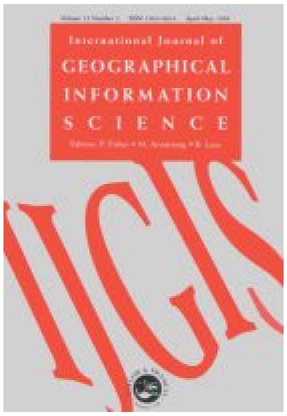


Individual Movement - Sequence Analysis Method (IM-SAM): Characterizing Spatio-Temporal Patterns of Animal Habitat Use across Landscapes

Journal:	*International Journal of Geographical Information Science*
Manuscript ID	IJGIS-2018-0277.R4
Manuscript Type:	Special Issue Paper
Keywords:	sequence dissimilarity, dendrogram, ungulates, spatio-temporal habitat use, mechanistic movement model

Individual Movement - Sequence Analysis Method (IM-SAM): Characterizing Spatio-Temporal Patterns of Animal Habitat Use across Landscapes

We present methodological advances to a recently developed framework to study sequential habitat use by animals using a visually-explicit and tree-based Sequence Analysis Method (SAM), derived from molecular biology and more recently used in time geography. Habitat use sequences are expressed as annotations obtained by intersecting GPS movement trajectories with environmental layers. Here, we develop IM-SAM, where we use the individual reference area of use as the reference spatial context. To assess IM-SAM's applicability, we investigated the sequential use of open and closed habitats across multiple European roe deer populations ranging in landscapes with contrasting structure. Starting from simulated sequences based on a mechanistic movement model, we found that different sequential patterns of habitat use were distinguished as separate, robust clusters, with less variable cluster size when habitats were present in equal proportions within the individual reference area of use. Application on real roe deer sequences showed that our approach effectively captured variation in spatio-temporal patterns of sequential habitat use, and provided evidence for important behavioral processes, such as day-night habitat alternation. By characterizing sequential habitat use patterns of animals, we may better evaluate the temporal trade-offs in animal habitat use and how they are affected by changes in landscapes.

Keywords: sequence dissimilarity; dendrogram; ungulates; spatio-temporal habitat use; mechanistic movement model.

Introduction

Understanding which habitat features are used by animals through space and time is important to establish cost-effective and flexible policies that are essential for species conservation and wildlife management purposes. For example, several ungulate species show higher activity and intensified movement at dusk and dawn, resulting in more road-crossings (Kämmerle *et al.* 2017) and, hence, vehicle collisions during twilight.

Similarly, by alternating between access to food and cover resources over the day, several wild species have adapted to agro-ecosystems (Aulak and Babinska-Werka 1990; Cibien *et al.* 1989; Hewison *et al.* 2001, Podgórski *et al.* 2013) or even urbanized areas (i.e., the phenomenon of 'urban wildlife'; Magle *et al.* 2012). Most animal movement methods have predominantly focused on how to analyze the spatial component so far, while the temporal dependence of habitat use is often analyzed less elegantly by pooling samples into classes, such as night/day and active/inactive, typically ignoring the sequential nature of habitat selection. Consequently, we need robust methodological approaches to understand the sequential temporal patterns in the use of complementary habitats in order to take appropriate conservation and management actions.

Habitats provide the resources (e.g. food, cover, thermal protection) that species need for survival and reproduction (Manly *et al.* 2002). European-level mapping products (e.g. Corine Landcover, Copernicus; see eea.europa.eu) are often used by movement ecologists to quantify such resources (e.g. forest cover) by linking these maps with GPS locations obtained from animal tracking projects. Improved spatio-temporal resolution and range of both remote sensing products and animal tracking datasets are allowing ecologists to derive ever more detailed animal trajectories annotated with habitat information, and, hence, facilitate the study of the animal-habitat relationship over time (Cagnacci *et al.* 2010, Demšar *et al.* 2015, Kays *et al.* 2015).

In geo-informatics, such habitat information is referred to as *biological and environmental context* and the integrated analysis of movement trajectories in relation to such contexts are termed *context-aware movement analysis* (Andrienko *et al.* 2011; Dodge *et al.* 2013; Demšar *et al.* 2015). Geo-informaticians and ecologists have together developed several context-aware methods to visualize and analyze movement in relation to habitat type (see Demšar *et al.* 2015 for overview). The concept was first proposed and applied to animal movement data by Andrienko *et al.* (2011). In a case study using roe deer GPS movement data, the latter presented several methods to visualize aggregated hourly use of open habitats for spatial clusters of locations. Xavier and Dodge (2014) developed DYNAMO (Dynamic Multivariate Visualization of Movement), a tool for animating trajectories annotated by habitat variables. Demšar *et al.* (2015b) proposed a 3D visualization of a home range, where the x,y-plane is space, and the z-axis is time, and aggregated information of used habitats could characterize the space-time cube. Toor *et al.* (2016) developed a trajectory segmentation algorithm based on temporal changes in habitat use using random forest models. All these context-aware approaches, one way or another, investigate how the contextual information is used through time.

One way of considering time dependency is to investigate sequentiality, which takes into account the temporal order in which behavioral, environmental or movement states occur. In the field of movement ecology, and especially for the study of recursions (i.e., revisitations of the same places), several promising methods have been proposed, often relying on approaches developed in different research fields or by combining several concepts (Berger-Tal and Bar-David 2015). Fourier and wavelet transforms have been used to simultaneously detect recursions at multiple temporal scales (Wittemyer *et al.* 2008, Bar-David *et al.* 2009, Polansky *et al.* 2010, Riotte-Lambert *et al.* 2013, 2017). Minimal conditional entropy was used to identify the temporal scale of repetitiveness in resource patch visitation and to quantify the degree of predictability in movement sequences (i.e, traplining, Riotte-Lambert *et al.* 2017). The latter study also presented an algorithm based on time-lag matrices (Goto 2003) to mine for repeated movement sub-sequences. Utilization distribution in combination with residence time was also proposed to identify areas of high recursion rate from movement data (Benhamou and Riotte-Lambert 2012). Recently, model-based continuous-time movement metrics were suggested to find recursion signatures (Péron *et al.* 2017), as well as periodograms (Péron *et al.* 2016).

The Sequence Analysis Methods (SAM) approach was recently developed to measure similarity between temporally ordered sequences of habitat use within individuals or populations (De Groeve *et al.* 2016). SAM is a tree-based approach developed in computer science to measure dissimilarity between multiple strings of characters (Wagner and Fischer 1974) and has subsequently been used in different fields of study. It was first adopted to measure dissimilarity between DNA and protein sequences (Needleman and Wunsch 1970). The popularity of the technique in molecular biology resulted in several applications in other fields such as in sociology to study life courses (e.g. Abbott 1995, Wilson 2006, Gabadinho *et al.* 2011), in time geography for transportation science (Wilson 2008), in tourism research (Shoval and Isaacson 2007), in indoor navigation (Delafontaine *et al.* 2012), in choreography research (Chavoshi *et al.* 2015), in human mobility (Brum-Bastos *et al.* 2018); and, recently, in the field of animal movement ecology (De Groeve *et al.* 2016).

De Groeve *et al.* (2016) showed that, for a given proportion of habitat used, animals can show very different sequential space use patterns. For example, while animals may equally use open and closed habitats over a given time-window, their sequential use patterns were markedly different (from random to day-night alternating patterns). We describe here a methodological framework building on De Groeve *et al.* (2016), where we analyze sequential patterns of habitat use of animals, while accounting for individual-level variability in landscape heterogeneity, or IM-SAM (Individual Movement - Sequence Analysis Method). In essence, the baseline SAM-framework uses simulated sequential habitat use patterns to determine the classification of real habitat use sequences, where simulated sequential habitat use patterns were generated based on a biased and correlated random walk movement model. De Groeve *et al.* (2016) generated the patterns of sequential habitat use in *artificially generated landscapes* that *mimic* habitat composition and structure of an animal's movement context. Instead, in IM-SAM we generated these sequential patterns in the *real landscape context for an individual animal*, or individual reference area of use, and hence accounted for the *true* variation in habitat composition and structure among individuals.

In this analysis we use IM-SAM to expand sequential habitat use analysis from animal trajectories derived from a single local context (i.e. single population) to a continental scale (i.e. multiple populations at the European level), specifically using GPS trajectories of 404 individual European roe deer (*Capreolus capreolus*) from nine populations with contrasting landscape structure across Europe. As roe deer are generally described as a forest species, but often feed on rich resources available in more open habitats (e.g. meadow, crop), we described sequential use of two simple habitat classes, open and closed, converting regularized animal trajectories into multiple character sequences, where each character in the sequence corresponds to the habitat used at a given timestamp.

Material and methods – habitat use sequential analysis

To describe sequential use of open and closed habitats for individual animals, we followed a workflow modified from De Groeve *et al.* (2016) that can be summarized in four steps (Fig. 1). First, we produced an exploratory tree for *each individual roe deer* based on biweekly trajectories annotated with habitat categories and used to formulate hypotheses of expected patterns of sequential habitat use (Fig. 1a). The number of sequences *per individual* exploratory tree depends on the monitoring period of the individual. Next, we generated stochastic movement rules for such expected patterns of sequential habitat use and ran the movement models within each individual's reference area of use (here computed as the 100% MCP) in order to produce individual specific simulated trajectories (Fig. 1b), and individual-level simulation trees (Fig. 1c). Simulation trees were therefore based on the true landscape context where each individual actually ranged. Finally, we combined real and simulated trajectories to produce trees where real and simulated habitat use sequences with a high degree of similarity were grouped together. After computing the proportions of simulated patterns in each cluster, we could define sequential habitat use cluster types and assign these identified tags to the real sequences included in that cluster (Fig. 1c/d). Finally, we pruned the output tree and visualized only the classified real trajectories, to facilitate interpretation (Fig. 1d). We now describe each step in detail. [Figure 1 near here]

Real trajectories – exploratory trees (Fig. 1a)

We extracted roe deer trajectories from the EuroDEER database (Cagnacci *et al.* 2011, euroungulates.org) and subsampled them into 16-day GPS trajectories with a fixed four-hour relocation interval (0, 4, 8, 12, 16 and 20h) over a fixed yearly schedule starting on January 1st (e.g. 01/01-16/01, 17/01-01/02, etc.). To increase the sample size, we also included relocations within one hour from the above four-hour intervals, after a sensitivity analysis verified that this did not affect exploratory tree clustering (see Appendix S2). If multiple GPS locations were within this time window, we selected the closest one in time to the reference time stamp. Otherwise, if no fix was obtained for a reference time stamp, we annotated the gap with a missing value (NA). The number of sequences per individual ranged between 2 and 52 and depended on the monitoring time and completeness of the individual trajectory. We annotated each 16-day trajectory with the percentage tree cover (0-100%) extracted from the High-Resolution Layer-Tree Cover Density 2012 (TCD, EEA 2012, 20m spatial resolution), thus obtaining the *biweekly sequences* of habitat use. We reclassified TCD into two distinct classes, using a cut-off point for pixel-level tree cover density of 50%, distinguishing closed (C, ≥50%) and open (O, <50%) habitats. The final dataset consisted of 5402 biweekly habitat use sequences of 404 animals (236 females and 168 males) from nine European roe deer populations characterized by varying forest composition (see map in Appendix S1: Southcentral Norway, NK5; Southwest France, FR8; Switzerland, CH25; Southern Germany, DE15, DE31; Southeast Germany, DE1; Northern Italy, IT1, IT24; Eastern Austria, AU17). After processing, the dataset consisted of 14,607 missing values (2.82%) and 503,985 true GPS locations (97.18%), of which 273,230 (52.69%) were classified as open and 230,755 (44.50%) as closed habitat. See Appendix S2 for the complete data preparation procedure.

We generated an initial visualization of the habitat use patterns by creating exploratory trees separately for each individual (Fig. 1a). These trees were used to describe sequential patterns and helped to build hypotheses for expected models of sequential habitat use (see below). Sequence Analysis Methods use a dissimilarity algorithm to compute the distance between all possible pairs of sequences. All these pairwise distances are written into a dissimilarity matrix. Here, we computed the pairwise distances using the Hamming dissimilarity algorithm (HD), which calculates the minimum number of character substitutions (i.e., O and C) required to match a number of sequences of equal length (Gabadinho *et al.* 2011). From the HD dissimilarity matrix, we subsequently calculated dissimilarity trees using a hierarchical clustering algorithm (Ward's method, Gabadinho *et al.* 2011, De Groeve *et al.* 2016).

The above described analytical steps were performed using the R-package TraMineR (Gabadinho *et al.* 2011).

From SAM to IM-SAM

a. Individual-specific simulated trajectories (Fig. 1b)

We computed individual reference areas of use as 100% Minimum Convex Polygons (MCP) for each roe deer separately as a simple representation of the available space in which movement of that individual could occur, including occasional excursions outside the usual range. We then intersected each MCP with the TCD raster reclassified as open/closed habitat as described above, and characterized all MCPs by their habitat proportions. For illustrative purposes (i.e. Fig. 2), we also classified MCPs into 5 classes (0.1 to 0.5) of 'relative proportion of open/closed habitat'. So, for example, the 0.1 class indicates both 10% open - 90% closed, and 10% closed - 90% open habitat.

Within each MCP we simulated sequential habitat use patterns using a simple spatially explicit stochastic movement model to express expected sequential habitat use patterns that were determined by a memory-based movement model with parametrized selection coefficients for open and closed habitats (see Appendix S3 for the full description). In particular, according to the exploratory trees obtained from real sequences with six locations per day (Fig. 1a), we identified six characteristic patterns of sequential habitat use (Fig. 1b): homogeneous use of closed habitat, or pattern 'C', the resulting day-night sequence (DS) being: DS=CCCCCC; homogeneous use of open habitats, or pattern 'O' (DS=OOOOOO); and three patterns of day-night alternating use of both open and closed habitats, or patterns 'A'. The alternating patterns were generated on the assumption that roe deer use open and closed habitats in relation to the daylight cycle, with use of open habitat more prevalent at night (Mysterud *et al.* 1999; Bonnot *et* *al.* 2013). In addition, we accounted for variation in day length over the different sites and seasons included in the study. In particular, to account for spatial and temporal variation in day length, we distinguished the following three patterns: dominant use of open habitat from 16:00 to 08:00 (pattern 'a24', DS=OOCCOO; winter condition in most sites), equal use of both habitats - open from 20:00 to 08:00, closed from 08:00 to 20:00 (pattern 'a33', DS=OOCCCO; fall and spring in most sites), and dominant use of closed habitat from 04:00 to 20:00 (pattern 'a42', DS=OCCCCO; summer in most sites). In the pattern names, the character ‘a’ refers to daylight-night habitat alternation, where the first number refers to the number of four-hour time periods spent in closed habitat during daytime and the second to the number of four-hour time periods spent in open habitat during the night. Finally, we defined a pattern of random use of the landscape (hence reflecting its structure), or pattern 'U', as a control (i.e. constant selection coefficient for open and closed, see Appendix S3). The seeds of stochastic simulations were random release locations within each individual MCP, whereas successive steps were based on six sets of habitat selection rules matching the aforementioned sequential behaviors. For completeness, we ran the movement simulations with three selection coefficients for each selection pattern to account for behavioral variability (selection coefficient ratios: low, 1:0.2; intermediate, 1:0.1; high, 1:0.01). Each movement simulation was repeated 50 times for each parameter set (i.e., six habitat selection rules by three selection coefficients), varying release location between repetitions but holding release location constant across parameter sets for any given repetition. We thus obtained 900 simulated sequences per individual MCP (six habitat selection rules by three selection coefficients by fifty repetitions).

b. Individual-specific simulation trees (Fig. 1c)

We obtained 404 individual dissimilarities trees (see above) based on individual-specific simulated sequences, thus illustrating the dissimilarity among expected habitat use sequences for each individual separately. Note that HD can be customized by assigning weights to character substitutions when comparing sequences. For example, in De Groeve *et al.* (2016) substitution weights were derived from spatial autocorrelation of habitat classes within simulated landscapes. However, in IM-SAM we considered substitution weights to be redundant because simulated sequences were modeled within each individual MCP and habitat was classified as a binary category. Hence, spatial structure, and thus spatial autocorrelation, was directly accounted for by using this individual-specific modeling approach.

c. Robust classification of individual-specific simulation trees: defining the cut-off distance

In the obtained trees, the leaves are the sequences, and remaining nodes represent clusters (groups of sequences) for which the branch lengths measure the extent of dissimilarity. In other words, the longer the branch length, the higher the dissimilarity between clusters. Hence, the number of clusters that are identified in a tree depends on a cut-off value that should be selected to obtain the most robust dissimilarity tree (Hennig *et al.* 2007). To assess robustness, an iterative procedure of sequence reclustering such as bootstrapping is generally used. Bootstrapping metrics express the consistency in reclassification of sequences in the same clusters. The same procedure can be repeated for different cut-off values (and therefore number of clusters) using the optimization of bootstrapping metrics as a criterion to decide upon the best cut-off point. In IM-SAM, we propose a standardized procedure to identify the most robust and informative tree.

For the trees based on real trajectories only (Fig. 1a) we did not include any cut-off, since we used them for exploratory purposes only. For the clustering of simulated trajectories only, and both simulated and real trajectories in the final classification (Fig.

1c), instead, we allowed all cut-off values (that correspond to distances from the last common node) that generated from a minimum of 2 to a maximum of 20 clusters per tree. For each cut-off value, we computed the Jaccard bootstrapping index for each cluster using bootstrap resampling of *n* number of sequences (BJ, or bootmean; see Hennig 2007, R-package fpc, clusterboot) using 1000 iterations (De Groeve *et al.* 2016), and then we computed the median BJ of all clusters (BJ_{MD} € [0, 1]), and the BJ interquartile range (BJ_{IQ} € [0, 1]). We finally defined a combined bootstrapping index (BJ_{IQMD} € [0, 1]) that we computed for each cut-off value:

$$BJ_{IQMD} = \frac{BJ_{MD} + (1 - BJ_{IQ})}{2}$$

To evaluate the optimized cut-off value, we used a semi-automated selection procedure based on the optimization of the BJ_{IQMD} index. Specifically, we plotted BJ_{IQMD} in relation to the number of clusters for each individual (Fig. 2, top panel and Appendix S4, Fig. S4.3). In most cases, the plot showed two peaks: a primary peak, corresponding to a cut-off value that generated trees with two to three clusters, that separates sequences based on general dissimilarity (for example: homogeneous sequences from all others); and a secondary peak, corresponding to a cut-off value that generated trees with five to eight clusters, catching the complexity of the sequences, i.e. the diversity of sequential habitat use patterns generated by the simulations (see also De Groeve *et al.* 2016). Hence, we excluded the primary peak and defined the cut-off value based on maximum values of BJ_{IQMD} within the secondary peak (Fig. 2, upper panel, light blue region). Then, we did a visual check of all trees derived from the cut-off values selected as above to identify potential inconsistencies, for example if some obvious clusters were overlooked by the cut-off criterion, or if the pruned tree of real trajectories (see below) showed a consistent structure.

d. Identification of cluster types ('cluster tagging')

Once we obtained the 'optimal' tree, we classified each cluster on the basis of the sequential patterns that composed the cluster ('cluster tagging'). First, we calculated the proportion of each sequential habitat use pattern in each cluster. Then, to identify cluster types, we appended each pattern representing at least 10% of the cluster to a tag, ordered by abundance (first criterion) and giving priority to homogeneous patterns (second criterion). For example, if a cluster consisted of 40% a24 sequences, 30% a33 sequences, and 20% a42 sequences, the resulting tag was a24_a33_a42 (first criterion); or, a composition of 30% of O sequences, 30% of a24, 20% of a33, and 15% of a42, was tagged as o_a24_a33_a42 (second criterion). We also used a simplified classification by retaining the most abundant sequential habitat use pattern only as the cluster tag (e.g. a24 and O in the two examples above, respectively).

Simulated and real trajectories - classification trees (Fig. 1c/1d)

In a final step, we obtained the classification tree for each individual by rerunning the HD algorithm on both real and simulated sequences, and by using the cut-off distance as defined above. This way, simulated sequences were used as a *guide for classification of real sequences* to their most similar cluster type (real habitat use sequence tagging, Fig. 1c). For visualization purposes, we pruned the classification tree by only retaining real sequences (Fig. 1d).

Results

Classification of individual-specific simulation trees

The application of the HD algorithm to simulated biweekly sequences generated 404 trees, one per individual MCP. According to the simulation procedure (see above, From

SAM to IM-SAM, par. a; Fig 1b), we expected trees to be composed of six clusters with 150 sequences each. Instead, we found a substantial deviation from this expectation (see Appendix S5 for a sample of 35 trees) that we attributed to individual differences in the environmental context, and notably in the relative proportion of open vs. closed habitat within the individual's MCP.

Using the IM-SAM cut-off criterion (i.e., the second peak in the plot of BJ_{IQMD} for all cut-off values; Fig. 2, top panel), we automatically identified 394 simulation trees, mainly composed of five (164 trees, 40.59%), six (192 trees, 47.52%) or seven (38 trees, 9.41%) clusters, whereas trees with three or eight clusters were rare (7 and 3 respectively, or 1.73% and 0.74%). The overall average BJ_{IQMD} of these trees was 0.749 ± 0.111 which corresponded to a high BJ_{MD} (0.850 ± 0.066), indicating high inter-cluster dissimilarity, and a low BJ_{IQ} (0.123 ± 0.078), indicating low variability in inter-cluster dissimilarity, thus confirming the robustness of the cluster classification. For 32 trees, the number of clusters was manually adjusted to distinguish clear and obvious clusters (from five or six to six, seven, or eight clusters), but this resulted in negligible change in average BJ_{IQMD} (decrease of 0.007, 0.742 ± 0.115). Furthermore, we noted that the relative proportion of open vs. closed habitat in the MCP affected cluster quality (i.e. clustering robustness). Specifically, BJ_{IQMD} increased when the proportion of closed and open habitats was more or less equal (Fig. 2, top panel: higher to lower values from dark to light trend lines). The corresponding BJ_{MD} values also increased when the proportions of open and closed habitats were similar (Fig. 2, top panel: purple shade on the trend lines), whereas the corresponding BJ_{IQ} values decreased (Fig. 2, top panel: orange to blue points along the trend lines). Despite these differences between classes of habitat proportion, the trend in BJ_{IQMD} was consistent, with a second peak for values of between five and seven clusters, with a rapid drop after that. [Figure 2 near here]

The proportion of open vs. closed habitat in the MCP also affected the cluster size within simulation trees. When the proportion of open and closed habitat in the MCP was similar, the trees for simulated sequences were more evenly classified between different clusters. However, when a given habitat type was preponderant in the MCP, some clusters were composed of a larger number of sequences than others (Fig. 2, lower panel, e.g. median lower than 150, and more outliers for class 0.1).

Identification of cluster types ('cluster tagging')

We identified 16 main cluster types among the 404 simulated trees, each representing more than 0.5% of all clusters (Fig. 3, circles in the lower panel). The 2300 clusters consisted of one or more sequential habitat use patterns in differing proportions. Specifically, 1613 clusters consisted of one sequential habitat use pattern (70% of total number of clusters, Fig.3, top-left panel, and legend of cluster types in the lower part: a24, a33, a42, c, o, u, present in 378, 272, 286, 290, 170, and 125 clusters, respectively). 513 (22%) and 139 (6%) clusters consisted of two or three sequential habitat use patterns, respectively (o_u, a33_a42, c_u, a33_a24, a24_a33, a42_a33, u_a33, o_u_a24, c_u_a42, present in 170, 125, 89, 52, 29, 21, 13, 66, and 27 clusters, respectively). The remaining 33 cluster types represented less than 3% of all clusters in total and were also used to classify real trajectories (these rare cluster types were omitted from the legend in Fig. 3; see Appendix S6 for the full set of classified cluster types). [Figure 3 near here]

The number of sequential habitat use patterns occurring in clusters was affected by the relative proportion of open vs. closed habitat in the MCP (Fig. 3, top-left panel). That is, we observed clear-cut cluster identification (i.e., one pattern per cluster) for trees derived from MCPs with similar relative proportions of open and closed habitat. More 'unclear' cluster identification (i.e., with a mix of 3 to 4 sequential habitat use patterns) was observed for trees derived from MCPs with a preponderance of one habitat type only.

Similarly, the occurrence and relative importance of cluster types within trees also depended on the relative proportion of open vs. closed habitat in the MCP (Fig. 3, top-right panel). Indeed, sequences with random and alternating habitat use patterns grouped together with homogeneous sequences when one habitat was prevalent in the MCP (Fig. 3, top-right panel; the patterns are mirrored for high proportion of open or closed habitat). Importantly, only a small proportion of alternating patterns clustered together with a random pattern of use, indicating that alternating patterns rarely occurred at random.

To sum up, we showed that sequences of habitat use patterns generated using predefined habitat use processes within individual MCPs mostly clustered amongst themselves, as expected, but there was some variation in the pattern. Clusters were characterized by one or more sequential habitat use pattern, and this 'cross-pattern' clustering was dependent on the relative composition of open and closed habitats in the MCP.

Classification of real animal trajectories into cluster types

After classification of simulated sequences, we reran SAM also including real trajectories so that these were grouped with the most similar cluster types. After pruning (i.e., filtering out of the simulated sequences), we obtained the classification tree of the real sequences for each individual (See Fig. 1d for an example, and Appendix S7 for a sample of 35 trees). In total, 69.40% of the real habitat use sequences were classified into only six cluster types (i.e., o_u, o_u_a24, c_u, a33, a42, a24, Table 1, in bold), and 97.55% of all real sequences into 17 cluster types (Table 1). The remaining 2.45% of all real habitat use sequences matched another 21 cluster types. All six a-priori simulated sequential habitat use patterns were represented amongst the real data sequences. Specifically, 40.11% of the sequences were classified as one of the six cluster types including a single sequential habitat use pattern (for example, 4.83% of the real sequences were classified as homogeneous closed, c, and 11.85% were classified as alternation a33, Table 1). Another 38.20% were classified into cluster types that included a combination of two patterns, especially a combination of homogeneous open/closed with random patterns (i.e., 29.40%, o_u, c_u, Table 1), or a combination of two alternating patterns (i.e., 8.26%, a33_a42, a33_a24, a42_a33, a24_a33). Finally, the remaining 21.69% of the real sequences were classified into cluster types that included a combination of three or more sequential habitat use patterns. When these results were considered with the most simplified classification (i.e. retaining the most abundant pattern only), the majority of the sequences were classified as homogeneous open (o, 31.51%), and homogeneous closed (c, 24.38%). More than 40% of the sequences were classified as one of the three types of habitat alternation (a33, 18.68%; a42, 10.90%; a24, 10.07%), while the smallest proportion of habitat patterns corresponded to random sequential use of habitat (u, 4.46%). [Table 1 near here]

Discussion

In this paper, we propose an ecological application of Sequence Analysis Methods, IM-SAM to describe sequential habitat use of animal trajectories applied to European roe deer across contrasting landscapes. Below, we first discuss the methodological advances of IM-SAM. Second, we consider the ecological relevance of the observed spatio-temporal patterns of roe deer sequential habitat use across Europe. Finally, we discuss the broader applicability of IM-SAM for other ecological and geographical data.

IM-SAM procedure

IM-SAM provides a suitable method to detect similarity in sequential patterns in movement data of animal species. The IM-SAM framework involves three methodological steps. First, exploration trees are built using real sequences only (Fig. 1a). Then simulation trees are generated taking into account the individual spatial context using simulated sequences only (Fig. 1b/c). Finally, classification trees are produced based on real and simulated sequences combined (Fig. 1c/d). While this three-step conceptual framework is identical to De Groeve *et al.* (2016), scaling up to a multi-population approach involved several fundamental methodological adjustments which we summarize in Table 2, and that improved the procedure. One of the most important advances of IM-SAM compared to other previous ecological applications of the SAM framework (De Groeve *et al.* 2016) is to account for the *individual* spatial context in which an animal moves, by generating individual-specific, spatially-explicit simulated sequences. In this way, individual sequential patterns of habitat use can be identified in a comparable manner across a diversity of landscapes, as done here, facilitating multi-population comparisons. Moreover, in this application, we generated simulations based on expected day-night habitat use patterns. While many natural processes follow alternating and repetitive rhythms (e.g. Wittemyer *et al.* 2008, Bar-David *et al.* 2009, Benhamou and Riotte-Lambert 2012, Polansky *et al.* 2010, Péron *et al.* 2016, 2017, Riotte-Lambert *et al.* 2013, 2017), this might not always be a pattern of interest for other studies. In general, the simulation rules must be based on the question addressed, on the behavioral traits of the species, and the spatio-temporal resolution of the study. For example, when studying migration timing and the use of summer vs winter ranges, simulations might be better based on a weekly timeframe. [Table 2 near here]

The IM-SAM procedure only detects sequential patterns that are coded within the simulation rules. While this appears to be a constraint at first sight, the approach based on simulated movement rules and exploratory trees enables classifying real trajectories within a hypothetical-deductive framework, i.e. based on reproducible expectations. In this sense, exploratory trees represent the empirical observations on which to build the set of hypotheses.

The underlying behavioral choice mechanism of our movement model used to simulate sequential patterns within the individual reference area of use (MCP) is simultaneous (*sensu* Van Moorter *et al.* 2013), therefore a given habitat type will be used more with increasing availability. It is important to underline that the preference for the habitat is *fixed*, hence variable use with availability does not correspond to a functional response. An alternative behavioral choice mechanism, for instance hierarchical, could be integrated in the movement model. In such an approach, the use of a given habitat type would be independent of its availability (Van Moorter *et al.* 2013). As of today, movement models have overwhelmingly used a simultaneous behavioral choice mechanism (Van Moorter *et al.* 2013). Although further research is required, empirical evidence suggests that simultaneous choice is appropriate (Schuck-Paim and Kacelnik 2007). In terms of IM-SAM, the above indicates that it is easier to distinguish different sequential habitat use patterns when the relative proportion of habitats available to the individual is similar.

In comparison with the simplified model used in De Groeve *et al.* (2016), we here simulated sequences within the true landscape context, giving more complex and variable patterns, and hence resulting in more complex dissimilarity trees, requiring methodological refinements to obtain robust classifications. In classification trees, the optimal number of clusters can be derived using many different approaches. Several SAM applications (e.g. Shoval and Isaacson 2007) define the cut-off value by visual exploration of clusters in dissimilarity trees without considering the robustness of the clusters. More objective methods use within- and between-cluster quality assessments, such as silhouette plots (Rousseeuw *et al.* 1987) and the Calinski-Harabaz Index (Caliński and Harabasz 1974), or, as often used in DNA analysis, cluster stability procedures based on bootstrapping (e.g. Jaccard bootstrapping, BJ). In our case, the sequences corresponded to simulated behaviors (i.e., discrete trajectories) that were obtained through a set of stochastic rules applied to real and highly heterogeneous environments. Hence, some variability in the output sequences, and so in the clustering, can be expected, especially when the proportion of alternative habitats is highly unequal within individual MCPs (see Fig. 3). For this reason, we extended the approach of De Groeve *et al.* (2016), based on median values in BJ, by combining a central tendency (BJ_{MD}), and a dispersion measure (BJ_{IQ}) of cluster quality into a unique index. We think that this procedure could be appropriate for other SAM applications, as it represents a semi-automated standardized approach.

One of the most important advantages of IM-SAM is the possibility to express sequences as multi-level habitat categories, as showed in other studies (De Groeve *et al.* 2016, Brum-Bastos *et al.* 2018). De Groeve *et al.* (2016) annotated trajectories with contextual information derived from two habitat variables (elevation and habitat openness) expressed as single character codes (i.e., high-open, high-closed, low-open, low-closed), and Brum-Bastos *et al.* (2018) instead generated character codes for each context variable which were then analyzed as multi-channel sequences. Here, we used a simple case of two alternative habitat types (open vs. closed) that showed promising sequential pattern variability in a single roe deer population (De Groeve *et al.* 2016). Note that for continuous or discretized habitat variables, which are expected to be spatially correlated, substitution weights are essential to correct for classes that are more similar to each other. For example, in the case of four habitat classes with different forest cover density (<25%, 25-50%, 50-75%, >75%), a forest cover density of <25% is more similar to the 25-50% category than to the >75% one. In the R package TraMineR such substitution weights can be based on automatic computation of transition rates, probability or user-defined (Gabadinho *et al.* 2011). While automatic computation of substitution weights is sufficient for exploration trees, we recommend assessing them directly by measuring spatial correlation at the relevant scale (i.e., median step length; see De Groeve *et al.* 2016) for simulation and classification trees. However, while the exploration phase can handle a large number of classes, the complexity of simulated sequential habitat use patterns increases with the number of habitat categories, hence we suggest using the exploration phase to identify the most essential for simulation and classification. Multi-channel sequence analysis, as proposed by Brum-Bastos *et al.* (2018), offers interesting future avenues for more complex combinations of habitat (or contextual) variables.

In IM-SAM, simulations of sequential habitat use patterns are performed in the individual reference area of use that were obtained with a simple geometric method (MCP 100 %). However, there is no limitation on using other methods to assess the area of use, for example, to overcome the sample size dependence of MCP (Spencer *et al.* 1990, Powell *et al.* 2000).

Ecological insights and geographical applications

In our study, 40% of the real sequences from all roe deer populations were classified as alternating patterns between open and closed habitats. This suggests that not only the proportion, but also the sequential order in which open and closed habitats are used, is an important metric for characterizing the space use strategy of individual roe deer. Activity and physiological circadian cycles, such as feeding-rumination, may explain the observed alternation between open and closed habitats. Indeed, roe deer are known mainly to select for forest and cover habitats during rumination and resting (Cederlund 1981, Mysterud and Østbye 1995), and to favor edges and open areas during peak foraging activity, at twilight and during the night (Pagon *et al.* 2013). This pattern may be less pronounced in areas with less human disturbance, such as for a Canadian elk population (Ensing *et al.* 2014). Indeed, in human-dominated European landscapes, habitat alternation is likely a behavioral response of ungulates to both landscape heterogeneity and temporal variation in human activities. Because rich open landscapes are often associated with higher risk of predation or disturbance, in such human-dominated environments, prey species must generally trade their acquisition of high-quality resources against risk avoidance (Godvik *et al.* 2009). By alternating between rich open areas and more closed forest habitats, with less forage but a higher degree of shelter, prey may hence resolve the risk-resource trade-off (Fraser and Huntingford 1986). In particular, wild ungulate species, including roe deer, generally use closed refuge habitats during daytime, when human disturbance is greater, and rich open habitats during night time (e.g. Bonnot *et al.* 2013; Padié *et al.* 2015).

A large proportion of real habitat use sequences of roe deer were also classified as homogeneous open, meaning that a non-negligible number of deer intensively used human-exploited agricultural lands during both day and night (i.e., crops: South-France, Southern Germany; husbandry: Switzerland; Aulak and Babinska-Werka 1990). Indeed, agricultural areas may simultaneously provide both high-quality food and cover resources for roe deer, at least during certain parts of the year (Hewison *et al.* 2001, Cimino and Lovari 2006, Bjørneraas *et al.* 2011). Homogeneous closed sequences, on the other hand, were more common in forest landscapes, but also occurred in agricultural landscapes, suggesting that some individuals are strictly associated with a given habitat type. Finally, our results indicate that most of the time, the sequential use of open and closed habitats by roe deer was not random, since only a small proportion of real sequences were tagged with a random pattern of habitat use.

A further step towards understanding the ecological significance of sequential habitat use would involve accounting for seasonal and individual variability, and to include further covariates, such as specific landscape features (e.g., fragmentation), or individual life-history traits (sex, age). IM-SAM 'tags' can be used with levels of a categorical variable in established statistical modeling frameworks, such as multivariate statistics (e.g. Jongman *et al.* 1995) or Generalized Linear Modeling (e.g. Pinheiro and Bates 2000; Zuur *et al.* 2009). SAM was originally applied to the ecological context as a spatio-temporal exploratory tool (De Groeve *et al.* 2016). IM-SAM takes this a step forward, opening up the potential to use spatio-temporal patterns as a variable in spatial ecological modeling.

IM-SAM applicability

While IM-SAM was applied here on animal habitat use sequences, human geography may also benefit from this novel framework. From a technical point of view, while several sequence analysis studies exist in time geography, to the best of our knowledge, robust classifications such as those obtained by bootstrapping and the use of BJ_{IQMD} have rarely been explored. Moreover, the routine applicability of IM-SAM could be supported by the use of a popular data analysis software (R version 3.4.1., R Core Team 2017; package TraMineR, Gabadinho *et al.* 2011). Conceptually, with the ongoing advances in human and animal tracking techniques, IM-SAM could ultimately be used as a tool to simultaneously compare patterns of space use in animals and humans. For example, mapping sequential animal and human space use in the same area could help understand if and how they differ or conflict. Alternatively, potential effects of traffic, recreation, hiking, cycling and other human activities could be assessed by modeling them as environmental drivers of sequential habitat use. Furthermore, after characterizing animal and human sequential space use, one could explore the sequential pattern of non-movement related metrics obtained through biologging, such as activity, heart rate, body temperature, or food intake (see Ropert-Coudert *et al.* 2005).

Acknowledgement

We thank Fondazione Edmund Mach and Special Research Fund (BOF) of Ghent University for their financial support to JDG under grant number 01sf2313. We thank Research Foundation – Flanders (FWO) for the travel grant for study abroad to Harvard University under grant number V417616N to JDG. FC was supported by the Sarah and Daniel Hrdy Fellowship during part of the development of this ms. We wish to thank the EuroDEER collaborative project to provide geospatial and environmental data (paper no. 009 of the EuroDEER series; www.euroungulates.org). The EuroDEER spatial database is hosted by Fondazione Edmund Mach. We are grateful to the Copernicus project for the use of free data (European Union, EEA - European Environment Agency). Here we confirm that all conditions are met required for free, full and open access to this data set, established by the Copernicus data and information policy Regulation (EU) No 1159/2013 of 12 July 2013. We are particularly grateful to Ferdinando Urbano for his support with data management, and to Lino Ometto for precious advice in data exploration.

The GPS data collection of the Fondazione Edmund Mach was supported by the Autonomous Province of Trento under grant number 3479 to FC (BECOCERWI—Behavioral Ecology of Cervids in Relation to Wildlife Infections), and by the help of the Wildlife and Forest Service of the Autonomous Province of Trento and the Hunting Association of Trento Province (ACT). The Norwegian participation was funded by the

Research Council of Norway (grant number 251112). The GPS data collection of the CEFS-INRA was supported by the "Move-It" ANR grant ANR-16-CE02-0010-02.

References:

Abbott, A., 1995. Sequence analysis: new methods for old ideas. *Annual review of sociology*, 21 (1), 93-113.

Andersen, R., Duncan, P., and Linnell, J.D.C., 1998. *The European roe deer: the biology of success.* Oslo: Scandinavian University Press.

Andrienko, G., Andrienko, N. and Heurich, M., 2011. An event-based conceptual model for context-aware movement analysis. *International Journal of Geographical Information Science*, **25** (9), 1347-1370.

Aulak, W. and Babińska-Werka, J., 1990. Use of agricultural habitats by roe deer inhabiting a small forest area. *Acta Theriologica*, 35 (1-2), 121-127.

Bar-David, S., Bar-David, I., Cross, P.C., Ryan, S.J., Knechtel, C.U. and Getz, W.M., 2009. Methods for assessing movement path recursion with application to African buffalo in South Africa. *Ecology*, 90 (9), 2467-2479.

Benhamou, S. and Riotte-Lambert, L., 2012. Beyond the Utilization Distribution: Identifying home range areas that are intensively exploited or repeatedly visited. *Ecological Modelling*, 227, 112-116.

Berger-Tal, O. and Bar-David, S., 2015. Recursive movement patterns: review and synthesis across species. *Ecosphere*, 6 (9), 1-12.

Bjørneraas, K., Solberg, E.J., Herfindal, I., Moorter, B.V., Rolandsen, C.M., Tremblay, J.P., Skarpe, C., Sæther, B.E., Eriksen, R. and Astrup, R., 2011. Moose Alces alces habitat use at multiple temporal scales in a human-altered landscape. *Wildlife Biology*, 17 (1), 44-54.

Bonnot, N., Morellet, N., Verheyden, H., Cargnelutti, B., Lourtet, B., Klein, F. and Hewison, A.J.M., 2013. Habitat use under predation risk, hunting, roads and human dwellings influence the spatial behaviour of roe deer. *European Journal of Wildlife Research*, 59, 185-193.

Brum-Bastos, V.S., Long, J.A. and Demšar, U., 2018. Weather effects on human mobility: a study using multi-channel sequence analysis. *Computers, Environment and Urban Systems*.

Cagnacci, F., Boitani, L., Powell, R.A. and Boyce, M.S., 2010. Animal ecology meets GPS-based radiotelemetry: a perfect storm of opportunities and challenges. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365, 2157-2162.

Cagnacci, F., Focardi, S., Heurich, M., Stache, A., Hewison, A.J., Morellet, N., Kjellander, P., Linnell, J.D., Mysterud, A., Neteler, M. and Delucchi, L., 2011. Partial migration in roe deer: migratory and resident tactics are end points of a behavioural gradient determined by ecological factors. *Oikos*, 120 (12), 1790-1802.

Caliński, T. and Harabasz, J., 1974. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, **3** (1), 1-27.

Cederlund, G., 1981. Daily and seasonal activity pattern of roe deer in a boreal habitat [Capreolus capreolus, Sweden]. *Swedish Wildlife Research*, 11 (8), 315-353.

Chavoshi, S.H., De Baets, B., Neutens, T., De Tré, G. and Van de Weghe, N., 2015. Exploring dance movement data using sequence alignment methods. *PloS one*, 10 (7), e0132452.

Cibien, C., Bideau, E., Boisaubert, B. and Maublanc, M.L., 1989. Influence of habitat characteristics on winter social organisation in field roe deer. *Acta Theriologica*, 34 (14), 219-226.

Cimino, L. and Lovari, S., 2003. The effects of food or cover removal on spacing patterns and habitat use in roe deer (Capreolus capreolus). *Journal of Zoology*, 261 (3), 299-305.

De Groeve, J., Van de Weghe, N., Ranc, N., Neutens, T., Ometto, L., Rota-Stabelli, O. and Cagnacci, F., 2016. Extracting spatio-temporal patterns in animal trajectories: an ecological application of sequence analysis methods. *Methods in Ecology and Evolution*, 7 (3), 369–379.

Delafontaine, M., Versichele, M., Neutens, T. and Van de Weghe, N., 2012. Analysing spatiotemporal sequences in Bluetooth tracking data. *Applied Geography*, 34, 659-668.

Demšar, U., Buchin, K., Cagnacci, F., Safi, K., Speckmann, B., Van de Weghe, N., Weiskopf, D. and Weibel, R., 2015a. Analysis and visualisation of movement: an interdisciplinary review. *Movement ecology*, **3** (1), 5.

Demšar, U., Buchin, K., van Loon, E.E. and Shamoun-Baranes, J., 2015b. Stacked space-time densities: a geovisualisation approach to explore dynamics of space use over time. *GeoInformatica*, 19 (1), 85-115.

Dodge, S., Bohrer, G., Weinzierl, R., Davidson, S.C., Kays, R., Douglas, D., Cruz, S., Han, J., Brandes, D. and Wikelski, M., 2013. The environmental-data automated track annotation (Env-DATA) system: linking animal tracks with environmental data. *Movement Ecology*, **1** (3), 1-14.

Ensing, E.P., Ciuti, S., de Wijs, F.A., Lentferink, D.H., ten Hoedt, A., Boyce, M.S. and Hut, R.A., 2014. GPS based daily activity patterns in European red deer and North American elk (Cervus elaphus): indication for a weak circadian clock in ungulates. *PLoS One*, 9 (9), e106997.

Fraser, D.F. and Huntingford, F., 1986. Feeding and avoiding predation hazard: The behavioral response of the prey. *Ethology*, 73, 56e68.

Jongman, R.H..G., Ter Braak, C.J.F. and Van Tongeren, O.F.R., 1995. *Data analysis in community and landscape ecology*. Cambridge: Cambridge University Press.

Gabadinho, A., Ritschard, G., Mueller, N.S. and Studer, M., 2011. Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*, 40 (4), 1-37.

Godvik, I.M.R., Loe, L.E., Vik, J.O., Veiberg, V., Langvatn, R., and Mysterud, A. 2009. Temporal scales, trade-offs and functional responses in habitat selection of red deer. *Ecology* 90, 699-710.

Goto, M. 2003. A chorus-section detecting method for musical audio signals. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) Proceedings; April 2003. 437-440.

Hennig, C., 2007. Cluster-wise assessment of cluster stability. *Computational Statistics and Data Analysis*, 52, 258-271.

Hewison, A.J., Vincent, J.P., Joachim, J., Angibault, J.M., Cargnelutti, B. and Cibien, C., 2001. The effects of woodland fragmentation and human activity on roe deer distribution in agricultural landscapes. *Canadian journal of zoology*, 79 (4), 679-689.

Kämmerle, J.L., Brieger, F., Kröschel, M., Hagen, R., Storch, I. and Suchant, R., 2017. Temporal patterns in road crossing behaviour in roe deer (Capreolus capreolus) at sites with wildlife warning reflectors. *PLoS One*, 12 (9), 1-18.

Kays, R., Crofoot, M.C., Jetz, W. and Wikelski, M., 2015. Terrestrial animal tracking as an eye on life and planet. *Science* 348, aaa2478.

Magle, S.B., Hunt, V.M., Vernon, M. and Crooks, K.R., 2012. Urban wildlife research: past, present, and future. *Biological Conservation*, 155, 23-32.

Manly, B.F.J., McDonald, L.L., Thomas, D.L., McDonald, T.L. and Erickson, W.P., 2002. Introduction to Resource Selection Studies. In: *Resource selection by animals: statistical design and analysis for field studies*. Dordrecht: Kluwer Academic Publishers, 1-15.

Mysterud, A. and Østbye, E. 1995. Bed-site selection by European roe deer (*Capreolus capreolus*) in southeastern Norway during winter. *Canadian Journal of Zoology* 73, 924-932.

Mysterud, A., Lian, L.B. and Hjermann, D.Ø., 1999. Scale-dependent trade-offs in foraging by European roe deer (Capreolus capreolus) during winter. *Canadian Journal of Zoology*, 77 (9), 1486-1493.

Needleman, S.B. and Wunsch, C.D., 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology*, 48 (3), 443-453.

Padié, S., Morellet, N., Hewison, A.J.M., Martin, J.L., Bonnot, N., Cargnelutti, B. and Chamaillé-Jammes, S., 2015. Roe deer at risk: teasing apart habitat selection and landscape constraints in risk exposure at multiple scales. *Oikos*, 124 (11), 1536-1546.

Pagon, N., Grignolio, S., Pipia, A., Bongi, P., Bertolucci, C. and Apollonio, M., 2013. Seasonal variation of activity patterns in roe deer in a temperate forested area. *Chronobiology international*, 30 (6), 772-785.

Péron, G., Fleming, C. H., de Paula, R. C., & Calabrese, J. M. 2016. Uncovering periodic patterns of space use in animal tracking data with periodograms, including a new algorithm for the Lomb-Scargle periodogram and improved randomization tests. Movement Ecology, 4, 19.

Péron, G., Fleming, C. H., de Paula, R. C., Mitchell, N., Strohbach, M., Leimgruber, P., & Calabrese, J. M. 2017. Periodic continuous-time movement models uncover behavioral changes of wild canids along anthropization gradients. Ecological Monographs, 87, 442–456

Pinheiro, J.C. and Bates, D.M., 2000. *Mixed-effects models in S and S-PLUS*. New York: Springer.

Podgórski, T., Baś, G., Jędrzejewska, B., Sönnichsen, L., Śnieżko, S., Jędrzejewski, W. and Okarma, H., 2013. Spatiotemporal behavioral plasticity of wild boar (Sus scrofa) under contrasting conditions of human pressure: primeval forest and metropolitan area. *Journal of Mammalogy*, 94 (1), 109-119.

Polansky, L., Wittemyer, G., Cross, P.C., Tambling, C.J. and Getz, W.M., 2010. From moonlight to movement and synchronized randomness: Fourier and wavelet analyses of animal location time series data. *Ecology*, 91, 1506-1518.

Powell, R. A. 2000. Animal home ranges and territories and home range estimators. - In: Boitani, L. and Fuller, T. K. (eds), Research techniques in animal ecology: controversies and consequences. Columbia University Press, New York, 65–103.

R Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Riotte-Lambert, L., Benhamou, S. and Chamaillé-Jammes, S., 2013. Periodicity analysis of movement recursions. *Journal of theoretical biology*, 317, 238-243.

Riotte-Lambert, L., Benhamou, S. and Chamaillé-Jammes, S., 2017. From randomness to traplining: a framework for the study of routine movement behavior. *Behavioral Ecology*, 28 (1), 280-287.

Ropert-Coudert, Y. and Wilson, R.P., 2005. Trends and perspectives in animal-attached remote sensing. *Frontiers in Ecology and the Environment*, 3 (8), 437-444.

Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20, 53-65.

Schuck-Paim, C. and Kacelnik, A., 2007. Choice processes in multialternative decision making. *Behavioral Ecology*, **18** (3), 541-550.

Shoval, N. and Isaacson, M., 2007. Sequence alignment as a method for human activity analysis in space and time. *Annals of the Association of American geographers*, 97 (2), 282-297.

Spencer, S.R., Cameron, G.N. and Swihart, R.K., 1990. Operationally defining home range: temporal dependence exhibited by hispid cotton rats. *Ecology*, *71*(5), 1817-1822.

Toor, M.L., Newman, S.H., Takekawa, J.Y., Wegmann, M. and Safi, K., 2016. Temporal segmentation of animal trajectories informed by habitat use. *Ecosphere*, **7** (10), e01498.

Van Moorter, B., Visscher, D., Herfindal, I., Basille, M. and Mysterud, A., 2013. Inferring behavioural mechanisms in habitat selection studies getting the null-hypothesis right for functional and familiarity responses. *Ecography*, **36** (3), 323-330.

Wagner, R.A. and Fischer, M.J., 1974. The string-to-string correction problem. *Journal of the Association for Computing Machinery (JACM)*, **21** (1), 168-173.

Wilson, C., 2006. Reliability of sequence-alignment analysis of social processes: Monte Carlo tests of ClustalG software. Environment and Planning A, 38 (1), 187-204.

Wilson, C., 2008. Activity patterns in space and time: calculating representative Hagerstrand trajectories. *Transportation*, 35 (4), 485-499.

Wittemyer, G., Polansky, L., Douglas-Hamilton, I. and Getz, W.M., 2008. Disentangling the effects of forage, social rank, and risk on movement autocorrelation of elephants using Fourier and wavelet analyses. *Proceedings of the National Academy of Sciences*, 105 (49), 19108-19113.

Xavier, G. and Dodge, S., 2014. An exploratory visualization tool for mapping the relationships between animal movement and the environment. *In: MapInteract 2014 – Proceedings of the 2nd ACM SIGSPATIAL International Workshop on MapInteraction,* 36-42.

Van Moorter, B., Visscher, D., Benhamou, S., Börger, L., Boyce, M.S. and Gaillard, J.M., 2009. Memory keeps you at home: a mechanistic model for home range emergence. *Oikos*, *118* (5), 641-652.

Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A., Smith, G.M., 2009. *Mixed Effects Models and Extensions in Ecology with R*. Berlin: Springer.

Table 1. Percentage of real sequences classified into the 17 most common cluster types (different from the most common cluster types for simulated sequences), and in the simplified classification. See paragraph d. in Methods 'Identification of cluster types' for the description of the acronyms.

o_u	**o_u_a24**	o_u_a24_a33	o	**c_u**	c	c_u_a42	c_u_a42_a33_a24	c_u_a42_a33	**a33**	a33_a42	a33_a24	**a42**	a42_a33	**a24**	a24_a33	u	**Total**
16.91	**9.96**	3.05	1.22	**12.49**	4.83	4.55	1.22	1.00	**11.85**	5.26	0.91	**10.12**	0.52	**8.07**	1.57	4.02	**97.55**
o **31.51**				**c** **24.38**					**a33** **18.68**			**a42** **10.90**		**a24** **10.07**		**u** **4.46**	

Table 2. Comparison between SAM and IM-SAM, distinguishing the general procedure (a), the three-step framework (exploration, b; simulation, c; classification, d) and the pivotal summarized advantage of IM-SAM (e). The steps that were identical between SAM and IM-SAM are in *italics*.

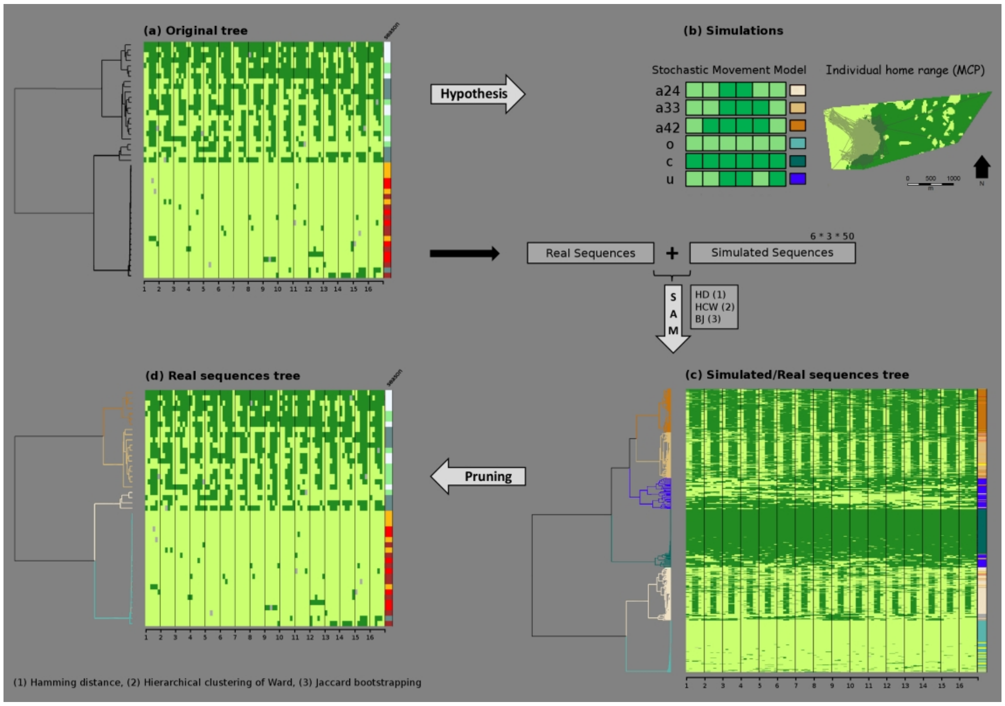
		SAM	**IM-SAM**
a	**Distance metric**	*Hamming Distance*	*Hamming Distance*
	substitution weights	Yes	No
	Extendible to multiple variables	*Yes*	*Yes*
	Clustering Algorithm	*Hierarchical Clustering of Ward*	*Hierarchical Clustering of Ward*
	Bootstrapping	Bootstrap median	Bootstrap median & interquartile range
b	**Exploration trees**	Population level tree	Individual or population level tree
c	**movement simulation context**	Nine simulated arenas *	Real movement context (MCP)
	Movement model	*Biased and correlated random walk*	*Biased and correlated random walk*
		Simulated patterns: a33, o, c, r	Simulated patterns: a24, a33, a42, o, c, r §
		single selection rule	three selection rules
	Simulation trees	Tree for each arena	Tree for each individual
	Matching of real sequences to a simulation arena	yes **	No need (individual trees)
d	**Classification trees**	Sequences from multiple individuals	Sequences from single individual
e	**Multi-population framework**	No	Yes

** Nine simulated arenas that represent the distribution and composition of real home ranges.*
*** Real sequences are matched to simulated arenas by measuring the proportion of available habitat at sequence level.*
§ Under the hypothesis that ungulates maintain a disturbance and predator avoidance strategy, using mainly open habitat during the night and closed habitat during the day, three types of alternation were simulated in relation to temporal and spatial variation in sunrise and sunset.

Figure 1. Workflow chart of the procedure to classify spatio-temporal habitat use patterns of individual animal trajectories using Individual Movement-Sequence Analysis Method (IM-SAM). Trees represent sequence dissimilarities between habitat use sequences, indicated by branch lengths. Each tree leaf corresponds to one biweekly sequence, which was visualized beside the tree, together with a color-coded bar representing a variable related to each sequence. See main text for details.

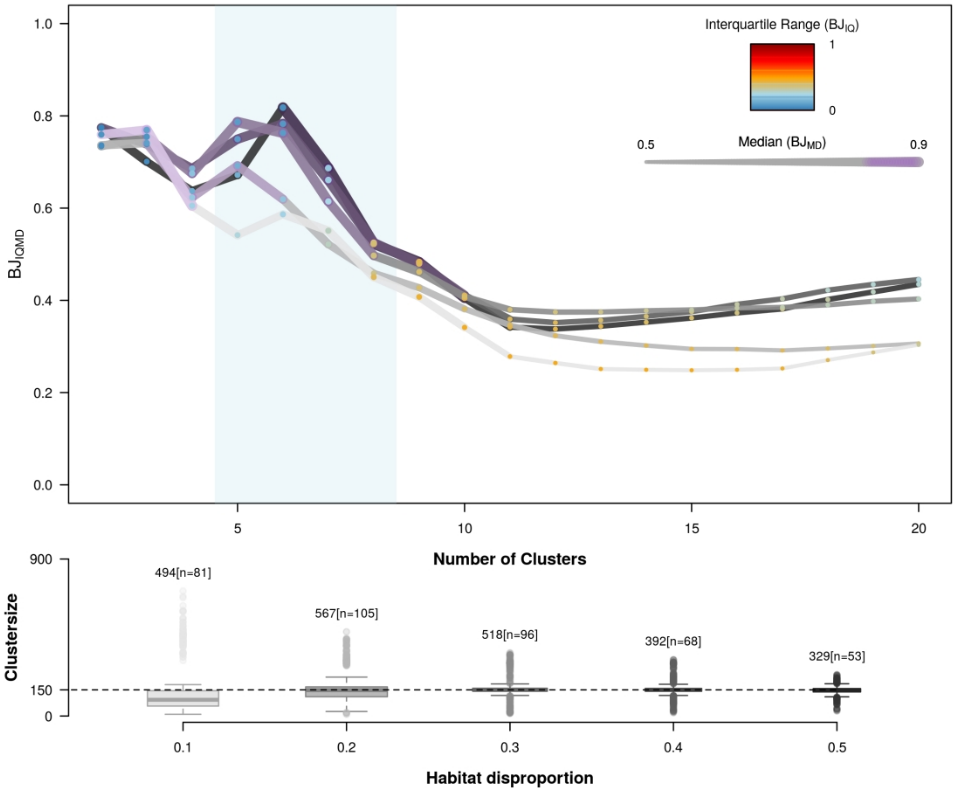
Figure 2. **Top plot:** Trend lines representing the combined bootstrap index BJ_{IQMD} (combining the bootstrap median BJ_{MD} and the bootstrap interquartile range BJ_{IQ}), for different cut-offs (2-20 clusters) in all 404 individual-based simulation trees. The output BJ_{IQMD} values are classified by the relative habitat proportion in the individual MCPs (gradient from light to dark gray, from 0.1 to 0.5). The plot also represents the contribution of BJ_{MD} and BJ_{IQ} to the combined index $BJ_{IQMD.}$ Specifically, the thickness of the lines corresponds to BJ_{MD}; when BJ_{MD} is larger than 0.8, a purple shading is added to the gray BJ_{IQMD} trend lines. The color of the dots along the trend lines represents BJ_{IQ}, with values decreasing from orange to blue (bright blue: $BJ_{IQ} < 0.2$). The transparent light blue region is the window corresponding to the second peak in BJ_{IQMD} that was chosen as the cut-off criterion for final simulation trees. **Bottom plot:** Boxplots visualize how the total 900 sequences simulated for each 404 MCPs are distributed between clusters, when the cut-off based on BJ_{IQMD} is used to define the corresponding simulated trees. The trees are classified by the relative habitat proportion in the MCPs (from 0.1, to 0.5- gradient of gray as in the top panel). When there is no habitat preponderance (i.e., relative habitat proportion of 0.4 or 0.5), the sequences are almost equally distributed between clusters (i.e., median cluster size close to 150, with very few outliers).

Figure 3. **Bottom panel - Cluster types:** colored circles represent all sixteen main cluster types identified in simulation trees, scaled by the number of clusters of that type (actual number of clusters in brackets). The color is the legend for the top-right figure (see below). **Top panel left – Cluster composition (a)**: frequency of occurrence of cluster types composed by up to five sequential habitat use patterns. The colored gradient indicates the proportion of open habitat in MCPs on which simulation trees were based. **Top panel right – Cluster composition (b)**: relative proportion of cluster types in simulated trees (main y-axis), plotted against the proportion of open habitat in MCPs (main x-axis).



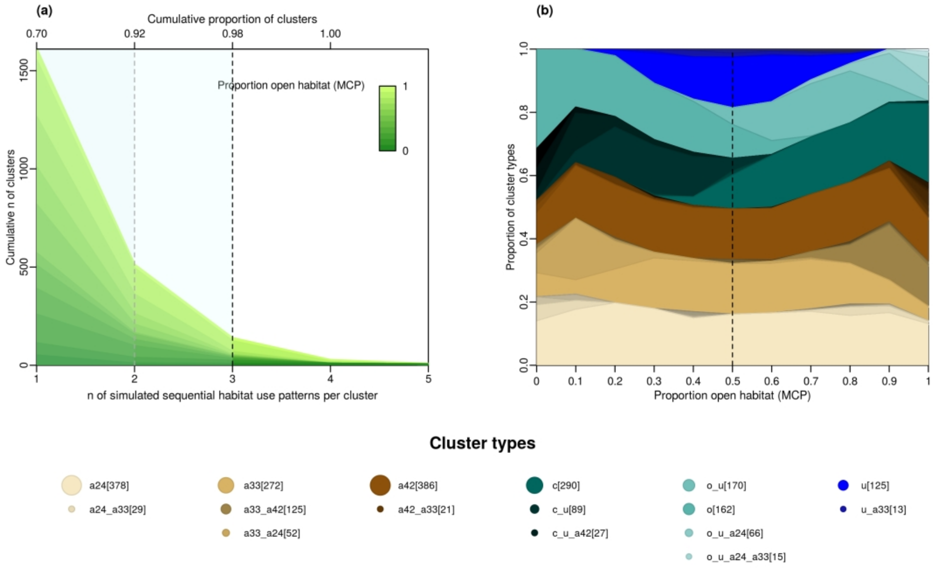
Workflow chart of the procedure to classify spatio-temporal habitat use patterns of individual animal trajectories using Individual Movement-Sequence Analysis Method (IM-SAM). Trees represent sequence dissimilarities between habitat use sequences, indicated by branch lengths. Each tree leaf corresponds to one biweekly sequence, which was visualized beside the tree, together with a color-coded bar representing a variable related to each sequence. See main text for details.

239x167mm (300 x 300 DPI)



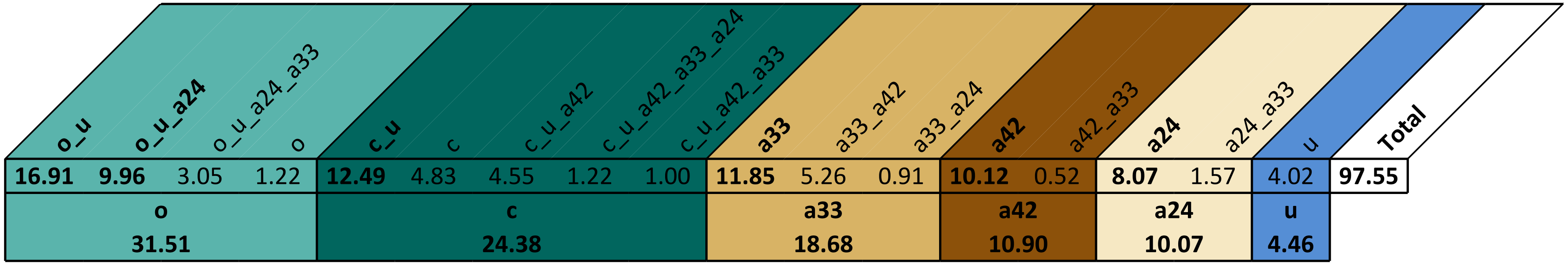
Top plot: Trend lines representing the combined bootstrap index BJ_{IQMD} (combining the bootstrap median BJ_{MD} and the bootstrap interquartile range BJ_{IQ}), for different cut-offs (2-20 clusters) in all 404 individual-based simulation trees. The output BJ_{IQMD} values are classified by the relative habitat proportion in the individual MCPs (gradient from light to dark gray, from 0.1 to 0.5). The plot also represents the contribution of BJ_{MD} and BJ_{IQ} to the combined index BJ_{IQMD}. Specifically, the thickness of the lines corresponds to BJ_{MD}; when BJ_{MD} is larger than 0.8, a purple shading is added to the gray BJ_{IQMD} trend lines. The color of the dots along the trend lines represents BJ_{IQ}, with values decreasing from orange to blue (bright blue: $BJ_{IQ} < 0.2$). The transparent light blue region is the window corresponding to the second peak in BJ_{IQMD} that was chosen as the cut-off criterion for final simulation trees. **Bottom plot:** Boxplots visualize how the total 900 sequences simulated for each 404 MCPs are distributed between clusters, when the cut-off based on BJ_{IQMD} is used to define the corresponding simulated trees. The trees are classified by the relative habitat proportion in the MCPs (from 0.1, to 0.5- gradient of gray as in the top panel). When there is no habitat preponderance (i.e., relative habitat proportion of 0.4 or 0.5), the sequences are almost equally distributed between clusters (i.e., median cluster size close to 150, with very few outliers).

203x169mm (300 x 300 DPI)

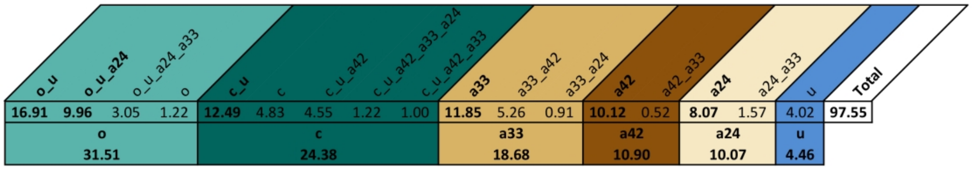


Bottom panel - Cluster types: colored circles represent all sixteen main cluster types identified in simulation trees, scaled by the number of clusters of that type (actual number of clusters in brackets). The color is the legend for the top-right figure (see below). **Top panel left – Cluster composition (a):** frequency of occurrence of cluster types composed by up to five sequential habitat use patterns. The colored gradient indicates the proportion of open habitat in MCPs on which simulation trees were based. **Top panel right – Cluster composition (b):** relative proportion of cluster types in simulated trees (main y-axis), plotted against the proportion of open habitat in MCPs (main x-axis).

254x169mm (300 x 300 DPI)



o_u	**o_u_a24**	o_u_a24_a33	o	**c_u**	c	c_u_a42	c_u_a42_a33_a24	c_u_a42_a33	**a33**	a33_a42	a33_a24	**a42**	a42_a33	**a24**	a24_a33	u	**Total**
16.91	**9.96**	3.05	1.22	**12.49**	4.83	4.55	1.22	1.00	**11.85**	5.26	0.91	**10.12**	0.52	**8.07**	1.57	4.02	**97.55**
o **31.51**				**c** **24.38**					**a33** **18.68**			**a42** **10.90**		**a24** **10.07**		**u** **4.46**	



o_u	**o_u_a24**	o_u_a24_a33	o	**c_u**	c	c_u_a42	c_u_a42_a33_a24	c_u_a42_a33	**a33**	a33_a42	a33_a24	**a42**	a42_a33	**a24**	a24_a33	u	**Total**
16.91	**9.96**	3.05	1.22	**12.49**	4.83	4.55	1.22	1.00	**11.85**	5.26	0.91	**10.12**	0.52	**8.07**	1.57	4.02	**97.55**
o				**c**					**a33**			**a42**		**a24**		**u**	
31.51				**24.38**					**18.68**			**10.90**		**10.07**		**4.46**	

Percentage of real sequences classified into the 17 most common cluster types (different from the most common cluster types for simulated sequences), and in the simplified classification. See paragraph d. in Methods 'Identification of cluster types' for the description of the acronyms.