Quantifying Animal Trajectories Using Spatial Aggregation and Sequence Analysis: A Case Study of Differentiating Trajectories of Multiple Species

Peng Gao, John A. Kupfer, Xi Zhu, Diansheng Guo

Department of Geography, University of South Carolina, 709 Bull Street, Columbia, SC 29208, USA

The increasing availability of telemetry data with high spatial and temporal resolution promises to greatly advance scientific understandings of the movement patterns of individual organisms across space and time. The amount of data provided by such methods, however, can be challenging to analyze and interpret. In this study, we present a new approach for analyzing animal movements that aggregates telemetry locations into spatial clusters and extracts the information from sequences formed by individuals passing through these spatial clusters. We applied this integrated approach of spatial aggregation and sequence analysis to quantify and compare trajectories of cattle (Bos taurus), mule deer (Odocoileus hemionus), and elk (Cervus elaphus) tracked by automated telemetry at the Starkey Experimental Forest and Range in northeastern Oregon, USA. Our approach effectively differentiated movement patterns of the three species. It provides a useful mean of quantifying movement patterns of species in a landscape.

Introduction

Animal movement is a fundamental process that determines the fate of individual organisms, the structure and dynamics of biological populations, and the nature of species interactions and biological community assembly (Nathan et al. 2008; Miller 2012). The increasing availability of telemetry data with high spatial and temporal resolution promises to greatly advance the scientific understanding of how spatial and temporal factors impact the movements of individual organisms and thereby affect species persistence in heterogeneous landscapes. However, the degree to which such data may yield important insights is dependent on the development of appropriate and effective means of analysis that capture important aspects of animal movement.

The earliest analyses of animal movement using data collected by GPS or telemetry involved the estimation of animal home range, the "area traversed by the individual in its

Correspondence: Peng Gao, Department of Geography, University of South Carolina, 709 Bull Street, Columbia, SC 29208, USA
e-mail: gaop@mailbox.sc.edu

1

normal activities of food gathering, mating, and caring for young" (Burt 1943: p 351). Since then, various approaches have been developed for characterizing aspects of animal movement (e.g., the relative angles between successive moves, the distance between successive relocations: Calenge, Dray, and Royer-Carenzi (2009)) and, less commonly, analyzing the distribution of descriptive movement parameters (e.g., by comparing metrics computed for observed trajectories with those expected based on simulated trajectories: Miller (2012)). Few studies, however, have considered the potentially useful information provided by the *sequences* of animal movement through sampled positions, particularly as a means of quantifying and comparing movement patterns of different species across landscapes.

In this study, we provide a new approach to analyzing trajectories of animal movement to quantify the movement patterns of different species. Our approach aggregates telemetry locations into groups of points (spatial clusters) and then considers trajectories as sequences passing through these spatial clusters. Valuable information including spatial and aspatial properties of these trajectories can then be extracted from these sequences. We specifically tested the hypothesis that these properties can capture and discriminate the movement characteristics of different species by classifying the trajectories of different species using these properties. Finally, we compared the performance of our method with another trajectory classification method that used the same data set.

Related work

Movement pattern analysis, which has been developing alongside time geography, provides an efficient means of extracting useful information and recognizing patterns in movement data sets expressed as locations (typically x, y coordinates) collected at discrete time intervals (Long and Nelson 2013). Insights into animal movement patterns and processes have come not just from the field of biology but from methods developed in fields such as computer science, GIS, and geovisualization to analyze the movement of a wider range of objects. In the visual analytics of movement, Andrienko, Andrienko, and Gatalsky (2000) designed a collection of widgets connected to dynamic map displays that allows users to investigate instant movement patterns by selecting a time moment or to explore the dynamic of movement patterns in time intervals that the users set. They applied the tool to telemetric observations of stork migration to Africa in autumn 1998 and back to Europe in spring 1999. By animating the routes of storks in different time intervals, they found the routes looked like "worms" crawling on the maps and that these "worms" captured dynamic characteristics of movement (e.g., variation of speed in different time intervals). The authors argued that such dynamic displays are superior to an ordinary animated presentation in capturing these dynamic characteristics. Such observations have been more fully developed through the computation of parameters that describe the basic unit of an individual movement trajectory (a "step") in terms of its length and orientation in space, for example, the distance, speed or angle between the coordinates of the two relocations defining the step (Brillinger et al. 2004; Calenge, Dray, and Royer-Carenzi 2009).

As with other types of trajectories, the large amount of data associated with telemetry data for animal movements exerts high demands on the methods of visual analytics, pattern detection and recognition. A variety of computational algorithms have been proposed to reduce the complexity inherent in large data sets, improve the efficiency of trajectory visualization, and facilitate pattern recognition. Clustering is a widely used technique for these purposes in analyzing and comparing trajectories. In much the same way that nonspatial objects can be

clustered on the basis of similarity measures, indices describing the path similarity of trajectories can be represented by measures such as the Hausdoff distance (Huttenlocher, Klanderman, and Rucklidge 1993), the Longest Common SubSequence (LCSS) (e.g., Vlachos, Gunopoulos, and Kollios 2002; Cheriyadat and Radke 2008) and Dynamic Time Wrapping (DTW) (e.g., Usabiaga et al. 2007). While popular, these measurements have documented weaknesses as well: the Hausdoff distance cannot consider chronological order of the points (Zhang, Huang, and Tan 2006), the LCSS is sensitive to the threshold selected to determine whether two elements match, and the DTW is not robust to noise (Chen, Özsu, and Oria 2005).

Alternatively, another set of algorithms reduces the complexity of large data sets by grouping telemetry locations into subsets such as spatial clusters (e.g., Andrienko and Andrienko 2011) or regions (e.g., Guo, Liu, and Jin 2010). By aggregating movement between locations, users are able to obtain an overall view of the spatial and temporal distribution of multiple movements and to uncover potential patterns (Andrienko and Andrienko 2013). This kind of approach, which is less influenced by the geometry of movement paths, can detect hidden patterns in the movement data and might be particularly suitable for animal movements because organisms, unlike vehicles, usually exhibit free movement. Verhein and Chawla (2008) established multiple spatiotemporal association rules to detect *stationary* and *high traffic* regions and described how mobile objects move between regions over time. When they applied their method to the movement of caribou in northern Canada, group and individual movements were distinguished by different regions where the movement occurred.

Beyond methods that can quantify aspects of individual movement, there is a need for analytical approaches that can statistically differentiate patterns of movement and habitat utilization for predefined groups of individuals. Comparisons of the spatial ecologies of multiple species (their movements and home ranges) can be used to demonstrate contrasting patterns of space and habitat use (e.g., Ryan et al. 2008; Jaeger and Cobb 2012). However, while approaches that examine the joint-space use between groups by calculating an index of home range overlap or volume of intersection using utilization distributions can be helpful (e.g., Millspaugh et al. 2004; Fieberg and Kochanny 2005), they are largely descriptive and may fail to fully capitalize on the wealth of information provided by most modern telemetry data sets. Methods that can statistically contrast the movement of multiple individuals belonging to more than one group may provide a better means for comparing species movements across a landscape. Such approaches could also potentially contribute to a well-recognized need for methods that can be used to quantify interactions between animals (Long et al. 2014).

Data and methods

Study site and data

The telemetry data used in this study were collected in June 1995 at the Starkey Experimental Forest and Range (EFR), which is located in the Blue Mountains ca. 50 km southwest of LaGrande, Oregon, USA. The study area is managed by the U.S. Forest Service and includes 78 km^2 of forests and mountain meadows enclosed by a 2.4 m high ungulate-proof fence that restricts the immigration of large herbivores. Forests are dominated by Douglas fir (*Pseudotsuga mensiesii*), ponderosa pine (*Pinus ponderosa*), grand fir (*Abies grandis*), and lodgepole pine (*Pinus contorta*) and cover about 70% of the study area. The remaining 30% is occupied by grasslands and meadows dominated by bunchgrasses (*Festuca idahoensis*r, *Poa sandbergii*,
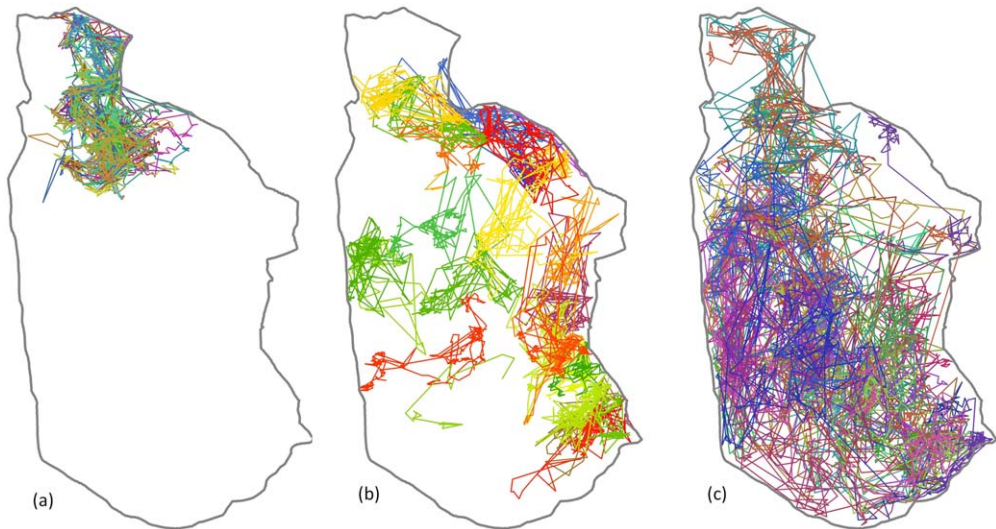
Figure 1. Trajectories of (a) cattle, (b) mule deer, and (c) elk at the Starkey Experimental Forest and Range in June 1995. Different colors represent trajectories of individual animals.
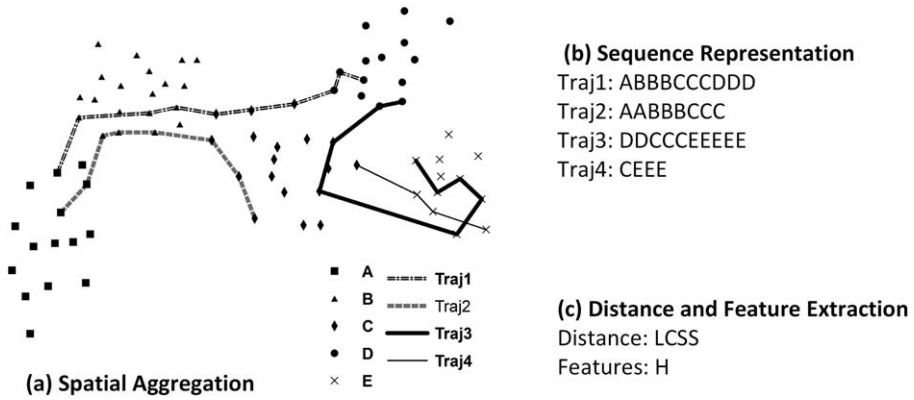
and *Agropyron spicatum*). Elevation ranges between 1,100 and 1,500 m, and the average annual precipitation from 1993 to 1996 was 636 mm (Rowland et al. 1998; Coe et al. 2001).

The data comprise 14,990 *x-y* coordinates for 34 cattle (*Bos taurus*), 30 mule deer (*Odocoileus hemionus*), and 38 elk (*Cervus elaphus*) that were tracked by an automated telemetry system based on a rebroadcast civilian long-range navigation (LORAN-C) technology. Temporal resolution of the movement data (which are described in greater detail in Rowland et al. 1998) is 45–90 min, and the spatial error is ca. 200 m. These data were also examined by Lee, Han, and Whang (2007) for trajectory clustering and Lee et al. (2008) for trajectory classification.

The three species chosen for our analysis have distinct habitat preferences, foraging ecologies and movement restrictions. Cattle avoid steep slopes and tend to select places in lower elevations compared to mule deer and elk. During the time that the data were collected, cattle inhabited the northeast corner of the Starkey EFR, and their movement was restricted by barbed-wire fences (Coe et al. 2001). Elk used both mesic and logged forests while mule deer avoid xeric grasslands (Stewart et al. 2002). Although mule deer and elk ranged freely within the study area without any movement restrictions, individual movements tended to be more highly concentrated along the eastern (mule deer) or western (elk) edge of the study area for the different species (Fig. 1).

Data analysis

Our method has two steps. In the first step, telemetry locations are aggregated into groups of points termed spatial clusters (Fig. 2a). This step is to reduce data complexity and achieve a practical computational efficiency by reducing the large number of telemetry points to a manageable number of spatial clusters. After aggregation, trajectories are considered as sequences passing through these spatial clusters (Fig. 2b). The second step is to extract spatial or aspatial properties from these sequences (Fig. 2c). The construction of spatial clusters is first introduced below.

(b) Sequence Representation
Traj1: ABBBCCCDDD
Traj2: AABBBCCC
Traj3: DDCCCEEEEE
Traj4: CEEE

(c) Distance and Feature Extraction
Distance: LCSS
Features: H

(a) Spatial Aggregation

Figure 2. Illustration of the steps used to define and quantify animal movement sequences: (a) Telemetry locations are aggregated into spatial clusters A to E; (b) Trajectories are represented as sequences passing through the spatial clusters; (c) Distances and/or features are calculated or extracted from these sequences. See the text for the explanation of spatial clusters, the Longest Common SubSequence (LCSS) and entropy (H).

Radio telemetry data take the form of time-indexed spatial locations of individual animals. Animal movement can be represented by a set of trajectories $T = \{T_i\}$ $(1 < i < n)$ of n individuals, wherein each T_i comprises m_i points $P_{ij} = \{<s_{ij}, t_{ij}>\}$ $(1 < j < m_i)$, and s_{ij} and t_{ij} represent the spatial coordinates and time of an individual telemetry location.

To aggregate points into spatial clusters, we adopt the spatial clustering method introduced in Guo et al. (2012), which involves three components: defining the similarity of a pair of points (*A* and *B*) based on their Shared Nearest Neighbors (SNN), building a Delaunay Triangulation (DT) to enable efficient clustering, and a constrained hierarchical spatial clustering procedure to derive clusters of similar sizes. Let *NN* (*A, k*) and *NN* (*B, k*) be the *k* Nearest Neighbors (NN) of *A* and *B* including itself in space, respectively, while SNN is the number of points shared by both NN (*A, k*) and NN (*B, k*). The similarity of *A* and *B* in *k* nearest neighbors is the ratio between the intersection of *NN* (*A, k*) and *NN* (*B, k*) (i.e., *SNN* (*A, B, k*)) and the union of *NN* (*A, k*) and *NN* (*B, k*):

$$\text{Similarity } (A, B) = (NN \ (A, \ k) \ \cap \ NN \ (B, \ k))/(NN \ (A, \ k) \ \cup NN \ (B, \ k)) \quad (1)$$

Before clustering, a DT is built for all points (Guo, Peuquet, and Gahegan 2003). A DT for a set *P* of points in the Euclidean plane is a triangulation DT (*P*) such that no point in P is inside the circumcircle of any triangle in DT (*P*) (Tsai 1993). The construction of a DT efficiently reduces the time complexity of finding the *k* nearest neighbor points of each point and speeds up the clustering (see below for explanation). The clustering step is a bottom–up procedure based on the single-linkage hierarchical clustering, which recursively merges the nearest clusters. It defines the distance between two clusters as the shortest edge that connects the two clusters, the clustering procedure only needs to evaluate the DT edges (instead of all pair-wise edges among the points). This significantly speeds up the clustering process and achieves a time complexity of O (*n* log *n*), where *n* is the number of telemetry locations (Guo et al. 2012). Beginning with each point as a single cluster, clusters are merged according to their descending order of similarity. The clustering process continues until the number of points contained in each cluster reaches a user-defined minimum number of points (*q*) (See Algorithm 1 in

Appendix and more details in Guo et al. (2012)). The boundaries of the spatial clusters are delineated by dissolving the Thiessen polygons of the points in the clusters.

The clustering procedure is governed by two parameters, the number of nearest neighbors (k) and the minimum size of clusters (q). When k is held constant and q varies, the structure of the dendrogram does not change, but an increase or decrease of q yields clusters at higher or lower levels of the hierarchy. When q is held constant, changing k impacts the smoothing effect. A larger value of k produces a stronger smoothing effect because the similarity of two points is considered in a larger neighborhood. The SNN based clustering has two merits (Guo et al. 2012): (1) unlike other methods, SNN is generally not biased toward producing clusters of a particular shape (e.g., k-means clusters are likely to be circles), and (2) it is adaptive to the uneven distribution of points over space and able to find more clusters in areas of high point density and fewer clusters where point density is low.

After the telemetry points are grouped into spatial clusters, each telemetry point is represented by the spatial cluster to which it belongs. Therefore, each trajectory that consists of telemetry points is considered as a sequence passing through the spatial clusters. We performed sequential data analysis to quantify the movement patterns of the three species considering these trajectories as sequences passing through the spatial clusters. Various properties can be extracted from the sequences of spatial clusters through which trajectories pass (Gabadinho et al. 2011). The longest common subsequence (LCSS), which measures spatial similarity between pairs of trajectories, and entropy (H), which measures the diversity of spatial clusters that a trajectory passes through, were used in this study. Both are described below.

Before describing how to determine the LCSS, we introduce the concept of a Subsequence (Gabadinho et al. 2011). Let x be a string that contains n characters. In this case, x represents a trajectory and the n characters indicate the spatial clusters representing each of n telemetry locations that the trajectory passes through. A subsequence v of x is formed by taking away $0 \leq k \leq n$ characters from x. If $k = 0$, v is x itself. If $k = n$, v is empty (i.e., the length of v is zero). Formally, $v = v_1 v_2 \ldots v_p$ is a subsequence of x, if there exists strings $u_1, u_2, \ldots, u_{p+1}$ that $u_1 v_1 u_2 v_2 \ldots v_p u_{p+1} = x$ (any u_1 to u_{p+1} could be empty). In other words, characters in v are not necessarily adjacent to the next one as they are in x, but the characters in v must have the same sequence as they are in x.

The LCSS for strings of x and y is the longest subsequence of both x and y. To make the LCSS of different lengths of strings (i.e., trajectories containing different number of telemetry locations) comparable, the length of the longest subsequence is standardized by the geometrical mean of the two sequence lengths (Studer et al. 2011). To facilitate visualization and comparisons with other measures, LCSS can also be converted to a dissimilarity measure by subtracting the standardized LCSS from 1 (equation 2). Thereafter, LCSS is a similarity measure, while LCSS distance (matrix) refers to a dissimilarity measure:

$$\text{LCSS distance} = 1 - \frac{\text{LCSS}\ (x,\ y)}{\sqrt{|x||y|}} \tag{2}$$

where LCSS $(x,\ y)$ is the LCSS for strings of x and y, and $|x|$ and $|y|$ are their length.

The entropy (H) of the string x that has m distinct characters is defined as:

$$H = -\sum_{i=1}^{m} p_i \log p_i \tag{3}$$

where p_i is the proportion of occurrences of the ith characters in x (Gabadinho et al. 2011).

Here, each character represents a spatial cluster, and x indicates the sequence of the spatial clusters a trajectory passes through. Entropy is considered as a measure of the diversity of spatial clusters that a trajectory passes through, with a high entropy value indicating high diversity. When the frequencies of a trajectory passing through each cluster are equal, the entropy reaches the maximum possible value for the m distinct spatial clusters. When all points that a trajectory passes through are in the same spatial cluster, the entropy equals zero. Both LCSS and entropy were calculated using the TraMineR package in R (Gabadinho et al. (2011); http://cran.r-project.org/web/packages/TraMineR/index.html, last accessed on May 3, 2015).

We used three approaches to examine the utility of using LCSS and H to quantify movement patterns of the three species. First, we visualized the LCSS distance matrix to distinguish individual movement patterns among the three species and the special movement patterns of some individuals. Second, we used boxplots to explore the ability of entropy (H) to differentiate the three species. Third, we examined the performance of trajectory classification based on LCSS and entropy (H) under the assumption that LCSS and entropy (H) can achieve high classification accuracy if they capture differences in the movement patterns of the three species.

Trajectory classification was conducted using Support Vector Machine (SVM) (Dimitriadou et al. 2008), a typical method used in trajectory analysis (e.g., Lee et al. 2008; Dodge, Weibel, and Forootan 2009). SVM separates objects into different class labels by constructing hyperplanes in a multidimensional space (Vapnik 2000). The two dimensions for SVM that we used were entropy (H) and the first axis from Metric Multidimensional Scaling (MMS) on the LCSS distance matrix. MMS is a dimensionality-reduction technique that places objects into an n dimensional space while attempting to preserve pairwise distances of objects in a distance matrix as much as possible. In MMS, the first axis represents the largest variance of LCSS distance matrix among the other axes (Salkind 2006). *Five-fold cross-validation* (the same portions of training and testing data as Lee et al. (2008)) was used to avoid bias that may be caused by the selection of training samples. The data set was evenly divided into five subsets. In each of the five single runs, one subset was hold for testing and the others were used to train the SVM, which was then applied to classify the trajectories in the testing subset as either cattle, elk or mule deer. The accuracy for that single run is the ratio of the number of correctly predicted trajectories and the number of trajectories in the testing subset. The SVM classification was conducted using the e1071 package in R (Meyer, Leisch, and Hornik (2003); http://cran.r-project.org/web/packages/e1071/index.html, last accessed on May 3, 2015). Classification accuracy was determined from the mean accuracy of five single runs.

Prior to the analyses, we expected that species movements should be constrained by two factors: (1) the presence of internal barbed wire fences, which should confine cattle movement to the northern portion of the range, and (2) the distribution of ecosystems across the study area, which should affect the movement pathways, home range size and location of suitable habitats for all three species, but especially for elk and mule deer. We examined the ability of LCSS and entropy (H) to capture such differences by extracting two factors from the trajectories. We first calculated the percentage of points in each trajectory that were within the internal barbed-wire fences. We assumed that LCSS, which measures spatial similarity between pairs of trajectories, should be able to capture the spatial constraint that these fences place on the movement of cattle. To test the correlation between the LCSS distance matrix and the percentage of points in each trajectory within the barbed-wire fences, we converted the percentage of points in each trajectory to the pairwise distance between trajectories, and then performed a Mantel test, which determines the correlation between two distance matrices (Legendre and

Fortin 1989). Mantel tests were performed using the ecodist package in R (Goslee and Urban (2007); http://cran.r-project.org/web/packages/ecodist/index.html, last accessed on May 3).

Next, we examined the home range which should be affected by species-specific adaptations and biological properties, such as weight, age, and gender (McNab 1963; Harestad and Bunnel 1979). Home range area has fundamental consequences for many ecological processes, such as the distribution and abundance of organisms and population regulation (Börger, Dalziel, and Fryxell 2008). So we tested whether entropy (H) "reflects" ecological meanings that home range carries by fitting entropy (H) and the area of home range with linear regression. We chose Minimum Convex Polygon (MCP) to estimate the area of home range. Although MCP has deficiencies and a variety of home range estimation methods have been developed (e.g., Getz and Wilmers 2004; Downs and Horner 2009), MCP remains a classic method that has been used either for home range estimation or used as a reference that any new home range estimation methods are compare to. In addition, we tested whether LCSS is correlated with the home range overlap of individual organisms using a Mantel test (Legendre and Fortin 1989). The overlapped area of home range for each pair of individuals was standardized by the geometrical mean of individuals' area of home range.

The assignment of telemetry locations into spatial clusters is a key process in our method. As discussed above, this procedure is controlled by two parameters, the number of nearest neighbors (k) and the minimum size of clusters (q). To test the robustness of our method to variations in the values of k and q, we generated spatial clusters using values of k and q that ranged from 10 to 100, with an interval of 10, resulting in 100 combinations of k and q in total. All tests introduced above were conducted for the 100 configurations of k and q.

Results

Spatial clustering

The number of nearest neighbors used to determine the similarity between each pair of points (k) and the minimum number of points contained in a cluster (q) control the number and shape of resulting spatial clusters (i.e., the assignment of points to spatial clusters). Among the 100 combinations of k and q, the minimum number of spatial clusters was 46 ($k = 10$ and $q = 100$) and the maximum number of spatial clusters was 716 ($k = 10$ and $q = 10$).

The size of spatial clusters was controlled primarily by q. Decreasing the value of q yielded more spatial clusters containing fewer points at lower levels of a dendrogram, which were nested in fewer but larger spatial clusters at higher levels (Fig. 3). In contrast, k primarily governed the shape of spatial clusters. When q was held constant, varying k yielded a similar number of spatial clusters in which the number of points was no less than q. However, the shape of these spatial clusters was different because the clustering was based on similarity governed by different values of k (Fig. 4).

Contrasting species trajectories: LCSS and entropy

Although the number and shape of spatial clusters varied under the configurations of k and q, there were significant correlations ($P < 0.05$) for all 100 configurations between: (1) the LCSS distance and the percentage of points in each trajectory within the fences, (2) entropy (H) and the area of home range, and (3) LCSS and the home range overlap. LCSS and entropy (H) extracted from the spatial clusters generated from the configuration of $k = 10$ and $q = 100$, which yielded the smallest number of clusters (Fig. 4a), were chosen to illustrate the difference
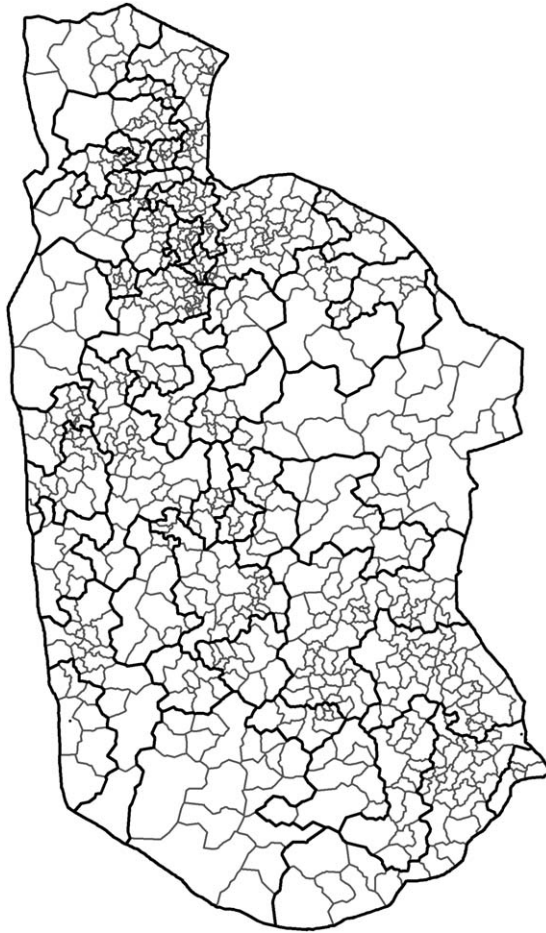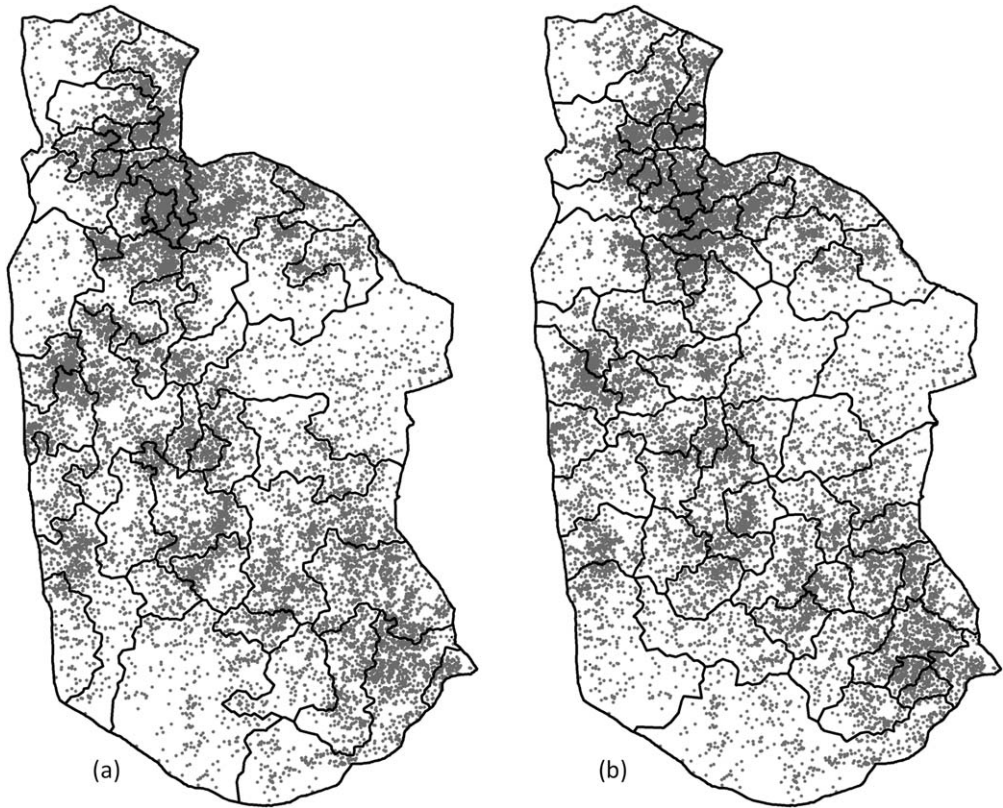
Figure 3. Spatial clusters created under configurations of $k = 10$ and $q = 10$ (thin lines in grey) and $k = 10$ and $q = 100$ (thick lines in black). When k is held constant, a lower value of q yields smaller and more clusters that are nested in larger and fewer clusters generated from a larger value of q.

of LCSS and entropy (H) among the three species. Other configurations showed similar patterns.

Presenting LCSS distance measures in a matrix facilitated the visual examination of trajectory dissimilarities across and within species (Fig. 5a). Dissimilarity was typically high between the trajectories of individual cattle and mule deer, individual mule deer and elk, and individual elk and cattle, which indicated that LCSS identified the differential movement patterns of the three species. Within species, trajectories of individual mule deer exhibited greater dissimilarity than those of individual cattle and elk (Fig. 5a), suggesting that the movement behavior differed more among individual mule deer than among individual cattle or elk. On average, trajectories of elk had the highest entropy, followed by those of cattle and mule deer (Fig. 6).

Beyond these general patterns, LCSS was able to highlight some exceptions. For example, most mule deer utilized different areas (Fig. 1) and therefore had relatively high dissimilarity
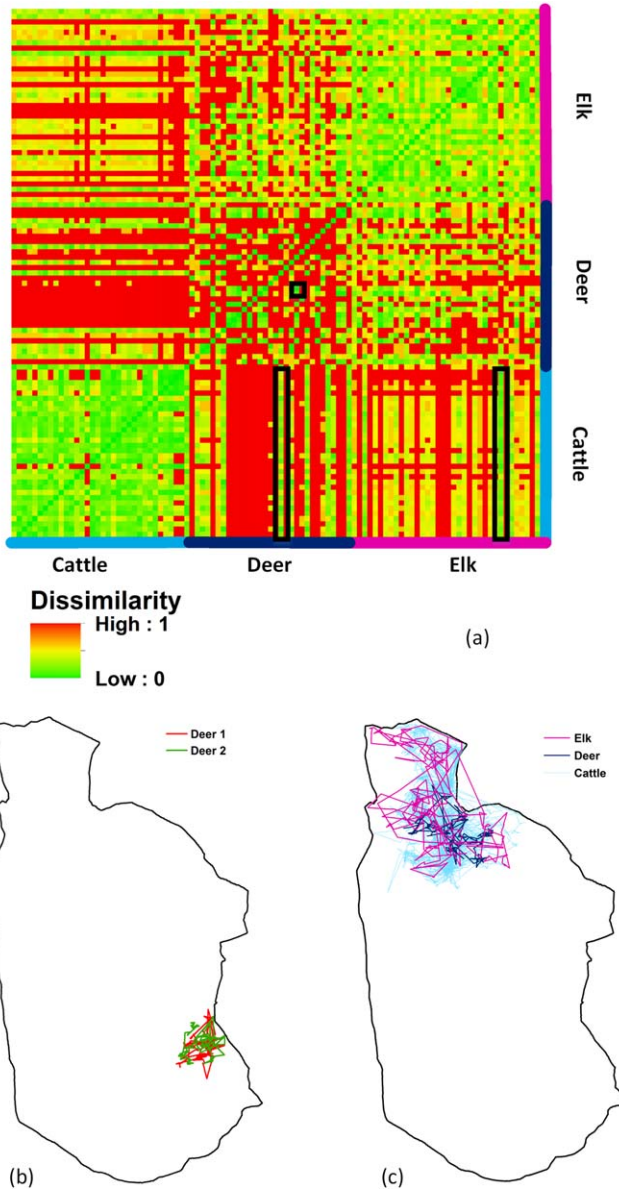
Figure 4. Spatial clusters created under configurations of (a) $k = 10$ and $q = 100$ ($n = 46$) and (b) $k = 100$ and $q = 100$ ($n = 63$). When q is held constant, varying k yields a similar number of spatial clusters but with different shapes. Dots are telemetry locations.

with other mule deer (red pixels representing dissimilarity within trajectories of mule deer in Fig. 5a). The pair of trajectories for the two mule deer highlighted by the square box in Fig. 5a, however, showed more similar movement patterns because the individuals moved through similar areas (Fig. 5b). Similarly, in contrast to the high dissimilarity between most trajectories of mule deer/elk and cattle, trajectories of one mule deer and one elk highlighted by the two rectangular boxes in Fig. 5a had higher similarity with cattle because they utilized similar areas (Fig. 5c).
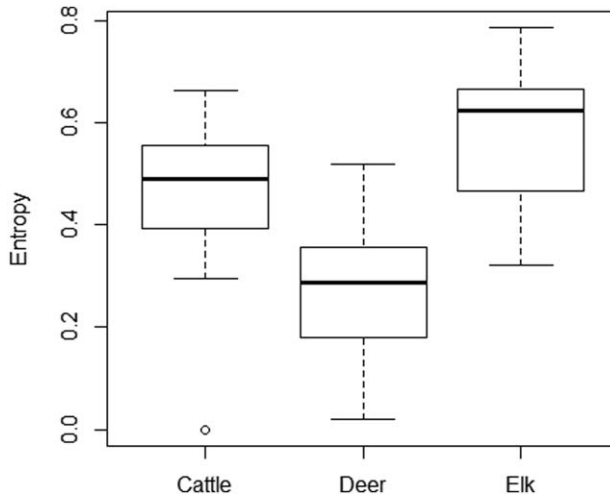
Trajectory classification

Finally, we examined the ability of our approach to differentiate the movement characteristics of different species using support vector machines to classify individual trajectories on the basis of shared movement sequences (LCSS) and the diversity of spatial clusters that were visited (entropy). Because both measures are dependent on the creation of spatial clusters, the results will vary depending on the number of nearest neighbors (k) and the minimum size of clusters (q) used in the clustering procedure. While we examined classification accuracy for all 100 combinations of k and q used to create spatial clusters, the results presented here are for $k = 60$, $q = 70$, one of the configurations that achieved the highest classification accuracy

Figure 5. Results of LCSS analysis. (a) LCSS distance matrix of trajectories of cattle, mule deer, and elk extracted from the spatial clusters generated under the configuration of $k = 10$ and $q = 100$. Each pixel represents the dissimilarity between a pair of trajectories. (b) Trajectories of two individual mule deer that exhibit high similarity as highlighted by the square box in (a). (c) Trajectories of one individual mule deer and one individual elk that had high similarity to trajectories of cattle (the two whitish strips highlighted by the two rectangular boxes in (a).

(92.2%), and $k = 10$, $q = 90$, the configuration achieving the lowest classification accuracy (83.3%).

Figure 6. Boxplot showing the entropy of trajectories for cattle, mule deer, and elk extracted from spatial clusters with the configuration of $k = 10$ and $q = 100$.

Trajectory classifications for both configurations of k and q showed a pattern in which individuals of specific species were clearly segregated on the basis of entropy (H) values and LCSS similarities, as captured by the first axis values from a MMS on pairwise LCSS distances (Fig. 7). For both configurations, deer were distinguished by lower entropy, which indicates that individuals traveled through a lower diversity of spatial clusters than individual cattle or elk. In contrast, cattle and elk both had higher mean entropy values than mule deer, but were differentiated on the basis of MMS Axis 1 values. Specifically, trajectories of cattle were clustered at positive MMS Axis 1 values, which were associated with a high percentage of points within the fences, while most elk trajectories had negative MMS Axis 1 values. Clustering of individual cattle, elk, and mule deer trajectories was more compact from the configuration of $k = 60$, $q = 70$ than that from $k = 10$, $q = 90$, which explained the higher classification accuracy. The trajectories of individual mule deer and elk with movement patterns that were similar to cattle (*discussed above and illustrated in Fig.* 5*c*) were those least distinguishable in Fig. 7.

When H and LCSS similarities were used to classify the individual trajectories, the accuracy of species classification for the 100 combinations of k and q ranged from 83.3% to 92.2%, with an average of 89.4%. These accuracy values were all greater than or equal to the 83.3% accuracy achieved by Lee et al. (2008) for the same data set but using an alternative method. The lowest accuracy resulted from the configuration of $k = 10$ and $q = 90$ (which used few nearest neighbors and a large minimum cluster size) while the highest accuracy was from the configurations of: $k = 50$, $q = 20$; $k = 60$, $q = 40$; $k = 60$, $q = 70$; $k = 60$, $q = 80$; $k = 70$, $q = 50$; $k = 70$, $q = 70$; and $k = 90$, $q = 90$.

Discussion

Our method provides an alternative approach for analyzing trajectories of animal movement by considering trajectories as sequences of spatial clusters, which are an aggregation of telemetry points. Two parameters, the number of nearest neighbors (k) and the minimum size of clusters (q), control the number and the shape of spatial clusters. Consequently, they affect how many
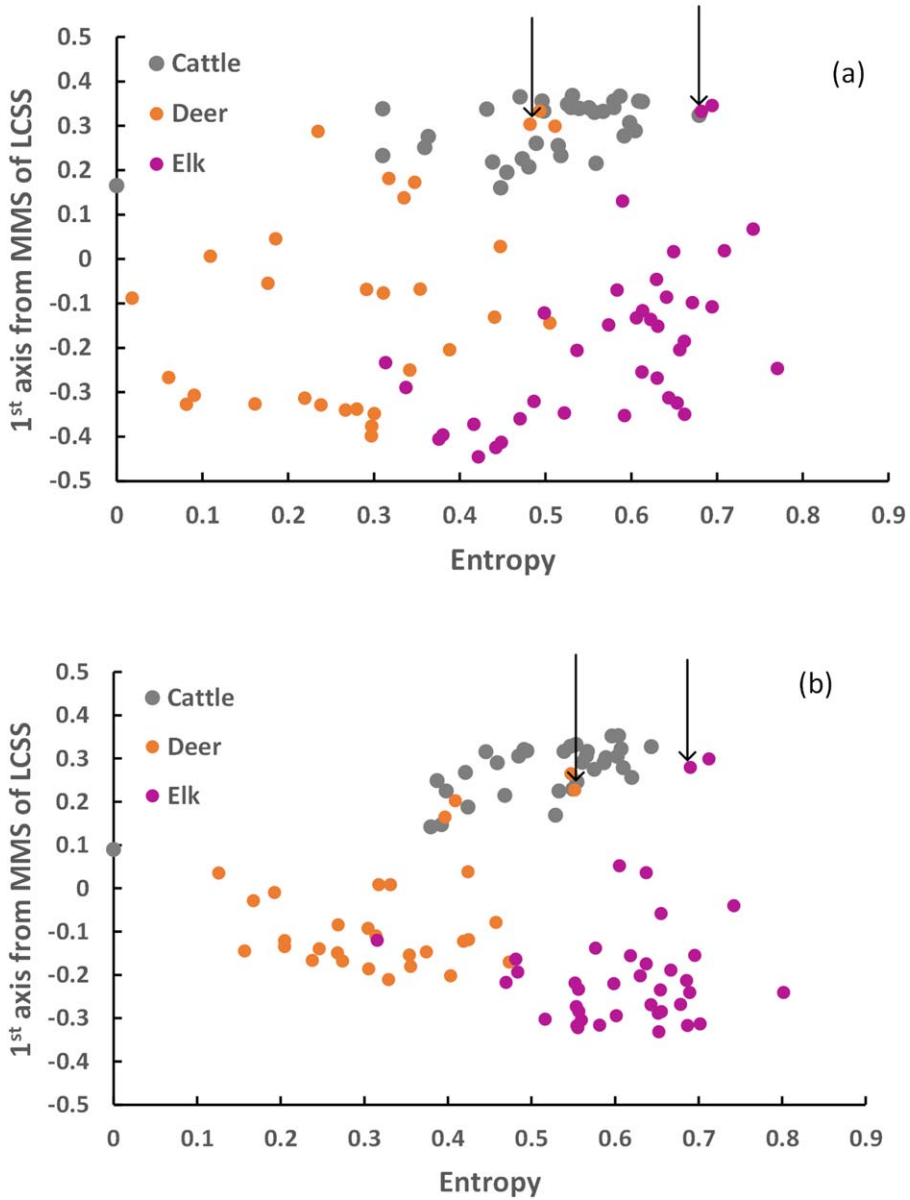
Figure 7. Plots of LCSS and entropy values used to classify individual trajectories of cattle, mule deer, and elk with the configuration of (a) $k = 10$ and $q = 90$ and (b) $k = 60$ and $q = 70$. Arrows denote the trajectories of mule deer and elk discussed in Fig. 5c.

labels (i.e., spatial clusters) are used to represent the trajectories and how these trajectories are labelled (i.e., the assignment of each point in trajectories to its corresponding spatial cluster). By examining LCSS and entropy extracted from spatial clusters generated by k and q in a wide range, we demonstrated that the patterns suggested by LCSS and entropy were robust to varied configurations of k and q.

Informative properties of trajectories can be extracted based on these sequences. We chose two particular measures, LCSS and entropy (refer to Gabadinho et al. (2011)) for other useful properties), and argue that the design of our method is particularly suitable for the analysis of animal movement. Entropy was highly correlated to home range area, which itself is related to biological properties (e.g., weight, age, and gender) and a range of ecological processes. The lower entropy of trajectories for mule deer suggested that individuals preferentially move through certain places (e.g., habitats or food sources) more regularly than cattle and elk (Figs. 1 and 6).

LCSS similarity emphasizes the similarity of spatial clusters shared by individual trajectories while paying less attention to non-shared clusters. For example, the LCSS for trajectories passing through spatial clusters *AACBDBDA* and *AEAFBBGA* is *AABBA*. The similar segments in trajectories might be places for water or food sources where all individuals need to visit, which also suggested by the correlation with home range overlap of individuals. While the dissimilar segments: (1) could be unusual or random movement caused by avoiding predators or unpredictable reasons, or (2) suggest that individuals have different preferences for habitat and food sources.

The high LCSS distance among the trajectories of mule deer suggested segregation of mule deer in the usage of resources (Bowyer 1984). Coe et al. (2001) found that elk usage of ponderosa pine/Douglas fir ecosystems decreased with the presence of cattle while mule deer usage of the same ecosystems increased as the elk's usage decreased in early summer. These findings were reflected in this study by the high LCSS distance among the trajectories of the three species, which suggested that their trajectories had little overlap in sequences moving across time or space (Fig. 5). LCSS also had a strong relationship with the percentage of points in each trajectory within fences, which indicated that LCSS captured the spatial constraint that fences placed on cattle movement.

Because LCSS and entropy captured quantitative aspects of individual movements that could be tied to ecological patterns and constraints (which themselves are species-specific), our method was able to effectively discriminate among species movements and achieved a high accuracy when classifying individuals by species according to their movement patterns. In fact, the classification accuracy based on all 100 combinations of k and q were higher than or equal to Lee et al. (2008), which supported the robustness of our method to these two parameters. Lee et al. (2008) employed a method that had two phases: region-based clustering and trajectory-based clustering. The region-based clustering phase grouped major class of trajectories into homogenous rectangular regions. The trajectory-based clustering phase then partitioned trajectories into the same class based on the density. Both of these phases not only involve the determination of two optimal parameters for each class of trajectories but also the selection of discriminative clusters. Their method achieved an accuracy of 83.3% using the two phases, but just 50% if only the trajectory-based clustering phase was used. Our method used only two parameters (i.e., k and q) to create spatial clusters and two properties (i.e., LCSS and entropy) to discriminate trajectories of different species and therefore, achieved higher classification accuracy in a much simpler way.

Conclusion

In this article, we present a new approach for the analysis of animal trajectories that considers animal movements as sequences passing through aggregations of telemetry locations and

applied the method to investigate movement patterns of cattle, mule deer, and elk at the Starkey EFR. The properties extracted from these sequences effectively distinguished the movement patterns of different species. Trajectory classification using these properties achieved higher accuracy than the previous study with great simplicity. Although we applied our method on a data set that has multiple species, our method is not limited to classify trajectories. Our method has the advantage of extracting ecologically meaningful features in animal movement and could be used to quantify animal movement, explore unknown patterns, detect trends and identify unusual behaviors.

Acknowledgement

Appendix

Algorithm 1: Creating spatial clusters (SC) from telemetry locations

Input: $P = \{p_i \mid 1 \leq I \leq n\}$: all of the collected telemetry locations;
    q: the minimum number of points in a SC (i.e., the minimum size of a SC);
    k: the number of nearest neighbors to determine the similarity between points.
Output: A set of $SC = \{SC_j \mid 1 \leq j << n\}$. Each SC_j contains a group of points (p_i) whose size is no less than q (i.e., $|SC_j| \geq q$).
Steps:

1. Create a Delaunay triangulation (DT) for all p_i in P;
2. Calculate the similarity of points at each DT edge according to equation (1);
3. Assign each p_i to a single SC_i;
4. Order the DT edges in descending order according to the similarity of the points at the two ends;
5. Pick up the next shortest edge $<p_i, p_j>$ in the DT, if the SC_i, SC_j that the two end points belongs to satisfies $SC_i! = SC_j$ and $\mid SC_i \mid < q$, $\mid SC_j \mid < q$ then merge SC_i, SC_j
6. Repeat step 5 until all the edges in the DT are processed then quit.

References

Andrienko, N., and G. Andrienko. (2011). "Spatial Generalization and Aggregation of Massive Movement Data." *IEEE Transactions on Visualization and Computer Graphics* 17, 205–19.

Andrienko, N., and G. Andrienko. (2013). "Visual Analytics of Movement: An Overview of Methods, Tools and Procedures." *Information Visualization* 12, 3–24.

Andrienko, N., G. Andrienko, and P. Gatalsky. (2000). "Towards Exploratory Visualization of Spatio-Temporal Data." In *Third AGILE Conference on Geographical Information Science*, Helsinki/Espoo, Finland, 137–42.

Börger, L., B. D. Dalziel, and J. M. Fryxell. (2008). "Are There General Mechanisms of Animal Home Range Behaviour? A Review and Prospects for Future Research." *Ecology Letters* 11, 637–50.

Bowyer, R. T. (1984). "Sexual Segregation in Southern Mule Deer." *Journal of Mammalogy* 65, 410–17.

Brillinger, D. R., H. K. Preisler, A. A. Ager, and J. G. Kie. (2004). "An Exploratory Data Analysis (EDA) of the Paths of Moving Animals." *Journal of Statistical Planning and Inference* 122, 43–63.

Burt, W. H. (1943). "Territoriality and Home Range Concepts as Applied to Mammals." *Journal of Mammalogy* 24, 346–52.

Calenge, C., S. Dray, and M. Royer-Carenzi. (2009). "The Concept of Animals' Trajectories from a Data Analysis Perspective." *Ecological Informatics* 4, 34–41.

Chen, L., M. T. Özsu, and V. Oria. (2005). "Robust and Fast Similarity Search for Moving Object Trajectories." In *Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data*, 491–502.

Cheriyadat, A. M., and R. J. Radke. (2008). "Detecting Dominant Motions in Dense Crowds." *IEEE Journal of Selected Topics in Signal Processing* 2, 568–81.

Coe, P. K., B. K. Johnson, J. W. Kern, S. L. Findholt, J. G. Kie, and M. J. Wisdom. (2001). "Responses of Elk and Mule Deer to Cattle in Summer." *Journal of Range Management* 54, 205.

Dimitriadou, E., K. Hornik, F. Leisch, D. Meyer, and A. Weingessel. (2008). "Misc Functions of the Department of Statistics (e1071), TU Wien." *R Package* Version 1.6-4. Available from url: http://cran.r-project.org/web/packages/e1071/index.html

Dodge, S., R. Weibel, and E. Forootan. (2009). "Revealing the Physics of Movement: Comparing the Similarity of Movement Characteristics of Different Types of Moving Objects." *Computers, Environment and Urban Systems* 33, 419–34.

Downs, J. A., and M. W. Horner. (2009). "A Characteristic-Hull Based Method for Home Range Estimation." *Transactions in GIS* 13, 527–37.

Fieberg, J., and C. O. Kochanny. (2005). "Quantifying Home-Range Overlap: The Importance of the Utilization Distribution." *Journal of Wildlife Management* 69, 1346–59.

Gabadinho, A., G. Ritschard, N. S. Mueller, and M. Studer. (2011). "Analyzing and Visualizing State Sequences in R With Traminer." *Journal of Statistical Software* 40, 1–37.

Getz, W. M., and C. C. Wilmers. (2004). "A Local Nearest-Neighbor Convex-Hull Construction of Home Ranges and Utilization Distributions." *Ecography* 27, 489–505.

Goslee, S. C., and D. L. Urban. (2007). "The Ecodist Package for Dissimilarity-Based Analysis of Ecological Data." *Journal of Statistical Software* 22, 1–19.

Guo, D., S. Liu, and H. Jin. (2010). "A Graph-Based Approach to Vehicle Trajectory Analysis." *Journal of Location Based Services* 4, 183–99.

Guo, D., D. Peuquet, and M. Gahegan. (2003). "ICEAGE: Interactive Clustering and Exploration of Large and High-dimensional Geodata." *GeoInformatica* 7, 229–53.

Guo, D., X. Zhu, H. Jin, P. Gao, and C. Andris. (2012). "Discovering Spatial Patterns in Origin-Destination Mobility Data." *Transactions in GIS* 16, 411–29.

Harestad, A. S., and F. Bunnel. (1979). "Home range and body weight–a reevaluation." *Ecology* 60, 389–402.

Huttenlocher, D. P., G. A. Klanderman, and W. J. Rucklidge. (1993). "Comparing Images Using the Hausdorff Distance." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15, 850–63.

Jaeger, C. P., and V. A. Cobb. (2012). "Comparative Spatial Ecologies of Female Painted Turtles (*Chrysemys picta*) and Red-Eared Sliders (*Trachemys scripta*) at Reelfoot Lake, Tennessee." *Chelonian Conservation and Biology* 11, 59–67.

Lee, J. G., J. Han, X. Li, and H. Gonzalez. (2008). "Traclass: Trajectory Classification Using Hierarchical Region-Based and Trajectory-Based Clustering." *Proceedings of the VLDB Endowment* 1, 1081–94.

Lee, J. G., J. Han, and K. Y. Whang. (2007). "Trajectory Clustering: A Partition-and-Group Framework." In *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data*, 593–604.

Legendre, P., and M. J. Fortin. (1989). "Spatial Pattern and Ecological Analysis." *Vegetatio* 80, 107–38.

Long, J. A., and T. A. Nelson. (2013). "A Review of Quantitative Methods for Movement Data." *International Journal of Geographical Information Science* 27, 292–318.

Long, J. A., T. A. Nelson, S. L. Webb, and K. L. Gee. (2014). "A Critical Examination of Indices of Dynamic Interaction for Wildlife Telemetry Studies." *Journal of Animal Ecology* 83, 1216–33.

McNab, B. K. (1963). "Bioenergetics and the determination of home range size." *American naturalist* 97, 133–140.

Meyer, D., F. Leisch, and K. Hornik. (2003). "The Support Vector Machine under Test." *Neurocomputing* 55, 169–86.

Miller, J. A. (2012). "Using Spatially Explicit Simulated Data to Analyze Animal Interactions: A Case Study with Brown Hyenas in Northern Botswana." *Transactions in GIS* 16, 271–91.

Millspaugh, J. J., R. A. Gitzen, B. J. Kernohan, J. A. Larson, and C. L. Clay. (2004). "Comparability of Three Analytical Techniques to Assess Joint Space Use." *Wildlife Society Bulletin* 32, 148–57.

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 of the United States of America* 105, 19052–9.

Rowland, M. M., P. K. Coe, R. J. Stussy, A. A. Ager, N. J. Cimon, B. K. Johnson, and M. J. Wisdom. (1998). "The Starkey Habitat Database for Ungulate Research: Construction, Documentation, and Use." *US Forest Service General Technical Report PNW-GTR-430*.

Ryan, T. J., C. A. Conner, B. A. Douthitt, S. C. Sterrett, and C. M. Salsbury. (2008). "Movement and Habitat Use of Two Aquatic Turtles (*Graptemys geographica* and *Trachemys scripta*) in an Urban Landscape." *Urban Ecosystems* 11, 213–25.

Salkind, N. J. (2006). *Encyclopedia of Measurement and Statistics*. New York: Sage Publications.

Stewart, K. M., R. T. Bowyer, J. G. Kie, N. J. Cimon, and B. K. Johnson. (2002). "Temporospatial Distributions of Elk, Mule Deer, and Cattle: Resource Partitioning and Competitive Displacement." *Journal of Mammalogy* 83, 229–44.

Studer, M., G. Ritschard, A. Gabadinho, and N. S. Müller. (2011). "Discrepancy Analysis of State Sequences." *Sociological Methods & Research* 40, 471–510.

Tsai, V. J. D. (1993). "Delaunay Triangulations in TIN Creation—An Overview and a Linear-time Algorithm." *International Journal of Geographical Information Science* 7, 501–24.

Usabiaga, J., G. Bebis, A. Erol, and M. Nicolescu. (2007). "Recognizing Simple Human Actions Using 3D Head Movement." *Computational Intelligence* 23, 484–96.

Vapnik, V. (2000). *The Nature of Statistical Learning Theory*. New York: Springer.

Verhein, F., and S. Chawla. (2008). "Mining Spatio-Temporal Patterns in Object Mobility Databases." *Data Mining and Knowledge Discovery* 16, 5–38.

Vlachos, M., D. Gunopoulos, and G. Kollios. (2002). "Discovering Similar Multidimensional Trajectories." In *Proceedings of the 18th International Conference on Data Engineering*, 673–84.

Zhang, Z., K. Huang, and T. Tan. (2006). "Comparison of Similarity Measures for Trajectory Clustering in Outdoor Surveillance Scenes." *In ICPR '06: 18th International Conference on Pattern Recognition (ICPR 2006)*, 1135–8. Washington, DC: IEEE.