pairing and analyses were conducted in Daily Diary Multi Trace v. 25/Feb/2018 (WildBytes, Inc.), Program R v. 3.3.1 (R Core Team, 2016; Appendix VI: 1:3,10,12:14) and MATLAB 2017a (MathWorks, Appendix VI: 4:9, 11, 15). The Random Forests were fit with the randomForest package (Liaw & Wiener, 2002; Appendix VI: 13:14).

3. RESULTS

As proportions of the entire data set, environmental interaction (37.62%) and locomotion (25.90%) were the most observed behaviors, followed by consumption (17.76%), resting (11.52%) and social interaction (7.21%). The average silhouette for the general SOM was maximized at $k = 9$ ($\bar{s}_l = 0.4895$; Appendix I: Figure S3). The Random Forest converged after 100 trees.

For the RF method, there was an overall 22% out-of-bag error rate irrespective of class. Each group was best predicted by a single behavior ranging from 97% (resting) to 70% (locomotion; Table 1). The "consumption" group (group 1) was most likely to be the consumption behavior (76%) followed by locomotion behavior (15%). The "locomotion" group (group 2) was most likely to be a locomotion behavior (70%) followed by a consumption behavior (17.5%). The "environmental interaction" group (group 3) was most likely to be an environmental interaction behavior (74%) followed by a consumption behavior (16%). This might suggest that behaviors associated with consumption in wild pigs produce several variable signals with some that appear more similar to a locomotion or environmental interaction signal than most consumption signals. This might be explained by blended signals where an animal is consuming food and thus its behavior is classified under consumption but is simultaneously walking or scanning (locomotion or environmental interaction respectively).

For the CSOM method, the clusters represented diverse behaviors with a high proportion of clusters that predicted behavioral combinations rather than a single behavior. Only 4/9 total clusters exhibited a >50% probability for a single behavior (2 environmental interaction, 1 locomotion, 1 consumption; Table 2). The remaining clusters had a dominant behavior that was larger than the other behaviors but <50% of the total probability of the cluster (3 environmental interaction, 1 locomotion, 1 resting). Environmental interaction was the most prevalent behavior in our data and was either the most or second most probable in all clusters. This may indicate that the CSOM struggled to distinguish environmental interaction from the other behaviors perhaps because the primary behaviors that compose environmental interaction (i.e. sniffing, scanning) may not have particularly distinctive signals. Further, pigs frequently interrupted one behavior with brief bouts of other behaviors that may have introduced variation in our

385 classification system or produced blended signals making it difficult to separate the behaviors
386 into discrete categories. Locomotion was the second most prevalent behavior with the majority
387 probability in clusters 2 and 8 and was second most probable in 3 clusters. Many non-locomotive
388 behaviors were frequently accompanied by meandering (slow tortuous movement) which may
389 share similar signal characteristics to walking (locomotion) or were punctuated by bouts of
390 walking which may mask the other behavior's signals completely. Consumption composed 60%
391 of cluster 3 and resting behavior was the most prevalent in cluster 9 at 38%. Although social
392 interaction did not dominate any cluster, it was the second most likely behavior at 25% in cluster
393 1.

394 The resampling procedure showed that there was an average maximum cluster
395 dissimilarity of $\overline{L^{1}}$ = 0.65 or 65% between the training and test data sets and an average mean
396 cluster dissimilarity of $\overline{L^{avg}}$ = 0.31 or 31% using the CSOM, and an average maximum cluster
397 dissimilarity of $\overline{L^{1}}$ = 0.20 or 20% and an average mean cluster dissimilarity of $\overline{L^{avg}}$ = .11 or 11%
398 for the RF (Fig. 6). This may be interpreted as 35% similarity between the test and training
399 probability matrices for the CSOM method and 80% similarity for the RF method, indicating that
400 the RF approach better predicted new data even though the hold-out data was 1/1000th the size of
401 the training data. This is good evidence to conclude that the RF, on average, accurately predicted
402 new data, and that the clusters identified as most distinctive by the 20 x 20 CSOM method did
403 not match our assigned behavioral labels. Thus, by performing this crucial validation step, we
404 can conclude that matrix **D** obtained from the trained RF, but not the CSOM, can be used to
405 approximate the probabilistic behavioral profiles $\mathbf{D}^{+}$ from unobserved data (but see Appendix IV
406 for additional CSOM analyses).

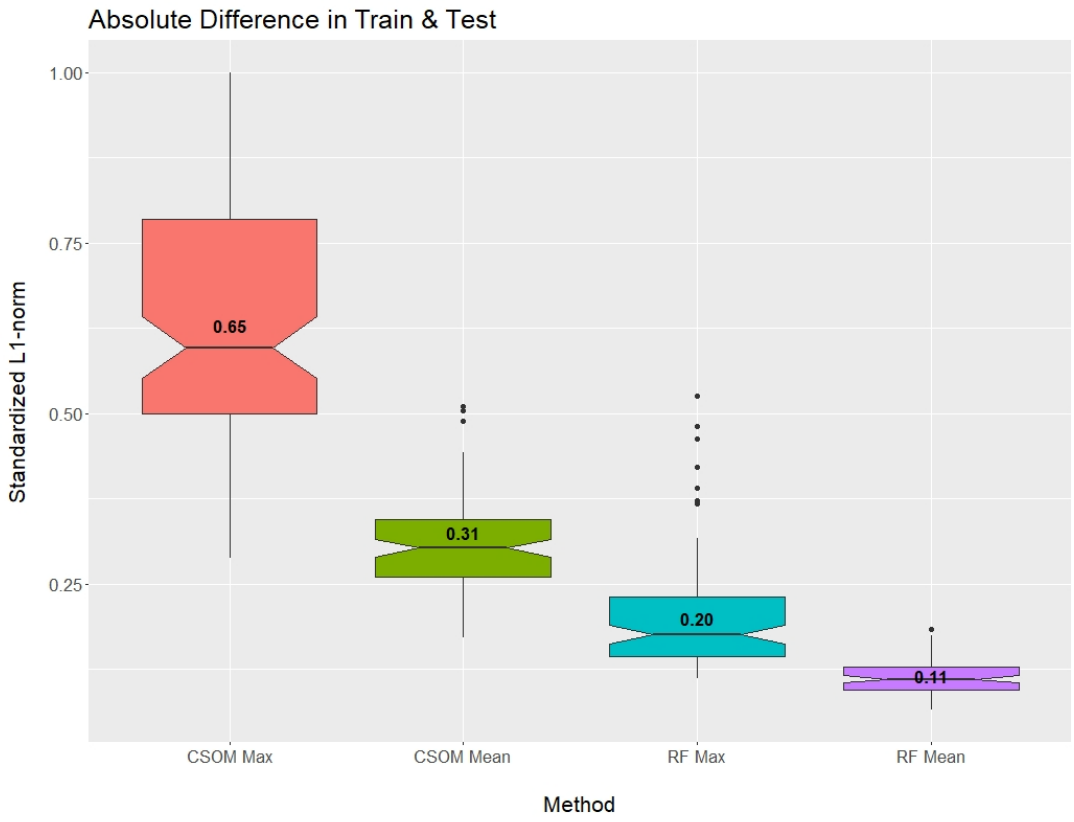


Figure 6. The distribution of the maximum and average group dissimilarity (standardized L^1-norm; L1) for the CSOM and RF methods.

4. DISCUSSION

Behavior is an important but often overlooked tool in interpreting animal distributions (Roever *et al.*, 2014; Abrahms *et al.*, 2017), and a variety of approaches have been used to link animal behavior to data sets of acceleration (Wilson, Shepard & Liebsch, 2008; Brown *et al.*, 2013). Here, we presented a probabilistic framework for interpreting animal behavior classified or clustered from AM signals. We applied two machine learning techniques: one supervised, where behaviors were paired with accelerometer data prior to classification, and one unsupervised, which first clustered the signals and then used the observed behaviors (Sakamoto

*et al.,*2009; Nathan *et al.*, 2012; Chimienti *et al.*, 2016). Both approaches ultimately produce a matrix identifying the probability of observing a behavior given a known acceleration signal (Fig. 3, Tables 1 & 2) but validation showed that the CSOM method performed poorly and thus the probability profile cannot be trusted. Though most current approaches assign a signal type to a target behavior (Nathan *et al.*, 2012), the RF showed many behaviors exhibited similar data signals, and several groups were likely to reflect multiple rather than single behaviors. This may be because some behaviors may be difficult to discriminate using collar attached accelerometers (e.g., Nathan *et al*., 2012; Wang & Xu, 2015) and by using a probability profile, we accommodate this explicitly.

Quantifying the variation in animal behaviors across landscapes is crucial for our understanding of animal movement (Johnson, 1980; Van Horne, 1982; Nathan *et al*., 2008; Matthiopoulos *et al*., 2015). Animals may use a single landscape for multiple purposes, complicating the ability of researchers to predict a single behavior from landscape features (Street *et al.*, 2016). Using a probability profile accounts for this but also allows for the empirical calculation of multiple behavioral probabilities for any landscape. Thus, this may be used to estimate the relative probability that an animal will engage in a behavior provided an environmental context. Conceptually similar to a Resource-Selection Function (RSF; Boyce & McDonald, 1999), one could estimate the probability of observing the behavior given a suite of covariates and thus predict the relative probability of observing a given behavior in space and time. We foresee 2 general approaches to producing such a model. First, a simple logit model as commonly applied to e.g., RSFs could produce the desired spatial model of behavior. However, the very high refresh rates for accelerometers (e.g., 1+ Hz) will typically produce many behavioral observations at a given spatial location, particularly if the resolution of the spatial

layers is large (as is common for remotely sensed landscape data products). Application of logit models will thus require researchers to consider how best to accommodate what is a clear violation of independence for such a model. This could be accomplished by sub-sampling data to a single observation per pixel, or by aggregating/averaging probabilities for behaviors within a given pixel. Neither outcome is particularly appealing due to loss of information, but the resulting logit models are well described in the literature and have ample tools for model validation, calibration, and sensitivity (e.g., Chivers *et al.*, 2014; Harrell, 2016). Alternatively, Hidden Markov Models (HMMs) have been recently applied to accelerometry data (Leos-Barajas, 2017). Though far less common than standard logit models, these or similar models that incorporate autocorrelation structures (e.g., temporal autoregressive models; Hooten, 2017) could be used to generate the desired spatial models while accommodating dependence between consecutive observations.

Provided that we have a properly parameterized probabilistic spatial model of behavior, multiplicative pairing of such models with simple models of space use (e.g., RSFs) combines the relative probability of occurrence with the relative probability of observing different behaviors in that location. Akin to a utilization kernel for animal behavior, this "behavior kernel" should reflect the relative probability of a known behavior, weighted by the probability of the animal using that space (Worton, 1989). These models, enabled by the probabilistic framework presented here, may offer tremendous opportunities for new insight. For example, the Ideal Free Distribution (IFD) posits the equilibrium distribution of a population will be achieved when all individuals have equal fitness relative to local density and environmental quality (Fretwell & Lucas, 1969). If animal fitness may be thought of as an energy budget whereby animals must balance energetic costs with inputs (Brown *et al*., 2004), then we should expect that animal

fitness will be highest in areas where energy intake exceeds expenditure. These areas of high relative fitness should be identifiable following the two-stage model described above as locations with high probabilities of occurrence and of behaviors promoting energy intake. We expect that the behavior kernel will strongly correlate with the IFD and provide a mechanistic link between the behaviors of individuals and the distributions of populations, a long-standing goal of predictive ecology (Matthiopoulos *et al*., 2015).

One concern with our approach is that the model may not accurately represent the behavioral probabilities of new data (Yates *et al*., 2018). By performing a resampling exercise and using the L^1 matrix-based statistic, we observed that data withheld from the RF were accurately classified but the CSOM method did not accurately identify new data. Thus, the CSOM model overall did not perform well and thus, is not suitable for either the direct assignment or probability profile approach. Our approach acknowledges that behaviors with similar characteristics should be difficult to identify in any machine learning approach. Although this problem may be more pronounced in unsupervised learning algorithms (Nathan *et al*., 2012), it is a common challenge of working with behavioral data. We emphasize that the type of machine learning technique used should be dictated by the questions the analyst has for the data set; however, use of a probability profile rather than a direct assignment can be adopted regardless of the algorithm used to group the data provided that the model validates well.

It is also worth noting that the apparent poor performance of the CSOM may be attributed to the validation metric L^1. This metric is mathematically well described and is calculated as the maximum observed summed difference between entries of the columns of two matrices. If two matrices are identical save for a single column, L^1 bases dissimilarity only on the differing column. L^1 is thus a highly conservative metric of matrix dissimilarity that may be interpreted as

488 the maximum dissimilarity observed across signal types. By contrast, we may instead evaluate
489 matrix dissimilarity using the arithmetic mean of the summed differences between columns
490 rather than the maximum. This metric is no longer conservative but may be a better reflection of
491 overall performance when few signal types perform poorly. Indeed, the average summed
492 differences showed overall better agreement between $\mathbf{D}_{\mathbf{test}}$ and $\mathbf{D}_{\mathbf{train}}$ than suggested by L^1 for the
493 both the CSOM and RF approaches (Fig. 6). Selection of the validation metric should be
494 determined by the specific needs of the project and the degree of predictive accuracy desired.
495 There was likely some error in label assignment as it was difficult to identify individual
496 pigs and to temporally synchronize DD observations to videos. Additionally, the behavioral
497 proportions observed via video may not represent the true frequency of pig behavior in the
498 enclosure and may differ from those of wild pigs (Burghardt *et al.*, 2012). Because captive pigs
499 were supplied with food, they may spend less time engaging in consumption behaviors as their
500 wild counterparts (Grandia, Van Dijk & Koene, 2001). However, of critical concern is the
501 possibility that certain behaviors may not be represented in the training data (e.g., swimming).
502 For such cases, our system would be incapable of identifying these behaviors. We reiterate that
503 data need to be representative of the total variation in data signals and behaviors and emphasize
504 that validation is critical for any remote sensing exercise (Congalton, 1991; Nathan *et al*., 2012).
505
506 ACKNOWLEDGEMENTS
507 Funding for this research was provided by the Mississippi Agriculture and Forestry Research
508 Station (MAFES) Strategic Research Initiative and the Noble Research Institute. We also thank
509 AJ Benney, Clay Gibson, Bryant Haley, Gwen Jones, Kelsey Paolini, and numerous other

510 volunteers for their contributions. All procedures were approved by the MSU Institutional
511 Animal Care and Use Committee (Protocol #16-062).
512
513 AUTHOR CONTRIBUTION STATEMENT
514 GMS, LB, and RPW conceived the ideas for the paper. MDH, BKS, BKS, RJM, GMS, and SFO
515 combined expertise to create the methodology. JED and DAN collected the data. JED performed
516 the analyses in collaboration with MDH, RJM, and GMS. JED and GMS led writing of the
517 manuscript. All authors contributed to the drafts and gave final approval for publication.
518
519 DATA ACCESSIBILITY
520 Data will be archived at the Mississippi State University Repository for permanent archiving. All
521 R and MATLAB scripts are available in the Appendix VI: 1:15.
522

523 TABLES

524 Table 1. The probability profile (matrix **D**) for the Random Forest. data set

RF: Matrix **D**					
	Group 1: "Consumption"	Group 2: "Environmental Interaction"	Group 3: "Locomotion"	Group 4: "Resting"	Group 5: "Social Interaction"
Consumption	0.7555	0.1593	0.1753	0.0072	0.0148
Environmental Interaction	0.0489	0.7383	0.0739	0.0014	0.0099
Locomotion	0.1480	0.0788	0.7009	0.0099	0.0081
Resting	0.0208	0.0073	0.0155	0.9749	0.0115
Social Interaction	0.0269	0.0163	0.0344	0.0067	0.9558

525

526

527 Table 2: The behavioral profile, matrix **D**, for the clustered self-organizing map.

CSOM: Matrix **D**									
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
Consumption	0.1581	0.0189	0.5989	0.1359	0.1112	0.0334	0.1585	0.0841	0.1623
Environmental Interaction	0.3159	0.1698	0.2447	0.4892	0.5063	0.5016	0.4051	0.2758	0.2026
Locomotion	0.1786	0.7359	0.1405	0.2251	0.3019	0.3177	0.2672	0.4575	0.1604
Resting	0.1006	0.0566	0.0102	0.1123	0.0351	0.1388	0.1260	0.0697	0.3814
Social Interaction	0.2468	0.0189	0.0057	0.0374	0.04545	0.0085	0.0432	0.1129	0.0934

528

529

530 REFERENCES

531 Abrahms, B., Sawyer, S. C., Jordan, N. R., McNutt, J. W., Wilson, A. M. & Brashares, J. S.
532 (2017). Does wildlife resource selection accurately inform corridor conservation? *Journal*
533 *of Applied Ecology*, *54*, 412–422.

534 Aebischer, N. J., Robertson, P. A. & Kenward, R. E. (1993). Compositional analysis of habitat-
535 use from animal radio-tracking data. *Ecology*, *74*, 1313–1325.

536 Altmann, J. (1974). Observational study of behavior: sampling. *Behaviour*, *49*, 227–267.

537 Auger-Méthé, M., Derocher, A. E., DeMars, C. A., Plank, M. J., Codling, E. A. & Lewis, M. A.
538 (2016). Evaluating the random search strategies in three mammals from distinct feeding
539 guilds. *Journal of Animal Ecology, 85*, 1411-1421.

540 Beyer, H. L., Gurarie, E., Börger, L., Panzacchi, M., Basille, M., Herfindal, I., Van Mooter, B.,
541 Subhash, R. L. & Matthiopoulos, J. (2016). “You shall not pass!”: quantifying barrier
542 permeability and proximity avoidance by animals. *Journal of Animal Ecology*, *85*, 43–53.

543 Boyce, M. S. & McDonald, L. L. (1999). Relating populations to habitats using resource
544 selection functions. *Trends in Ecology and Evolution*, *14*, 268–272.

545 Brown, D. D., Kays, R., Wikelski, M., Wilson, R. P. & Klimley, A. P. (2013). Observing the
546 unwatchable through acceleration logging of animal behavior. *Animal Biotelemetry*, *1*, 1–
547 16.

548 Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M. & West, G. B. (2004). Toward a
549 metabolic theory of ecology. *Ecology*, *85*, 1771–1789.

550 Burghardt, G., Bartmess-LeVasseur, J., Browning, S. A., Morrison, K. E., Stec, C. L., Zachau, C.
551 E. & Freeberg, T. (2012). Minimizing observer bias in behavioral studies: a review and
552 recommendations. *Ethology*, *118*, 511-517.

553 Chakravarty, P., Cozzi, G., Ozgul, A. & Aminian, K. (2019). A novel biomechanical approach
554 for animal behaviour recognition using accelerometers, *Methods in Ecology and Evolution,*
555 *10,* 802-814.
556 Chakravarty, P., Cozzi, G., Dejnabadi, H., Léziart, P., Manser, M., Ozgul, A. & Aminian, K.
557 (2020). Seek and learn: Automated identification of microevents in animal behaviour using
558 envelopes of acceleration data and machine learning. *Methods in Ecology and Evolution,*
559 *00,* 1-13.
560 Chimienti, M., Cornulier, T., Owen, E., Bolton, M., Davies, I. M., Travis, J. M. J. & Scott, B. E.
561 (2016). The use of an unsupervised learning approach for characterizing latent behaviors in
562 accelerometer data. *Ecology and Evolution*, *6*, 727–741.
563 Chivers, C., Leung, B. & Yan, N. D. (2014). Validation and calibration of probabilistic
564 predictions in ecology. *Methods in Ecology and Evolution, 5,* 1023-1032.
565 Chon, T. S. (2011). Self-organizing maps applied to ecological sciences. *Ecological Informatics,*
566 *6*, 50–61.
567 Cochran, W. W. & Rexford D. L. (1963). A radio-tracking system for wild animals. *Journal of*
568 *Wildlife Management, 27*, 9–24.
569 Congalton, R. G. (1991). A review of assessing the accuracy of classification of remotely sensed
570 data. *Remote Sensing of Environment, 37*, 35-46.
571 Cozzi, G., Maag, N., Borger, L., Clutton-Brock, T. H. & Ozgul, A. (2018). Socially informed
572 dispersal in a territorial cooperative breeder. *Journal of Animal Ecology*, *87*, 838-849.
573 Cutler, D. R., Edwards, T. C., Beard, K. J., Cutler, A., Hess, K. T., Gibson, J. & Lawler, J. J.
574 (2007). Random forests for classification in ecology. *Ecology, 88,* 2783-2792.

DeCesare, N. J., Hebblewhite, M., Bradley, M., Hervieux, D., Neufeld, L. & Musiani, M. (2014). Linking habitat selection and predation risk to spatial variation in survival. *Journal of Animal Ecology*, *83*, 343–352.

Elith, J. & Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annals of Ecology, Evolution, and Systematics*, *40*, 677–697.

Fortin, M. J., Keitt, T. H., Maurer, B. A., Taper, M. L., Kaufman, D. M. & Blackburn, T. M. (2005). Species' geographic ranges and distributional limits: pattern analysis and statistical issues. *Oikos*, *108*, 7–17.

Fretwell S. D. & Lucas, H. L. (1969). On territorial behavior and other factors influencing habitat distribution in birds. *Acta Biotheoretica*, *19*, 45-52.

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

Górecki, T., & Krzyśko, M. (2012). Functional principal components analysis. *Data Analysis Methods and its Applications, CH Beck, Warszawa*: 71-87.

Grandia, P., Van Dijk, J. J., & Koene, P. (2001). Stimulating natural behavior in captive bears. *Ursus*, *12*, 199–202.

Guisan, A. & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, *8*, 993–1009.

Halsey, L. G., Portugal, S. J., Smith, J. A., Murn, C. P. & Wilson, R. P. (2009). Recording raptor behavior on the wing via accelerometry. *Journal of Field Ornithology, 80,* 171–177.

Harrell, F. E. (2016). *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis.* Springer, New York, NY.

Hebblewhite, M. & Haydon, D. T. (2010). Distinguishing technology from biology: a critical review of the use of GPS telemetry data in ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences, 365*, 2303–2312.

Hooten, M. B., Johnson, D. S., McClintock, B. T. & Morales, J. M. (2017). *Animal Movement: Statistical Models for Telemetry Data*. CRC Press, Boca Raton, LA.

Jafari-Marandi, R., Khanzadeh, M., Smith, B. K. & Bian, L. (2017). Self-organizing and error driven (SOED) artificial neural network for smarter classifications. *Journal of Computational Design and Engineering*, *4*, 282-304.

Jafari-Marandi, R. & Keramati, A. (2014). Webpage clustering-taking the zero step: a case study of an Iranian website. *Journal of Web Engineering*, *13*, 333-360.

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013). *Introduction to Statistical Learning with Applications in R*. Springer, New York, NY.

Johnson, D. H. (1980). The comparison of usage and availability measurements for evaluating resource preference. *Ecology*, *61*, 65–71.

Kohonen, T. (2013). Essentials of the self-organizing map. *Neural Networks*, *37*, 52-65.

Ladds, M. A., Thompson, A. P., Kadar, J., Slip, D. J., Hocking, D. P. & Harcourt, R. G. (2017). Super machine learning improving accuracy and reducing variance of behavior classification from accelerometer. *Animal Biotelemetry, 5,* 1-9.

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

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

The MathWorks Inc. (2017a). MATLAB, Natick, Massachusetts, United States.

Matthiopoulos, J., Fieberg, J., Aarts, G., Beyer, H. L., Morales, J. M. & Haydon, D. T. (2015). Establishing the link between habitat selection and animal population dynamics. *Ecological Monographs*, *85*, 413–436.

Mitchell, L. J., Kohler, T., White, P. C. L. & Arnold, K. E. (2020). High interindividual variability in habitat selection and functional habitat relationships in European nightjars over a period of habitat change, *Ecology and Evolution, 10,* 5932-5945.

Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E. & Fryxell, J. M. (2004). Extracting more out of relocation data: building movements models as mixtures of random walks. *Ecology*, *85*, 2436–2445.

Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D. & Smouse, P. E. (2008). A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences*, *105*, 19052–19059.

Nathan, R., Spiegel, O., Fortmann-Roe, S., Harel, R., Wikelski, M. & Getz, W. M. (2012). Using tri-axial acceleration data to identify behavioral modes of free-ranging animals: general concepts and tools illustrated for griffon vultures. *Journal of Experimental Biology*, *215*, 986–996.

Paolini, K. E., Strickland, B. K., Tegt, J. L., VerCauteren, K. C. & Street, G. M. (2018). Seasonal variation in preference dictates space use in an invasive generalist. *PLoS ONE*, *13*, e0199078.

642 Qasem, L., Cardew, A., Wilson, A., Griffiths, I., Halsey, L. G., Shepard, E. L. C., Gleiss, A. C.,
643 & Wilson, R. P. (2012). Tri-axial dynamic acceleration as a proxy for animal energy
644 expenditure; should we be summing values or calculating the vector? *PLoS ONE*, *7*,
645 e31187.
646 R Core Team, (2016). A language and environment for statistical computing. R Foundation for
647 Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
648 Rodgers, A. R., Rempel, R. S. & Abraham, K. F. (1996). A GPS-based telemetry system.
649 *Wildlife Society Bulletin, 24*, 559-566.
650 Roever, C., Beyer, H., Chase, M. & Van Aarde, R. (2014). The pitfalls of ignoring behavior
651 when quantifying habitat selection. *Diversity and Distributions*, *20*, 322-333.
652 Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of
653 cluster data. *Computational and Applied Mathematics, 20*, 53-65.
654 Sakamoto, K. Q., Sato, K., Ishizuka, M., Watanuki, Y., Takahashi, A., Daunt, F., & Wanless, S.
655 (2009). Can ethograms be automatically generated using body acceleration data from free-
656 ranging birds? *PLoS ONE*, *4*, e5379.
657 Shalaginov, A. & Franke, K. (2015). A new method for an optimal SOM size determination in
658 Neuro-Fuzzy for the Digital Forensics applications. *IWANN 2015*: *Advances in*
659 *Computational Intelligence, 9095,* 549-563
660 Shepard, E. L. C., Wilson, R. P., Quintana, F., Gómez Laich, A., Liebsch, N., Albareda, D.,
661 Halsey, L. G., Gleiss, A., Morgan, D. T., Myers, A. E, Newman, C., & McDonald, D.
662 (2008). Identification of animal movement patterns using tri-axial accelerometry.
663 *Endangered Species Research*, *10*, 47–60.

664 Stanton, L. A., Sullivan, M. S., & Fazio, J. M. (2015). A standardized ethogram for the felidae:
665 A tool for behavioral researchers. *Applied Animal Behaviour Science*, *173*, 1-15.
666 Spanakis, G. & Weiss, G. (2017). Enhancing visual clustering using Adaptive Moving Self-
667 Organizing Maps (AMSOM). *ICAART 2016: Agents and Artificial Intelligence, 1016,* 189-
668 211.
669 Street, G. M., Fieberg, J., Rodgers, A. R., Carstensen, M., Moen, R., Moore, S. A., Windels, S.
670 K., & Forester J. D. (2016). Habitat functional response mitigates reduced foraging
671 opportunity: implications for animal fitness and space use. *Landscape Ecology*, *31*, 1939-
672 1953.
673 Van Horne, B. & Ford, R. G. (1982). Niche breadth calculation based on discriminant analysis.
674 *Ecology*, *63*, 1172-1174.
675 Vesanto, J. & Alhoniemi, E. (2000). Clustering of the Self-Organizing Map. *IEEE Transactions*
676 *on neural networks*, *11*, 586-600.
677 Viviani, R., Grön, G., & Spitzer, M. (2005). Functional principal component analysis of fMRI
678 data. *Human brain mapping*, *24*, 109-129.
679 Wang, C., & Xu, G. (2015). A mixture hierarchical model for response times and response
680 accuracy. *British Journal of Mathematical and Statistical Psychology*, *68*, 456-477.
681 Warton, D. I., & Shepherd, L. C. (2010). Poisson point process models solve the "pseudo-
682 absence problem" for presence-only data in ecology. *Annals of Applied Statistics*, *4*, 1383–
683 1402.
684 Williams, J. H., Taylor, L. A., Benhamou, S., Bijleveld, A. I., Clay, T. A., de Grissac, S.,
685 Demšar, U., English, H. M., Franconi, N., Gómez-Laich, A., Griffiths, R. C., Kay, W. P.,
686 Morales, J. M., Potts, J. R., Rogerson, K. F., Rutz, C., Spelt, A., Trevail, A. M., Wilson, R.

687 P., & Börger, L. (2020). Optimising the use of bio-loggers for movement ecology research.
688 *Journal of Animal Ecology, 89,* 186-206.
689 Wilson, R. P., Shepard, E. L. C., & Liebsch, N. (2008). Prying into the intimate details of animal
690 lives: use of a daily diary on animals. *Endangered Species Research*, *4*, 123–137.
691 Wilson, R. P., Holton, M. D., di Virgilio, A., Williams, H., Shepard, E. L. C., Lambertucci, S.,
692 Quintana, F., Sala, J. E., Balaji, B., Lee, E. S., Srivastava, M., Scantlebury, D. M. & Duarte,
693 C. M. (2018). Give the machine a hand: A Boolean time-based decision-tree template for
694 rapidly finding animal behaviours in multisensory data. *Methods in Ecology and Evolution,*
695 *9,* 2206-2215.
696 Wilson, R. P., Börger, L., Holton, M. D., Scantlebury, D. M., Gómez-Laich, A., Quintana, F.,
697 Rosell, F., Graf, P. M., Williams, H., Gunner, R., Hopkins, L., Marks, N., Gerladi, N. R.,
698 Duarte, C. M., Scott, R., Strano, M. S., Robotka, H., Eizaguirre, C., Fahlman, A., &
699 Shepard, E. L. C. (2020). Estimates for energy expenditure in free-living animals using
700 acceleration proxies: A reappraisal. *Journal of Animal Ecology, 89,* 161-172.
701 Worton, B.J. (1989). Kernel methods for estimating the utilization distribution in home-range
702 studies. *Ecology*, *70*, 164-168.
703 Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., Parnell, S., Fielding, A.
704 H., Bamford, A. J., Ban, S., Barbosa, A. M., Dormann, C. F., Elith, J., Embling, C. B.,
705 Ervin, G. N., Fisher, R., Gould, S., Graf, R. F., Gregr, E. J., Halpin, P. N., Heikkinen, R. K.,
706 Heinänen, S., Jones, A. R., Krishnakumar, P. K., Lauria, V., Lozano-Montes, H., Mannocci,
707 L., Mellin, C., Mesgaran, M. B., Moreno-Amat, E., Mormede, S., Novaczek, E., Oppel, S.,
708 Ortuño Crespo, G., Peterson, A. T., Rapacciuolo, G., Roberts, J. J., Ross, R. E., Scales, K.
709 L., Schoeman, D., Snelgrove, P., Sundblad, G., Thuiller, W., Torres, L. G., Verbruggen, H.,

710 Wang, L., Wenger, S., Whittingham, M. J., Zharikov, Y., Zurell, D., & Sequeira, A. M. M.
711 (2018). Outstanding challenges in the transferability of ecological models. *Trends in*
712 *Ecology & Evolution, 33*, 790-802.
713 Yoda, K., Naito, Y., Sato, K., Takahashi, A., Nishikawa, J., Ropert-Coudert, Y., Kurita, M. L., &
714 Maho, Y. (2001). A new technique for monitoring the behaviour of free-ranging Adélie
715 penguins. *Journal of Experimental Biology, 204*, 685–690.