

Identifying Movement States From Location Data Using Cluster Analysis

Authors: Van Moorter, Bram, Visscher, Darcy R., Jerde, Christopher L., Frair, Jacqueline L., and Merrill, Evelyn H.

Source: Journal of Wildlife Management, 74(3) : 588-594

Published By: The Wildlife Society

URL: <https://doi.org/10.2193/2009-155>

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.



Tools and Technology Note

Identifying Movement States From Location Data Using Cluster Analysis

BRAM VAN MOORTER,¹ *Department of Biology, Norwegian Institute of Technology in Trondheim, Høgskoleringen 5, NO-7491 Trondheim, Norway*

DARCY R. VISSCHER, *Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada*

CHRISTOPHER L. JERDE,² *Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada*

JACQUELINE L. FRAIR,³ *Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada*

EVELYN H. MERRILL, *Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada*

ABSTRACT Animal movement studies regularly use movement states (e.g., slow and fast) derived from remotely sensed locations to make inferences about strategies of resource use. However, the number of movement state categories used is often arbitrary and rarely inferred from the data. Identifying groups with similar movement characteristics is a statistical problem. We present a framework based on k-means clustering and gap statistic for evaluating the number of movement states without making a priori assumptions about the number of clusters. This allowed us to distinguish 4 movement states using turning angle and step length derived from Global Positioning System locations and head movements derived from tip switches in a neck collar of free-ranging elk (*Cervus elaphus*) in west central Alberta, Canada. Based on movement characteristics and on the linkage between each state and landscape features, we were able to identify inter-patch movements, intra-patch foraging, rest, and inter-patch foraging movements. Linking behavior to environment (e.g., state-dependent habitat use) can inform decisions on landscape management for wildlife.

KEY WORDS Alberta, *Cervus elaphus*, gap statistic, Global Positioning System (GPS), k-means, movement state.

Linking animal behavior to processes, such as dispersal, population dynamics, and habitat selection, is an important theme in ecology (Lima and Zollner 1996, Nathan et al. 2008). A major focus of the behavior-oriented approach has been to understand how landscape heterogeneity influences animal movements (Wiens 1989, Crist et al. 1992, Johnson et al. 2002). Whereas early animal movement studies primarily relied on directly observing animals in controlled arenas or animals that had narrow ranges of movement (Jones 1977, Roitberg and Mangel 1997), advances in technology of Global Positioning Systems (GPS) now allows fine-scale monitoring of movements with high accuracy for free-ranging animals. Technological advances in GPS have led to a surge in methods for characterizing movement patterns and identifying movement states (Johnson et al. 1992b, Fauchald and Tveraa 2003, Morales et al. 2004, Luque and Guinet 2007, Barraquand and Benhamou 2008).

One common approach, with roots in random walk theory (Kareiva and Shigesada 1983), uses sequential GPS locations to describe the animal's movement path in terms of distance moved between points (referred to as step lengths) and the turning angle between points (Turchin 1998). Based on changes in characteristics of these measures (e.g., differences in distribution of step lengths and turning angles), several studies have successfully identified multiple movement states, such as slow and fast states thought to reflect within and between patch movements or foraging and exploratory movements (Morales et al. 2004, Patterson

et al. 2008). State-space modeling provides a powerful means to model movement processes given multiple movement states and also can be used to explore the number of identifiable states using information criteria to select among models based on different numbers of states (Burnham and Anderson 2003; e.g., Morales et al. 2004). However, computational challenges associated with large numbers of parameters, which are needed for a realistic movement model, and assumptions involved in modeling of multistate movement processes are not trivial, especially for data with many variables and many states. Ideally, a technique for identifying the number of movement states would evaluate evidence in multivariate movement data for how many groupings are distinguishable without strong assumptions about the movement process itself. Such an exploratory analysis may be complementary to state-space modeling by providing initial parameter estimates associated with a set number of states.

We explored a multivariate clustering approach based on k-means for statistically identifying patterns of different groupings of movement characteristics, and we examined its use with movement data taken from elk (*Cervus elaphus*) in west central Alberta, Canada. We assumed no a priori underlying joint distributions, nor number of movement states, and we used the gap statistic (Tibshirani et al. 2001) to evaluate evidence for number of movement clusters.

METHODS

The purpose of cluster algorithms is to group observations with similar characteristics using multivariate information. Fisher (1958:789) stated the problem as follows: given n observations with d variables, find k similar groups. When there are 2 variables, clusters are identifiable using a scatter plot and by inspecting density of points (Fig. 1A). Clusters of similarly related observations have a high density and

¹ E-mail: bram.van.moorter@gmail.com

² Present address: Department of Biological Sciences, University of Notre Dame, P.O. Box 369, Notre Dame, IN 46556-0369, USA

³ Present address: Department of Environmental and Forest Biology, 250 Illick Hall, State University of New York College of Environmental Science and Forestry, Syracuse, NY 13210, USA

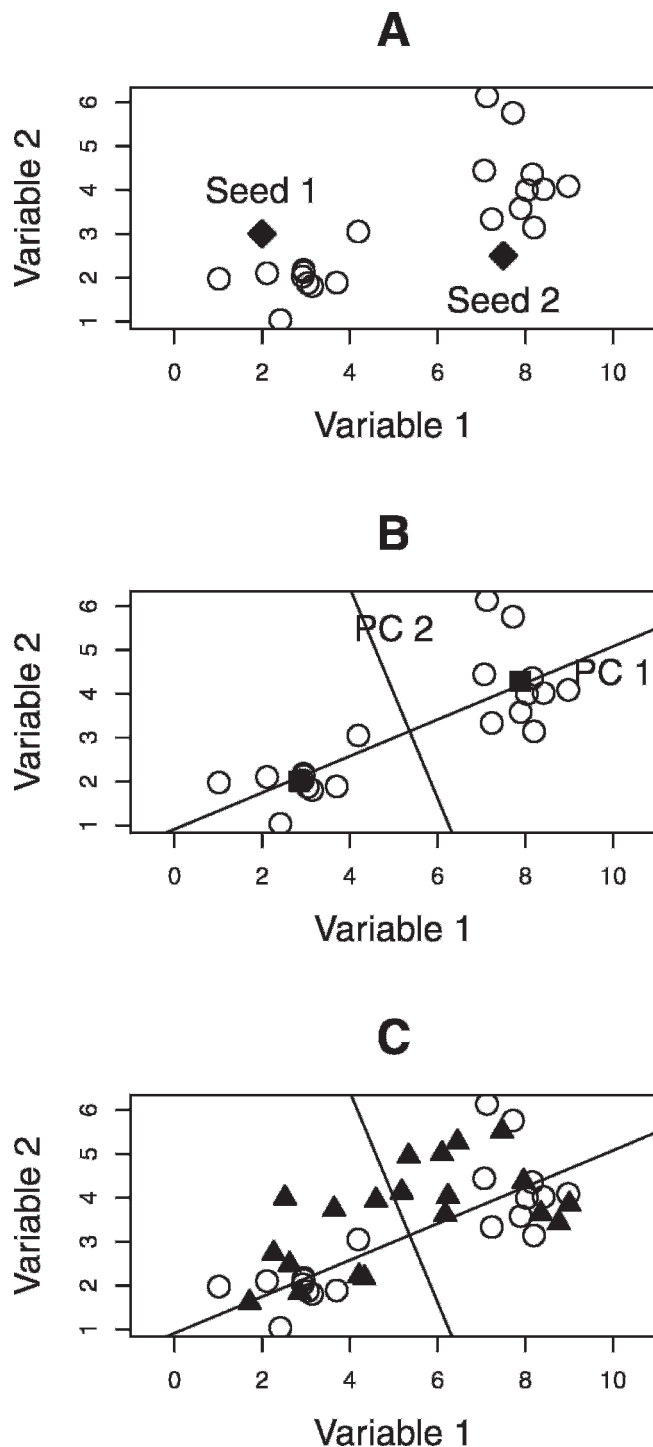


Figure 1. Illustration of k-means clustering (A, B) and gap statistic (B, C) using simulated data (i.e., variable 1 and 2). (A) Two-dimensional data (open circles) with initial seeds to be used for identifying center points of clusters (black diamonds) with k-means. (B) Center point of the clusters (no. clusters $k = 2$; black squares) after convergence of the k-means, and the principal components (PC1 and PC2) to be used for creating reference data. (C) Reference data (black triangles) needed to calculate the gap statistic.

separation between clusters shows low or no density. For the problem of identifying different movement states, we would have n observations (i.e., animal locations) each with d associated variables (e.g., step length and turning angle) where we attempt to identify k groups (e.g., movement

behaviors, modes, or states) based on movement characteristics.

One commonly used algorithm for clustering data is the k-means procedure (MacQueen 1967). This procedure starts by initializing the number of clusters, k , identified within the data. Once k is designated, k random, initial nodes, often referred to as seed points, are generated (Fig. 1A). Once the seeds are initialized, a multivariate distance measure (most often a squared Euclidean distance) is calculated from each data point to each seed. Observations are assigned to the seed with the shortest distance, which reflects similarity in the observations. Within-cluster means of all variables are then calculated. The vector of variable means then becomes the new seed point for each cluster. The process is iterated until membership of observations within clusters stabilizes (Fig. 1B). The k-means algorithm can be sensitive to initial seed points. One approach to solve this problem is to run the clustering routine multiple times with different initial seed points and compare results (both the no. clusters and the cluster to which each datum was assigned). Steinley (2006a) provided a comprehensive review of k-means clustering, and we refer the reader to Johnson and Wichern (1998) or Legendre and Legendre (1998) for a thorough treatment of k-means analysis.

Determining the variables important to identifying clusters is an important area of current research. Kaufman and Rousseeuw (1990) noted the addition of non-informative variables will confound underlying cluster structure. Building upon this observation, Lleti et al. (2004) developed a procedure to eliminate non-informative variables to investigate clustering when there are many variables available. However, as Johnson and Wichern (1998) suggested that the selection of variables to be included in the clustering is best justified by the process, in our case, the biology of movement variables related to between-patch and within-patch movements.

Studies have shown that the outcome of cluster analyses, including methods like k-means, are sensitive to distribution and range in values of the input (x) variables (Steinley 2006b, Yingqiu et al. 2007). Heavily skewed data can hide the cluster structure present in these data, whereas range differences among variables affect contributions from different variables to the clustering result. Clustering relies on multidimensional measurement of (Euclidean) distance; variables with a larger range will have a larger contribution to this distance measure (Kaufman and Rousseeuw 1990). Therefore, variables with a larger range will have an increased weight in cluster analyses. As a result, both data transformation and standardization have been recommended (Yingqiu et al. 2007). Range standardization is the recommended standardization procedure for clustering (Steinley 2006a):

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (1)$$

Due to the role of the minimum and maximum value of a variable in range standardization, it is crucial to inspect the

data for outliers (i.e., atypical data that are distant from the rest of the data) before performing this standardization.

It is critical in clustering observations into groups to determine number of groups present in the data. There are a few suggested methods for evaluating evidence for the number of clusters (Calinski and Harabasz 1974, Milligan and Cooper 1985). Recently, Tibshirani et al. (2001) proposed the gap statistic and found it correctly identified true number of clusters more regularly than other approaches when compared in simulation. The gap statistic (eq 2) provides a hypothesis testing framework that requires 4 steps. First, using the observed data, cluster analysis is performed for $k = 1, 2, 3 \dots n$. With the data clusters defined, a measure of dispersion of the observations is calculated by summing the within cluster squared Euclidean distances (i.e., $SS_{k,data}$ for a given k). Second, a number of reference data sets (β) are created by Monte Carlo simulation using the uniform distribution on the principal components (Fig. 1B, C). Principal components are used to ensure the reference data span the same space as observed data. Reference data sets are then clustered for $k = 1, 2, 3 \dots n$, and the $SS_{k,ref}$ is calculated for each reference data set.

$$\text{gap}(k) = E[\log(SS_{k,ref})] - \log(SS_{k,data}) \quad (2)$$

$$SE_k = \sqrt{\frac{1}{\beta} \sum_{b=1}^{\beta} \left(\log(SS_{k,ref}(b)) - \frac{1}{\beta} \sum_{b=1}^{\beta} \log(SS_{k,ref}(b)) \right)^2} \sqrt{1 + \frac{1}{\beta}} \quad (3)$$

Third, the standard error, SE_k (eq 3), from the reference data set is required to create a decision rule. Finally, the decision rule in the gap framework is to choose the smallest number of clusters, k , such that the gap statistic of the data with k groups is greater than or equal to the gap statistic of $k + 1$ clusters less the standard error of $k + 1$ clusters (eq 4).

$$\hat{k} = \min(k) : \text{GAP}(k) \geq \text{GAP}(k+1) - T \times SE_{k+1} \quad (4)$$

Tolerance T is analogous to setting the alpha level in the standard hypothesis testing framework, where increased tolerance is similar to selecting a smaller alpha rejection region. Tibshirani et al. (2001) used a tolerance of 1, but larger values of tolerance increase the strength of evidence required to include additional clusters (see Tibshirani et al. 2001 for full details and formulations). Tibshirani et al. (2001) demonstrated that the gap statistic performed well in detecting number of clusters when clusters are well separated; however, it was sensitive to the amount of overlap between clusters. Fortunately, the bias in sensitivity is such that it is likely to identify fewer clusters than there are in truth.

We illustrated this approach with data from elk in the Rocky Mountain foothills of west central Alberta, adjacent to Jasper and Banff national parks (Frair et al. 2005, Frair et al. 2007). The area was topographically diverse and predominately forested (>60%), making direct observation of an individual elk's movement behavior impractical. Forest harvesting in the area created small clear-cut patches of regenerating forest (hereafter cutblocks) throughout the forest matrix. These patches provided considerably more

foraging opportunities than the surrounding forest, making them attractive to elk (Visscher and Merrill 2009).

We used locations sampled every 2 hours from 15 June to 15 September 2002 for a resident female elk captured by aerial netgunning (University of Alberta Animal Care and Use Protocol no. 300201) and equipped with a GPS collar (GPS 2200, Lotek Wireless, New Market, ON, Canada). For more details on the data collection, see Frair et al. (2005). Throughout this 3-month period, 930 locations were recorded, with an 83% fix success rate. Using consecutive locations stored in the collar, we calculated the step length and turning angle for each movement segment (Turchin 1998, Jerde and Visscher 2005). We calculated all statistics only for 3 consecutive successful fixes to ensure comparable time intervals. Turning angles ranged from 0° to 180° , so they were always positive. The neck collar also contained an activity sensor that recorded number of times a tip switch was activated by an up and down (vertical) or side to side (horizontal) motion of the head during a preset 4-minute period directly before when the collar acquired an animal's location. Frequency of vertical or horizontal head movements ranged from 0–255 (max. count stored over 4 min). Step lengths and turning angles described movement trajectories (Turchin 1998); head movements have been shown to reflect feeding behavior (Adrados et al. 2003). We used both the measures of path trajectories (step length [SL] and turning angle [TA]) and head movements (vertical activity [ACT1] and horizontal activity [ACT2]) as variables in the cluster analysis because we expected they had the potential to distinguish behaviors reflecting within and between feeding patch movements.

We log-transformed both activity measures and step length to reduce positive skew. Furthermore, we standardized all variable values on their range (Steinley 2006a). Because we found no outliers (no data seemed outlying by visual inspection of the distribution, nor were any observations >3.3 SDs from the mean, which corresponds to a density of 0.001 in a normal distribution), we did not remove any data before data standardization.

To demonstrate the importance of data preprocessing, we first performed a cluster analysis on the raw data, these data after log transformation and then after range standardization. In subsequent analyses, we used the transformed and standardized data, which is the recommended strategy. Second, we replicated the k-means analysis 3 times to investigate the effect of variable inclusion on choice of number of groups using 1) all 4 movement measures, 2) only trajectory measures (i.e., SL and TA), and 3) only activity measures (i.e., ACT1 and ACT2). Finally, we performed one more analysis using all 4 measurements but with unequal weighting of the different variables. The use of differential weighting for further analysis of these data was suggested by the results from our previous analyses on the effect of variable inclusion. For each of these analyses, we used the gap statistic with 50 reference data sets (β in eq 3) and tolerance levels of 1 and 2. We conducted k-means cluster analyses 100 times with random seed points. All procedures for the gap framework we described are available

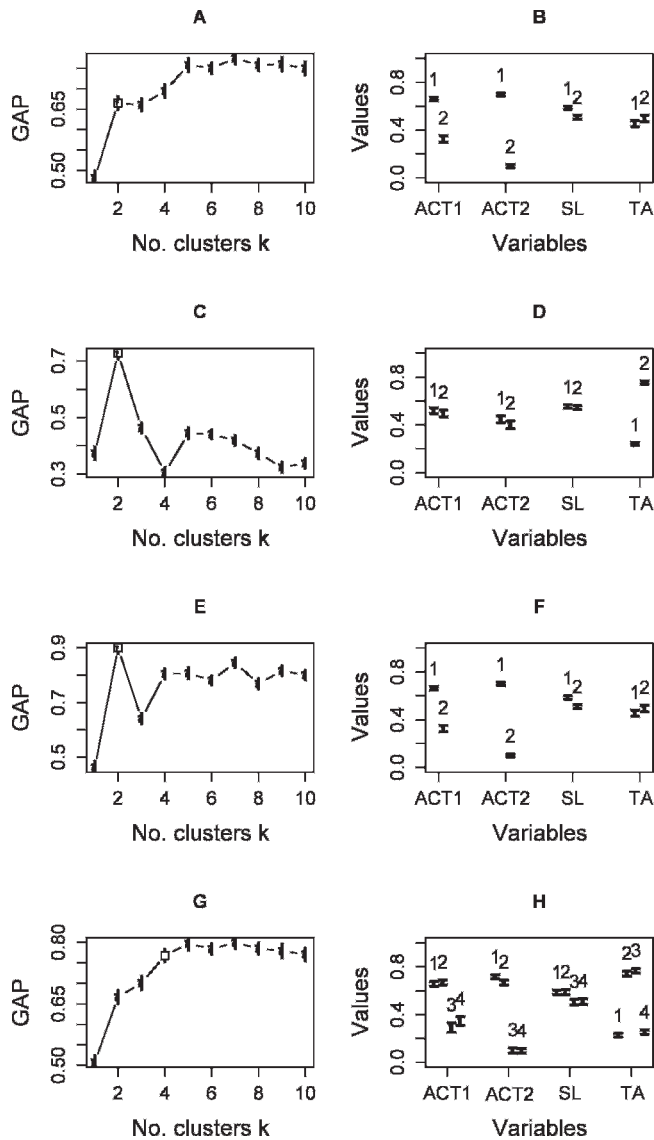


Figure 2. Gap statistic for different number of clusters and characteristics of the variables for the selected number of clusters for an elk during summer in Alberta, Canada. Panels on the left side (A, C, E, G) show the gap statistic with its standard error for 1 to 10 clusters. The selected number of clusters using a tolerance of 1 and 2 are marked with a black dot and a square, respectively (note that only in panel G a difference occurred between both tolerances with, respectively, 5 and 4 clusters selected). Panels on the right side (B, D, F, H) show for each cluster the mean and $2 \times$ standard error of the 4 clustering variables in our study (note that we did not necessarily use these variables all for the clustering itself): vertical and horizontal head movements (ACT1 and ACT2, respectively), step length (SL), and turning angle (TA). The cluster analysis in the top panels (A and B) was based on all 4 variables; the second row (C and D) on the trajectory measures (step length and turning angle); the third row (E and F) are the results of both head movement (i.e., activity) measures; and the bottom panels (G, H) result from all 4 measures with decreased weight of the activity measures.

in the open-source software for statistical computing and graphics R (R version 2.9.0, <<http://www.r-project.org/>>, accessed May 2009), using package clusterSim by Waleśiak and Dudek (version 0.36-3, <<http://keii.ue.wroc.pl/clusterSim>>, accessed May 2009). A modified version of the function index.Gap is available from B. Van Moorter's website (<<http://ase-research.org/moorter>>, accessed Nov 2009) together with our example data.

RESULTS

The raw data without any data preprocessing did not reveal any cluster structure; the gap statistic was maximum for one cluster (i.e., no cluster structure). After log transforming both activity measures and step lengths, we found a structure with 2 clusters. The 2 movement states differed only in their degree of directional persistence (mean TA: 43° versus 136° ; $t_{897} = -55$, $P < 0.001$). The larger range of the turning angles (from 0° to 180°) than the other measures (on a log-scale: 0 to 8.22 for the step lengths and 0 to 5.55 for both activity measures) explained the clustering based on turning angles. Figure 2A and B shows how range standardization of the different variables led to a distinctly different cluster structure (instead of on turning angles, the clustering is now mainly based on both activity measures).

Irrespective of whether we included only trajectory measures (SL and TA), only activity measures (ACT1 and ACT2), or all 4 variables in the k-means analysis, we found evidence for only 2 movement states at both levels of tolerance (Fig. 2A, C, E). However, when using all 4 measures, the gap statistic continued to increase for >2 clusters (Fig. 2A). Such an increase in gap values could indicate a more complex or even hierarchical structure in the data (Tibshirani et al. 2001). We used variable weighting to further explore the cluster structure in our data (see below).

Although we found agreement about the number of movement states among analyses with different input variables, we found also important differences in cluster assignment of the data. First, agreement between cluster membership based on trajectory measures and on all 4 measures were barely different from random (proportion agreement = 0.54; $\chi^2_1 = 3.6$, $P = 0.06$), whereas agreement between clusters of movement states from activity measures and from all input variables was nearly the same (proportion agreement = 1.00, $\chi^2_1 = 909.0$, $P < 0.001$). This indicated that when we analyzed all variables together, cluster assignment was driven primarily by activity measures and that movement behaviors related to head movement were more distinct than trajectory variables. Nonetheless, the pattern of turning angles differed among movement states based solely on trajectory data (Fig. 2D). One movement state ($n = 421$) was characterized by strong directional persistence, whereas the other ($n = 503$) by frequent direction reversals by elk.

The previous analyses suggested that the cluster structure in the trajectory measures (notably TA) was lost in the structure presented by both activity measures (see Discussion). This dominant effect of the activity measures could be explained by the high correlation between both activity measures ($r = 0.51$). Therefore, we continued our investigation of the cluster structure by decreasing the weight of both activity measures. To compensate for the variance shared by both variables ($r^2 = 0.26$), we decreased the weight of each activity measure by 0.13 (i.e., $r^2/2$). The reduced weight of the activity measures led to a more stable gap statistic pattern (see Fig. 2G). The value of the gap statistic increased until we identified 4 or 5 clusters, after

Table 1. Characteristics of the 4 movement states in an elk trajectory during summer in Alberta, Canada. The movement states were distinguished by the weighted k-means clustering on all 4 variables (i.e., activity measures [ACT1 and ACT2], step length [SL], and turning angle [TA]). For each state, the mean and standard error of each variable, proportion and standard error during twilight and within cutblocks, and total number of observations are shown.

State	State 1		State 2		State 3		State 4	
	\bar{x}	SE	\bar{x}	SE	\bar{x}	SE	\bar{x}	SE
ACT1	57.89	3.29	56.87	3.33	17.09	2.46	18.79	2.18
ACT2	71.95	3.4	58.33	3.36	2.16	0.41	1.77	0.26
SL	309.54	23.18	317.48	26.46	180.22	16.71	206.64	20.62
TA	41.23	1.51	134.16	1.77	139.19	1.67	46.56	1.74
Twilight	0.6	0.03	0.6	0.03	0.44	0.04	0.42	0.03
Cutblock	0.5	0.03	0.65	0.03	0.62	0.04	0.49	0.03
<i>n</i>	284		225		189		231	

which it stabilized. Whether 4 or 5 clusters were selected depended on the choice of tolerance value (2 or 1, respectively). Figure 2H and Table 1 summarize the movement characteristics of the 4 movement state clusters.

DISCUSSION

We showed the strength of the k-means approach with gap statistic as a framework for identifying movement states that allows movement data to be explored in a hypothesis testing framework, using a well formulated statistical criterion (i.e., gap statistic) to identify the number of movement states. From our results it also was clear that the outcome of the k-means was sensitive to distribution of the data, because without data preprocessing we identified no clusters in our data. It is therefore important to carefully inspect data distribution before undertaking a cluster analysis. Equally important is a comparable measurement scale for all variables. When large scale-differences exist, those variables with the largest range will contribute more to Euclidean distances among points, resulting in an unintended weighting of the variables. Intentional variable weighting can be useful when there are statistical (e.g., Steinley and Brusco 2008) or biological reasons to assign more weight in the clustering approach to specific variables. We demonstrated this by decreasing the weight of the head movement data to counter the effect of having included 2 such highly correlated measures in the analysis.

Using the approach we presented, the researcher can explicitly test for the number of distinctive states and can adjust the strength of evidence required to add clusters by changing the tolerance used in the gap statistic. Tolerance T in equation 4 determines the number of standard errors used to assess the change in gap statistic with increasing number of clusters (increasing T leads to a more conservative choice). In the analysis with weighted activity measures, we found some uncertainty regarding the number of clusters (i.e., no. clusters 4 vs. 5). Therefore, one should carefully inspect changes in gap statistic with number of clusters, instead of exclusively relying on an arbitrary tolerance value. If there is a clear peak (as in Fig. 2D), then low uncertainty in the number of clusters exists, with all clusters being clearly separated. However, a gap statistic increasing beyond the selected number of clusters (e.g., Fig. 2B) suggests a more complex cluster structure (Tibshirani et al. 2001). In

this situation, the researcher should investigate cluster structure further. Variable weighting, as we did, can be used with k-means cluster analysis. Alternatively, hierarchical clustering can be used to investigate hierarchical cluster organization (Kaufman and Rousseeuw 1990).

In the context of the analysis of GPS data, we must address 2 important issues: sampling interval and missing fixes. First, the sampling interval determines the temporal scale of the data (Nathan et al. 2008); therefore, it is important to select an interval appropriate for the research question. When travel distances become too short at short sampling intervals, inaccurate movement assessment may arise due to GPS error (Hurford 2009). With increasing sample intervals behaviors become more aggregated. For example, elk movement sampled at a daily interval aggregates rest and movement states (found in our data at 2-hr intervals, see below); instead, these data showed behaviors at a larger temporal scale, like encamped and exploratory states (Morales et al. 2004). Therefore, the error in the data and the temporal scale of the behavior of interest will dictate sampling interval. Second, missing fixes can lead to bias in estimates of habitat selection (e.g., Frair et al. 2004, Bourgoin et al. 2009). A similar problem can arise for movement states. In contrast with habitat selection studies, it is not possible to use static collars to assess movement related fix bias. However, when movement states are temporally autocorrelated, interpolation can be used to estimate fix bias related to movement behavior (similar to Bourgoin et al. 2009 for habitat selection).

One of the interesting possibilities offered by distinguishing movement states of animals in ecology is to investigate animal responses to the environment. As a simple illustration, we compared occurrence of the 4 movement states in cutblocks versus non-cutblocks, and during crepuscule (morning and evening) versus the rest of the 24-hour period. As discussed above, cutblocks offer important foraging opportunities for elk, and the major activity peaks for elk occur at dusk and dawn (Merrill 1991, Olsson et al. 2007). We found important differences among movement states with respect to their occurrence within cutblocks ($\chi^2_3 = 18.64, P < 0.001$) and during twilight ($\chi^2_3 = 25.58, P < 0.001$; see Table 1). The first movement state seemed to correspond to movements between cutblock patches, characterized by high activity and high directional

persistence. Alternatively, the second movement state was most likely composed of foraging movements, which combined frequent head movements with frequent direction changes. Not surprisingly, these foraging movements occurred primarily during twilight within cutblocks. The third movement state was characterized by few up and down head movements; therefore, feeding was unlikely to have occurred during this phase. The low directional persistence may have resulted from GPS-error during an animal's resting phase (Hurford 2009). In further support for this interpretation, we found this movement state occurred less during the main activity periods (i.e., twilight) and more within cutblocks, where high visibility may aid early detection of approaching predators. In contrast, little head movements with high directional persistence were typical of the final movement state. A clear biological interpretation for this state is not straightforward. This final movement state occurred less during twilight and less within cutblocks. We can speculate that this state consisted of between-patch movements without foraging, whereas occasional foraging occurred in the first movement state.

Hence, presence or absence of foraging during between patch movements would explain the amount of head movements. Thus, the strong linkage between movement states and landscape features corroborates the success of our technique in identifying biologically plausible states.

Ultimately, the usefulness of groupings of movement characteristics derived from the k-means approach is dependent on their correspondence to biological mechanisms and processes that produced the clusters. Indeed, the interpretation of different movement states follows from the descriptive characteristics of these states. Cluster analysis investigates the structure and patterns in the data. If behavioral observations were taken while the GPS collar was deployed, then once the data were clustered, the researcher could explore correspondence of observed behaviors to the different clusters.

An important asset of the k-means approach combined with the gap statistic is the ability to include multivariate information. Although previous methods (Johnson et al. 1992a, b; Morales et al. 2004) have included turning angles and step lengths, they are not conducive to using a range of variables, such as head movements we used in our elk example. Although it is possible to include other variables in state-space models, doing so results in increasing model complexity and computational challenges. With advancements in animal collar technology, we expect additional variables, such as physiological responses (e.g., heart rate), abiotic conditions (e.g., ambient temp.), to become particularly useful in remotely identifying behaviors such as inactivity—thermogenesis or predatory flight responses. However, care should be taken in selecting variables to be included, because nondiscriminating variables can clutter and hide cluster structure present in the data.

The k-means approach with the gap statistic allows the researcher to explicitly investigate the number of movement states in a transparent hypothesis testing framework without assumptions about the underlying movement process. In

previous methods, the number of states was not an explicit part of the research hypothesis (e.g., Fauchald and Tveraa 2003) or was identified using a parametric approach with assumptions regarding the movement process (e.g., Morales et al. 2004). It is important to note that these different approaches are not exclusive; they are complementary instead. For example, the first passage time (Fauchald and Tveraa 2003) can easily be included in a k-means cluster analysis in addition to or instead of other movement characteristics, whereas knowledge gained by investigating patterns in movement data will aid development of models for movement processes (Patterson et al. 2008).

Management Implications

For many management applications, knowledge of animal behavior in relation to its environment is important. For example, if the monitored organism forages in limited habitat and this behavior is identifiable from the GPS data, then identifying these presumably critical areas could have consequences for management actions that may disrupt critical forage habitat. Where direct observation is difficult or impossible, remotely sensed data can provide information on behavior states; however, indirect observation of behavior requires linking behavioral states to movement and activity metrics. The k-means clustering with the gap-statistic can offer a defensible approach to infer these behavioral states based on directly or remotely sensed data.

Acknowledgments

We thank M. Lewis, A. Hurford, H. Beyer, and M. Taper for helpful comments and suggestions. B. Van Moorter was financially supported by the Norwegian University of Science and Technology and the Norwegian Research Council's PredClim grant to B.-E. Saether. C. L. Jerde was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery to M. Lewis, NSERC Industrial Scholarship in partnership with Weyerhaeuser to D. R. Visscher, National Science Foundation (0078130), Rocky Mountain Elk Foundation, and the Alberta Conservation Association grants to E. H. Merrill and J. L. Frair. Suggestions by S. McCorquodale and 2 anonymous reviewers greatly improved the manuscript.

LITERATURE CITED

- Adrados, C., H. Verheyden-Tixier, B. Cargnelutti, D. Pepin, and G. Janeau. 2003. GPS approach to study fin-scale site used by wild red deer. *Wildlife Society Bulletin* 31:544–552.
- Barraquand, F., and S. Benhamou. 2008. Animal movements in heterogeneous landscapes: identifying profitable places and homogeneous movement bouts. *Ecology* 89:3336–3348.
- Bourgoin, G., M. Garel, D. Dubray, D. Maillard, and J.-M. Gaillard. 2009. What determines global positioning system fix success when monitoring free-ranging mouflon? *European Journal of Wildlife Research* 55:603–613.
- Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a Practical information-theoretical approach. Second edition. Springer-Verlag, New York, New York, USA.
- Calinski, R. B., and J. Harabasz. 1974. A dendrite method for cluster analysis. *Communications in Statistics* 3:1–27.
- Crist, T., D. Guertin, J. Wiens, and B. Milne. 1992. Animal movement in heterogeneous landscapes—an experiment with Eleodes Beetles. *Functional Ecology* 6:536–544.

- Fauchald, P., and T. Tveraa. 2003. Using first passage time in the analysis of area restricted search and habitat selection. *Ecology* 84:282–288.
- Fisher, W. 1958. On grouping for maximum homogeneity. *Journal of the American Statistical Association* 53:789–798.
- Frair, J., E. H. Merrill, D. Visscher, D. Fortin, and J. Morales. 2005. Scales of movement by elk in response to heterogeneity in forage resources and predation risk. *Landscape Ecology* 20:273–287.
- Frair, J., S. E. Nielsen, E. H. Merrill, S. R. Lele, M. S. Boyce, R. H. M. Munro, G. B. Stenhouse, and H. L. Beyer. 2004. Removing GPS collar bias in habitat selection studies. *Journal of Applied Ecology* 41:201–212.
- Frair, J. L., E. H. Merrill, J. R. Allen, and M. S. Boyce. 2007. Know thy enemy: experience affects elk translocation success in risky landscapes. *Journal of Wildlife Management* 71:541–554.
- Hurford, A. 2009. GPS measurement error gives rise to spurious 180° turning angles and strong directional biases in animal movement data. *PLoS ONE* 4: e5632.
- Jerde, C., and D. Visscher. 2005. GPS measurement error influences on movement model parameter estimation. *Ecological Applications* 15:806–810.
- Johnson, A., B. Milne, and J. Wiens. 1992a. Diffusion in fractal landscapes: simulations and experimental studies of tenebrionid beetle movements. *Ecology* 73:1968–1983.
- Johnson, A., J. Wiens, B. Milne, and T. Crist. 1992b. Animal movements and population dynamics in heterogeneous landscapes. *Landscape Ecology* 7:63–75.
- Johnson, C., K. Parker, D. Heard, and M. Gillingham. 2002. Movement parameters of ungulates and scale-specific responses to the environment. *Journal of Animal Ecology* 71:225–235.
- Johnson, R., and D. Wichern. 1998. *Applied multivariate statistical analysis*. Prentice-Hall Inc., Upper Saddle River, New Jersey, USA.
- Jones, R. 1977. Movement patterns and egg distribution in cabbage butterflies. *Journal of Animal Ecology* 46:195–212.
- Kareiva, P. M., and N. Shigesada. 1983. Analyzing insect movement as a correlated random walk. *Oecologia* 56:234–238.
- Kaufman, L., and P. Rousseeuw. 1990. *An introduction to cluster analysis*. Wiley, New York, New York, USA.
- Legendre, P., and L. Legendre. 1998. *Numerical ecology*. Elsevier Science, Amsterdam, The Netherlands.
- Lima, S., and P. Zollner. 1996. Towards a behavioral ecology of ecological landscapes. *Trends in Ecology and Evolution* 11:131–135.
- Lleti, A., M. Ortiz, L. Sarabia, and M. Sanchez. 2004. Selecting variables for k-means cluster analysis by using a genetic algorithm that optimises the silhouettes. *Analytica Chimica Acta* 515:87–100.
- Luque, S. P., and C. Guinet. 2007. A maximum likelihood approach for identifying dive bouts improves accuracy, precision and objectivity. *Behavior* 144:1315–1332.
- MacQueen, J. 1967. Some methods for classification and analysis of multivariate observations. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability* 1:281–297.
- Merrill, E. H. 1991. Thermal constraints on use of cover types and activity time of elk. *Applied Animal Behaviour Science* 29:251–267.
- Milligan, G. W., and M. C. Cooper. 1985. An examination of procedures for determining the number of clusters in a data set. *Psychometrika* 50:159–179.
- Morales, J., D. Haydon, J. Frair, K. Holsinger, and J. Fryxell. 2004. Extracting more out of relocation data: building movement models as mixtures of random walks. *Ecology* 85:2436–2445.
- Nathan, R., W. M. Getz, E. Revilla, M. Holyoak, R. Kadmon, D. Saltz, and P. E. Smouse. 2008. A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences USA* 105:19052–19059.
- Olsson, P. M. O., J. J. Cox, J. L. Larkin, D. S. Maehr, P. Widen, and M. W. Wichrowski. 2007. Movement and activity patterns of translocated elk (*Cervus elaphus nelsoni*) on an active coal mine in Kentucky. *Wildlife Biology in Practice* 3:1–8.
- Patterson, T. A., L. Thomas, C. Wilcox, O. Ovaskainen, and J. Matthiopoulos. 2008. State-space models of individual animal movement. *Trends in Ecology and Evolution* 23:87–94.
- Roitberg, B., and M. Mangel. 1997. Individuals on the landscape: behavior can mitigate landscape differences among habitats. *Oikos* 80:234–240.
- Steinley, D. 2006a. K-means clustering: a half-century synthesis. *British Journal of Mathematical and Statistical Psychology* 59:1–34.
- Steinley, D. 2006b. Profiling local optima in K-means clustering: developing a diagnostic technique. *Psychological Methods* 11:178–192.
- Steinley, D., and M. J. Brusco. 2008. A new variable weighting and selection procedure for k-means cluster analysis. *Multivariate Behavioral Research* 43:77–108.
- Tibshirani, R., G. Walther, and T. Hastie. 2001. Estimating the number of clusters in a dataset via the gap statistic. *Journal of the Royal Statistical Society B* 63:411–423.
- Turchin, P. 1998. *Quantitative analysis of movement*. Sinauer, Inc., Sunderland, Massachusetts, USA.
- Visscher, D. R., and E. H. Merrill. 2009. Temporal dynamics of forage succession for elk at two scales: implications of forest management. *Forest Ecology and Management* 257:96–106.
- Wiens, J. 1989. Spatial scaling in ecology. *Functional Ecology* 3:385–397.
- Yingqiu, L., L. Wei, and L. Yun-Chun. 2007. Network traffic classification using k-means clustering. *Second International Multi-Symposiums on Computer and Computational Sciences* 2007:360–365.

Associate Editor: McCorquodale.