

## Extracting spatio-temporal patterns in animal trajectories: an ecological application of sequence analysis methods

Johannes De Groeve<sup>1,2</sup>, Nico Van de Weghe<sup>1</sup>, Nathan Ranc<sup>2,3</sup>, Tijs Neutens<sup>1</sup>, Lino Ometto<sup>4</sup>, Omar Rota-Stabelli<sup>4</sup> and Francesca Cagnacci<sup>2</sup>

<sup>1</sup>Department of Geography, Ghent University, Krijgslaan 281, 9000 Gent, Belgium; <sup>2</sup>Department of Biodiversity and Molecular Ecology, Research and Innovation Centre, Fondazione Edmund Mach, Via E. Mach 1, 38010 San Michele all' Adige, Trentino, Italy; <sup>3</sup>Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA; and <sup>4</sup>Department of Sustainable Agro-Ecosystems and Bioresources, Research and Innovation Centre, Fondazione Edmund Mach, Via E. Mach 1, 38010 San Michele all' Adige, Trentino, Italy

### Summary

1. Digital tracking technologies have considerably increased the amount and quality of animal trajectories, enabling the study of habitat use and habitat selection at a fine spatial and temporal scale. However, current approaches do not yet explicitly account for a key aspect of habitat use, namely the sequential variation in the use of different habitat features.

2. To overcome this limitation, we propose a tree-based approach that makes use of sequence analysis methods, derived from molecular biology, to explore and identify ecologically relevant sequential patterns in habitat use by animals. We applied this approach to ecological data consisting of simulated and real trajectories from a roe deer population (*Capreolus capreolus*), expressed as ordered sequences of habitat use.

3. We show that our approach effectively captured spatio-temporal patterns of sequential habitat use by roe deer. In our case study, individual sequences were clustered according to the sequential use of the elevation gradient (first order) and of open/closed habitats (second order). We provided evidence for several behavioural processes, such as migration and daily alternating habitat use. Some unexpected patterns, such as homogeneous sequences of use of open habitat, could also be identified.

4. Our findings advocate the importance of dealing with the sequential nature of movement data. Approaches based on sequence analysis methods are particularly useful and effective since they allow exploring temporal patterns of habitat use in a synthetic and visually captive manner. The proposed approach represents a useful and effective way to classify individual movement behaviour across populations and species. Ultimately, this method can be applied to explore the temporal scale of ecological processes based on movement.

**Key-words:** autocorrelation, distance, exploratory analysis, habitat use, Hamming, roe deer, spatio-temporal sequences, trees

### Introduction

Recent advances in digital tracking technology and increased availability of high-resolution environmental data by remote sensing have facilitated the collection of spatio-temporal series of animal-borne data (Cagnacci *et al.* 2010). Application of satellite navigation technology (e.g. Global Positioning System, GPS) to individual animals allows recording temporal sequences of animal locations at an unprecedented spatio-temporal resolution. Moreover, by projecting these locations onto spatial layers, including satellite images, it is possible to obtain robust and standardized information about the habitat of these animals (Urbano *et al.* 2010).

At present, an array of both exploratory and inferential methods is available to the analyst to investigate the relation between animal movement and the use of habitat. Exploratory methods apply multivariate analysis techniques (e.g. general niche-environment system factor analysis, GNESFA, the K-select analysis, (canonical) outlying mean index analysis; see R-package AdehabitatHS of Calenge (2006) for an overview) to identify relevant variables describing the habitat (or the realized niche) of a population. Similarly, decision tree learning methods, such as random forest and CART modelling, are data mining techniques that present decision rules for classifying a set of data based on associated explanatory variables (see R-package rpart of Therneau, Atkinson & Ripley (2014)). Conversely, inferential methods mainly consist of a variety of regression models testing the disproportion between used and available habitat units (i.e. habitat selec-

\*Correspondence author. E-mail: degroevjohannes@gmail.com

tion; Johnson 1980), such as resource selection functions, RSF (Boyce *et al.* 2002) and step selection functions, SSF (Fortin *et al.* 2005). In essence, exploratory methods offer a description of animals' habitat, whereas inferential methods allow to test specific hypotheses (Calenge & Basille 2008). In this sense, the first can be used to select explanatory variables that are relevant for the application of the latter (e.g. Calenge 2006; Wittemyer *et al.* 2008; Dray, Royer-Carenzi & Calenge 2010).

Despite the proliferation of exploratory methods, current approaches rarely evaluate the sequential use of habitats by animals, that is the sequence of locations (trajectory) vs. the underlying ordered pattern of habitat use. In movement ecology, temporal patterns have been addressed, for example, by exploration of temporal autocorrelation of movement parameters (Wittemyer *et al.* 2008; Dray, Royer-Carenzi & Calenge 2010). We wish to draw attention on the meaningfulness of temporal patterns when describing habitat choices. For example, the same proportion of habitat use in a certain time interval may correspond to very different sequential patterns. An animal may continuously use a single habitat type, then switch to another, or, in contrast, alternate the use of both. Such spatio-temporal patterns may correspond to alternative space-use tactics, and find a deep ecological significance. Currently, insights in the spatial patterns of use of multiple habitat-related variables are easy to obtain, for example using suitability maps (e.g. Calenge 2006); however, very few methods provide insights into spatio-temporal patterns combined. An interesting publication in that direction comes from Benhamou & Riotte-Lambert (2012) presenting a framework using movement-based kernel density estimation (utilization distribution) and computation of residence time combined to explore the areas of intensive use. Here, we are interested into methods to visually explore the sequential and thus temporal structure of habitat use.

In other research areas, such as geo-visual analytics, important progress has been made in terms of visually exploring sequential data at variable spatio-temporal scales (Andrienko, Andrienko & Heurich 2011). Buchin, Dodge & Speckmann (2012), for example, developed a geometric algorithm for trajectory clustering that takes into account environmental context parameters such as temperature and habitat type. In sociology, on the other hand, the link between sequential order of human behaviour and space use has been investigated using sequence analysis methods (SAMs) (Abbott 1995). This technique is principally used in the field of bioinformatics to evaluate the degree of similarity among DNA or protein sequences, but has also been applied successfully in transportation science (Wilson 2008), tourism research (Shoval & Isaacson 2007) and indoor navigation (Delafontaine *et al.* 2012).

Sequence analysis methods to our knowledge has never been applied to explore spatio-temporal patterns in sequential habitat use by animals. The essence of this approach is the possibility to 'extract' ordered sequences of habitat classes occupied along trajectories by means of clusters, which can be conveniently visualized in trees and validated by measures of

statistical reliability. Moreover, SAM allows to deal with two common issues of GPS-based location data sets: missing data points (i.e. acquisition failures by GPS sensors; Frair *et al.* 2010) and spatial correlation (Dray, Royer-Carenzi & Calenge 2010).

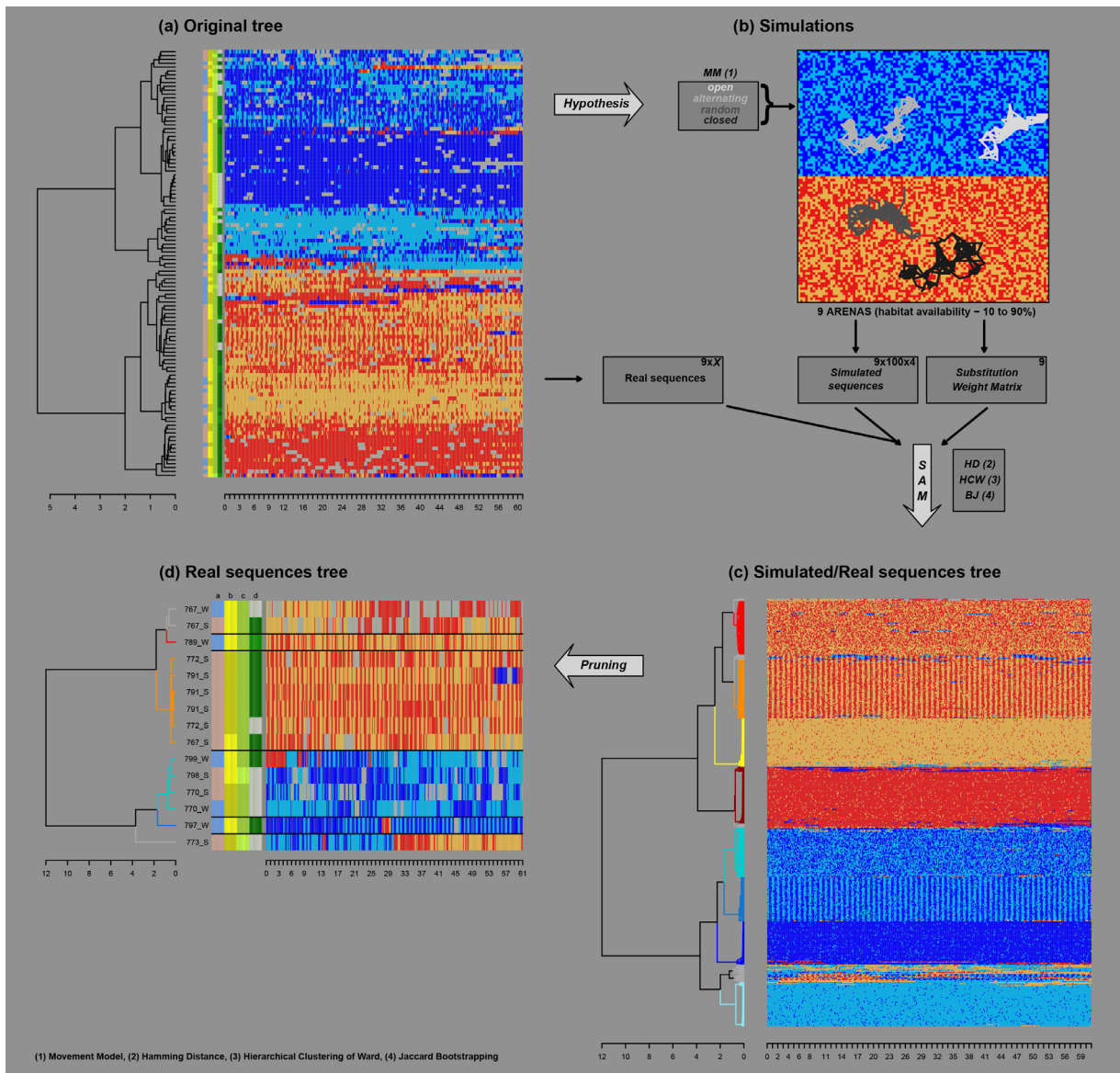
In this study, we aimed to evaluate the applicability of SAM to movement ecology data for exploratory purposes by analysing both simulated trajectories and time-stamped locations of individually tracked roe deer (*Capreolus capreolus*) from a partially migratory alpine population. The analysis consisted of several steps (Fig. 1). We first produced a classification tree based on bimonthly sequences of habitat use by individual roe deer. We used this first exploratory classification to hypothesize potential patterns of sequential habitat use. Then, we produced simulated trajectories with those patterns, at different proportion of habitat availability, and classified them in trees. We then re-added the real trajectories to simulation trees, while accounting for their relative proportion of habitat availability. Finally, we evaluated the biological relevance of such classification on the basis of ecological predictions. This way, we explored spatio-temporal patterns of real trajectories and evaluated them in a simulated experimental setting.

## Materials and methods

### STUDY AREA AND REAL TRAJECTORIES

The studied animal population consisted of 26 European roe deer equipped with a GPS collar (GPS-Plus D, Vectronic Aerospace GmbH), of which 16 were females and 10 males. Six of them were collared as fawns (i.e. less than 1 year old; 2 females and 4 males), one as yearling (i.e. one female between 1 and 2 year old), while all others were collared as adults. Figure 2 pictures the cumulative sum of the 90% fixed-kernel home ranges (KDE) with reference smoothing parameter (href; Worton 1995) of individual roe deer using Home Range Extension (Rodgers & Carr 1998). This area extends across the Monte Bondone–Monte Stivo range, west of Trento and Adige valley and east of Valle dei Laghi, in north-east Italy (46°4'N, 11°7'E). Elevation ranges between 200 and 2300 m above sea level (m a.s.l.). Along this altitudinal gradient, climate is extremely varied, ranging from semi-Mediterranean and temperate (<1000 m a.s.l. defined as 'low elevation') to semi-alpine and alpine (>1000 m a.s.l. defined as 'high elevation') conditions. The study area is mainly covered by broad-leaved, coniferous and mixed forest (defined as 'closed habitat', representing 50% and 75% of high and low elevations, respectively), alternated by pastures (defined as 'open habitat', representing 50% and 25% of high and low elevations, respectively). Relevantly, the high-elevation range is mainly constituted by protected land, whereas the low elevation is not and is characterized by more anthropic land use.

The sampling period spanned from 2005 to 2008, during which GPS collars yielded a total of 54 845 time-stamped locations. The interval and duration of sampling were different among individuals, depending on date of capture and battery exhaustion. Locations were acquired at a pre-determined temporal interval  $\Delta t$  of 4 h at fixed time stamps (0, 4, 8, 12, 16 and 20 h) except for cold winter months (January and February), when  $\Delta t$  was 6 h (0, 6, 12 and 18 h) in order to save battery. We linked GPS locations to two geographical parameters retrieved from

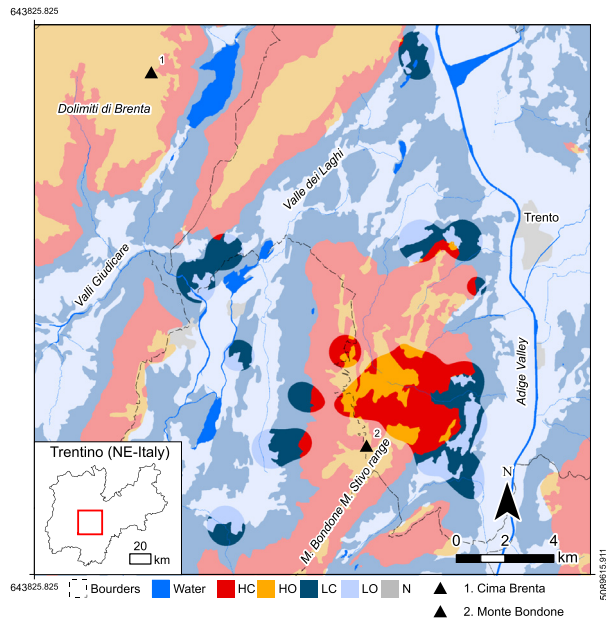


**Fig. 1.** Flowchart of the complete process to explore spatio-temporal habitat use patterns of real trajectories by sequence similarity analysis. In essence, based on simple classification trees of real trajectories (a), we simulated trajectories with a priori defined patterns of sequential habitat use (b), which in turn were used to produce classification trees accounting for spatial correlation, at different proportion of habitat availability (b/c). We then introduced the real trajectories (c), accounting for the relative proportion of habitat availability, and assessed their classification into tree clusters resulting from the simulation exercise (c/d).

remote sensing sources: habitat openness [EEA-Corine Landcover (CLC) 2006 IV, European Environment Agency, EEA (2010), resolution = 100 m] and elevation [NASA-ASTER GDEM, Ministry of Economy, Trade, and Industry of Japan, METI, National Aeronautics and Space Administration of the United States, NASA (2012) = 30 m]. We combined and reclassified environmental parameters into four classes: high closed (HC), high open (HO), low closed (LC), low open (LO) (Table 1 and Fig. 2). In this application, we used four classes to reduce the complexity of data analysis for demonstration. On the one side, the classes correspond to well-defined vegetation successional types in the alpine habitat of this population (Cagnacci *et al.* 2011); on the other side, they are meaningful for roe deer habitat use traits, since they are known to prefer ecotonal habitats and forest edges (Tufto, Andersen & Linnell 1996).

Then, we recoded sequences of locations into sequences of the environmental classes above, adding asterisks to account for missing locations due to acquisition failure. More precisely, the input sequences for SAM describe the habitat use by individual animals at regular time stamps (0, 4, 8, 12, 16 and 20 h) over a period of 2 months. While SAM is able to deal with missing data, if they are too frequent, they can over-fragment the sequence and thus bias the similarity measurement. To avoid a bias in the downstream analyses, we therefore excluded bimonthly sequences with more than 40% of missing data. After removal, the input file consisted of a total of 111 sequences (min 21, max 24 per bimonthly period), which were reclassified as belonging to summer (May–October) or winter (November–April) season, based on snowfalls and typical alpine climate (Ramanzin, Sturaro & Zanon 2007; Cagnacci *et al.* 2011). Only seven sequences could be retained for





**Fig. 2.** The study area, indicated by the darker irregular shaped polygons, was computed as the cumulative sum of home ranges of 26 individually marked roe deer (kernel density estimator, 90% polygons). For codes of habitat classes, see Table 1. N: urban areas.

**Table 1.** Reclassification of the environmental parameters elevation and habitat openness resulting in combined classes (coded as HC, HO, LC and LO)

Elevation	Habitat openness	Habitat use classes
Low (<1000 m)	Closed*	LC
Low	Open†	LO
High (>1000 m)	Closed	HC
High	Open	HO

\*Forest: Corine Landcover classes 311, 312, 313, 323, 324.

†No Forest: All other Corine Landcover classes, except inland water.

January–February; therefore, we excluded the whole period from further analysis. Finally, sequences were associated with sex, age (fawn, yearling or adult) and migration occurrence (migrant, non-migrant, no data), of each individual. We obtained a first visualization of the habitat use patterns represented in our sample by creating a simple classification tree of the 111 sequences (Fig. 1a). This classification tree was based on Hamming distance (Gabadinho *et al.* 2011; see below for more details), without accounting for spatial correlation, nor habitat availability. Thus, this first classification tree does not allow to objectively classify the sequential use of habitat types, but can be used to build hypotheses for expected models of habitat use.

#### SIMULATION ARENAS AND SIMULATED TRAJECTORIES

We simulated nine squared gridded arenas, each composed of 10 000 squared cells of 100 × 100 m mirroring the availability of environmental classes present in the real settings (Fig. 1b, Appendix S1, Supporting Information). Since elevation classes are highly clustered in our study area (see Fig. 2), we split each arena into two equal parts, corresponding to high and low elevations. We obtained the final four categories by

randomly assigning all cells to either open or closed habitats, with varying proportions from 10 to 90% in each arena. Thus, simulated landscapes covered all possible habitat prevalence, which may occur within individual home ranges in this specific study area.

Within each arena, we generated simulated sequential habitat use using a simple spatially explicit stochastic movement model (Fig. 1b, grey box 'MM'; Appendix S2). For the simulated trajectories, we chose 4 different patterns of habitat selection, based on preliminary observations on the classification tree of real trajectories (Fig. 1a), and previous knowledge on roe deer ecology. We thus distinguished homogeneous use of closed or open habitats, and random and alternating use (i.e. day–night patterns) of open and closed habitats. The random pattern represented the 'control' in our simulated experimental settings. The homogeneous closed and the alternation between closed and open were the expected patterns according to roe deer ecology and specifically the known preference for forest and ecotonal habitats (Tufto, Andersen & Linnell 1996). Homogeneous open represented the alternative hypothesis. We simulated 100 trajectory replicates for each of the nine arenas and four behaviours of habitat selection (total of 3600 sequences, Fig. 1b, grey box 'simulated sequences'). Because sequential use of elevation was strongly dependent on the release location of the simulated agents, each of the 100th set of simulations had the same seed random locations across arenas and behaviours (i.e. trajectories had the same seed in groups of 36). We then trimmed the simulated trajectories to match the length of the real bimonthly roe deer trajectories, that is 366 steps. Finally, simulated sequences of habitat use were extracted from these calibrated trajectories.

#### GENERAL PROCEDURE OF SEQUENCE SIMILARITY ANALYSIS FOR CALCULATION OF DISSIMILARITY TREES

Sequence analysis methods are based on sequence similarity measures that are used to identify groups of sequences showing similar behaviour. The input of such analysis always relies on a *dissimilarity matrix*, which provides the dissimilarity, or 'distance' among all possible pairs of sequences. Among the available algorithms, we chose Hamming distance (HD) to ascertain the dissimilarity matrix, as it is considered the most suitable for sequences with a temporal dimension. HD relies solely on two operations: identity and substitution, and in fact, it computes the minimum number of substitutions to equate a number of sequences of equal length (Gabadinho *et al.* 2011). In a more optimized HD, also weights can be assigned to substitutions; that is, HD computation can be based on a *substitution weight matrix*. Figure 3 gives a conceptual example of two alternative HD of the character strings 'Kapeolo' and 'Capriolo', respectively, the word for roe deer in Esperanto and Italian. Both distinguish six identities and two substitutions, but differ from each other in weights assigned to substitutions. The substitution between the letters K and C gets a lower weight (i.e. probability) in 'a' (substitution score = 0.4, Fig. 3a) than in 'b' (substitution score = 1, Fig. 3b), since the former HD takes into account the phonetic similarity. Consequently, the total dissimilarity in 'a' will be lower than in 'b'.

In our case, the dissimilarity matrix computed by the HD algorithm was based on substitution weights for all combinations of habitat classes, and constrained to their availability and distribution, which together determine the patterns of spatial correlation. Therefore, we derived the *substitution weights from spatial correlation* of habitat classes in the simulated arenas (Fig. 1b, grey box 'substitution weight matrix', see Appendix S3 for computation and simulations of the substitution weights).

The successive step in data analysis is using the HD dissimilarity matrix to calculate a dissimilarity tree. Here, we used Ward's method

	<b>K</b>	<i>A</i>	<i>P</i>	<i>R</i>	<b>E</b>	<i>O</i>	<i>L</i>	<i>O</i>
	<b>C</b>	<i>A</i>	<i>P</i>	<i>R</i>	<b>I</b>	<i>O</i>	<i>L</i>	<i>O</i>
<b>a</b>	0.4	0	0	0	1	0	0	1.4
<b>b</b>	1	0	0	0	1	0	0	2

**Fig. 3.** Two alternative computations of the Hamming distance, based on substitution weights given to each operation performed to equate the sequences (reported under each character). Identities are in italic and substitutions bold.

(Fig. 1b/c, HCW), the most common hierarchical agglomerative cluster procedure (Gabadinho *et al.* 2011). The resulting trees are a representation of the dissimilarity among the habitat use sequences of animals. The distance, or 'branch length', between leaves (which stay for individuals) and nodes (which stay for groups of individuals) can be used as a proxy for the dissimilarity between the portions of the tree: the longer the distance, the higher the dissimilarity.

The following step is the identification of 'clusters', or portions of the tree indicating association between sequences. This is the most important step to use trees for exploratory purposes and the subsequent interpretation of results. We determined the number of clusters present in a tree by a cut-off distance, based on a repeatable bootstrapping procedure (Fig. 1b/c,  $B_1$ ). For this application, we performed 1000 iterations and calculated the Jaccard bootstrapping index (i.e. bootmean; Hennig 2007) for a cut-off distance that separated from two to twenty clusters. This index compares the similarities of the original clusters to the most similar clusters in bootstrapped data sets. The cut-off distance was determined as the maximum number of clusters where the median bootmean is higher than 0.9, a value corresponding to highly stable clusters. This threshold is conservative in a sense that normally values above 0.75 correspond to robust classification.

#### DETERMINING CLASSIFICATION TREES OF SIMULATED TRAJECTORIES

Using this procedure, we computed a dissimilarity tree and identified clusters for each simulated landscape by using the 400 simulated sequences (100 replicates  $\times$  4 patterns of sequential habitat use, Fig. 1c). To account for spatial correlation of the four habitat classes, we recomputed the substitution matrix for each landscape. In fact, although all grid cells were assigned to habitat classes (open/closed) at random, this still corresponded to different spatial correlation structures across arenas, an effect of the different proportion of classes. Then, we investigated the sequence composition of individual clusters distinguished by the classification trees to assess the liability of the method to *correctly group sequences of different sequential behaviours*. This also allowed us to identify the name of each cluster.

#### DETERMINING SPATIO-TEMPORAL PATTERNS OF REAL TRAJECTORIES

In the next step, we introduced the real trajectories into one of the classification trees (Fig. 1c), according to the relative proportion of habitat availability as follows. For each real bimonthly sequence, we measured the availability of open/closed habitat within the corresponding bimonthly home range. We then associated them to the simulated trajectories referring to the arena with the same habitat proportion. For example, if an individual bimonthly home range showed an open habitat availability between 45 and 55%, the corresponding sequences were associated with the simulated sequences originated from the 50%

arena. Then, each tree was recalculated for the combined set of sequences (simulated and corresponding real sequences). Using this approach, simulation sequences could be used as a guide for classification of real sequences to their most similar sequential behavioural group (Fig. 1c,d).

#### EXPECTED SPATIO-TEMPORAL PATTERNS OF REAL TRAJECTORIES

Based on previous knowledge on roe deer ecology, and the individual descriptive variables, we formulated predictions of tree clustering. This was the core of our study, to assess the meaningfulness of SAM for exploration of spatio-temporal sequences of ecological data. In roe deer populations of northern and alpine environments, some individuals reach higher elevations in summer, when habitat suitability increases, and return at lower elevation in rigid winter conditions. Other individuals, instead, occupy the same low-elevation range all year round (Ramanzin, Sturaro & Zanon 2007; Cagnacci *et al.* 2011). This phenomenon is known as partial migration. On these premises, we expected individuals to classify into two main clusters according to the use of elevation: animals with a constant use of the same elevation range (winter and summer sequences) and animals with a seasonal shift in elevation range associated with migration (winter sequences separated from summer sequences) (P1a). Roe deer reproductive season is concentrated in summer, when both males and females exploit the best environmental conditions to meet the high energetic demand of mating and giving birth (Hewison, Vincent & Reby 1998). If migration is linked to habitat quality, we expect both sexes to show similar patterns of migration and thus of elevation range use (P1b, but see Ramanzin, Sturaro & Zanon 2007). Likewise, since fawns are not yet engaged in reproduction, we may expect a lower rate of migrating individuals (P1c). Alternatively, they may follow the mother in the migrating movements. Predictions on sequential use of open and closed habitats are less straightforward. Roe deer are known to prefer habitats providing cover and protection, especially intermediate stages of forest succession and ecotonal habitats (e.g. Tufto, Andersen & Linnell 1996). Therefore, we predict that animals would show a sequential use of habitats different from random (P2a). In particular, we predict a separation between animals using only closed habitat and those showing a combined use of closed and open habitats (P2b). When animals use both habitats, we expect open habitats to be used mainly during night, due to the anti-disturbance and anti-predatory behaviour of roe deer, translating in an alternating sequential use of open and closed (P2c) (Saïd & Servanty 2005). For similar reasons, a constant use of open habitat is instead less likely (P2d).

#### Classification trees of simulated trajectories

The application of HD algorithm to simulated bimonthly sequences generated 9 trees, each corresponding to landscapes with different habitat availabilities (i.e.  $T_{10}$ – $T_{90}$ ; see Fig. 1c and Appendix S4 for all trees). In all trees, the topological relationships between tree branches indicated two main orders of classification: first-order clusters, splitting the sequences into two groups according to the preferential use of different elevations ( $C_1$  and  $C_2$ ; e.g. Fig. 1c), and several second-order clusters separating animals with different sequential use of open and closed habitats (C, R, A, O, U; see text below, Fig. 1c and Appendix S4: second-order clusters are distinguished by coloured branches).

At first order, all trees showed a significant separation (Jaccard bootstrapping index  $B_J > 0.95$ ) between high- and low-elevation trajectories, thus defining a 'high-elevation' cluster ( $C_1$ ) and a 'low-elevation' cluster ( $C_2$ ). At second order, our defined bootstrap threshold (i.e.  $B_{J,median} > 0.9$ ) identified 6–8 main clusters ( $T_{10}$ – $T_{30}$  and  $T_{70}$ – $T_{90}$  = 6 clusters, with 40–80 sequences in each cluster;  $T_{40}$ – $T_{60}$  = 8 clusters, with about 40 sequences in each) and 3–7 small clusters (with less than 10 sequences each). Separation of the main second-order clusters was significant (most clusters  $B_J > 0.95$ ), whereas for smaller clusters, it was not (most clusters  $B_J < 0.75$ ). For both high and low elevations, the main second-order clusters corresponded to different patterns of sequential open/closed habitat use. The classification of sequences in such clusters was highly dependent on habitat availability in arenas. For trees deriving from arenas with similar open/closed habitat proportion (40–60%:  $T_{40}$ ,  $T_{50}$ ,  $T_{60}$ , Fig. 1c, Figs S4.4–S4.6 in Appendix S4), we distinguished all four simulated sequential behaviours: homogeneous closed (C; brown and dark blue branches), random (R; red and cyan branches), alternating (A; orange and blue branches) and homogeneous open (O; yellow and light blue branches). In this case, 99% of sequences in each cluster was of the same sequential behaviour. For example, all sequences in the blue cluster of Fig. 1c are homogeneous closed. Conversely, for landscapes where habitats were proportionally available (open < 30% or open > 70%:  $T_{10}$ ,  $T_{20}$ ,  $T_{30}$  and  $T_{70}$ ,  $T_{80}$ ,  $T_{90}$ , Fig. S4.1–S4.3, S4.7–S4.9 in Appendix S4),

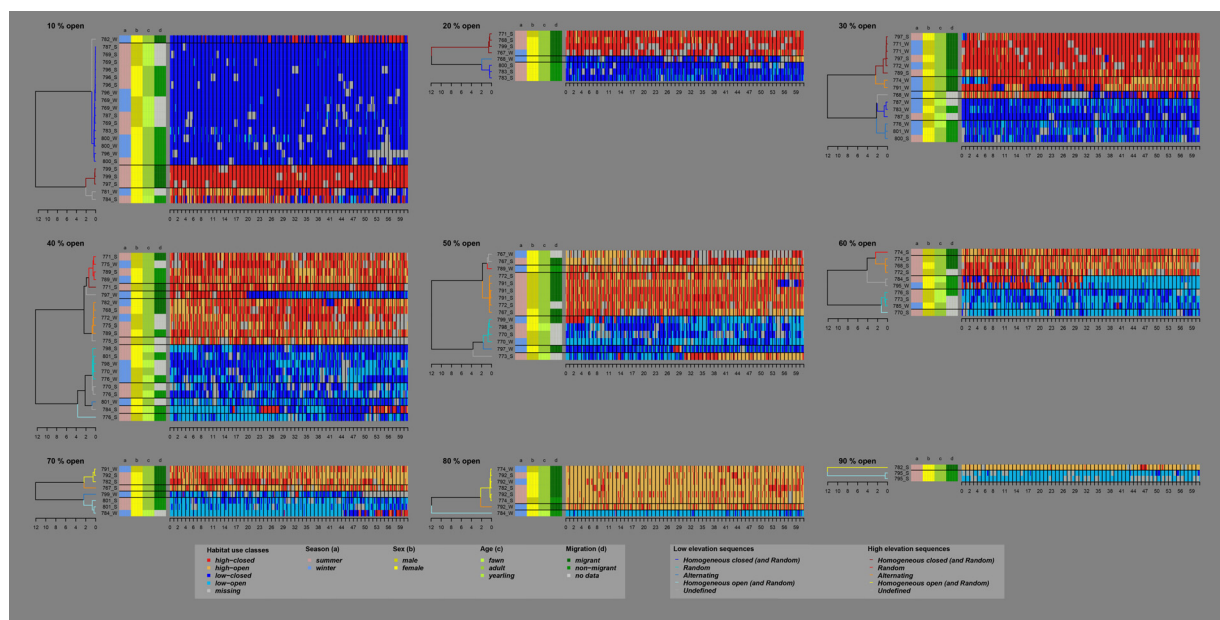
random sequences clustered with homogeneous sequences of the dominant habitat (C with R: brown and dark blue branches; O with R: light blue and yellow branches). The random sequential use of habitat according to availability (i.e. the control case) was therefore effectively represented by a separated random cluster only when sequences were not trivial (e.g.  $T_{10}$  and  $T_{90}$  obviously led to homogeneous sequences 'at random').

Finally, small clusters corresponded to trajectories indicating a mixed use of high and low elevations, for a specific pattern of sequential use of open and closed habitats (e.g. mixed-alternating). Alternatively, small clusters were undefined (U) due to a too small number of sequences (<5 sequences) and can be considered as outliers. Both mixed and undefined clusters were coloured grey in trees (Fig. 1c, Appendix S4).

## Spatio-temporal patterns of real trajectories and discussion of the study case

### SEQUENTIAL HABITAT USE PATTERNS

We used the trees based on mixed simulated and real trajectories to assess the classification of real trajectories according to sequential habitat use. To visualize how real trajectories were classified, we extracted (i.e. pruning) the real sequences from nine different trees (Figs 1d and 4, see also Appendix S5 for a different visualization).



**Fig. 4.** Roe deer bimonthly sequences of habitat use classes (right portion of the nine panels; daily scale reported bottom right), and resulting dissimilarity trees, based on the Hamming distance algorithm (left portion of the panels). The real sequences were extracted (pruned) from their corresponding simulated tree from which sequences were derived from a trajectory running over nine different arenas with varying habitat proportions (10–90%). Tree 'leaves' represent a real sequence, while 'nodes' indicate their clustering. Branch lengths (distance between leaves, and first common node: bottom left for the scale) indicate the dissimilarity between individual sequences. The split into two main clusters define first-order clusters separation, based on use of elevation classes (high,  $C_1$ ; low,  $C_2$ ). Second-order clusters are based on the use of open/closed habitats and distinguish homogeneous closed (C, brown and dark blue branches), random (R, red and cyan branches), alternating (A, orange and blue branches) and homogeneous open (O, yellow and light blue branches). Grey branches (U) are sequences with undefined classification or clusters with mixed sequences of high and low. The id gives the animal code and season (summer, S; winter, W). Variables season, age, sex and migration are represented as colour-coded bars between trees and sequences (see legend for meaning of the colours).



The first-order clusters  $C_1$  and  $C_2$  distinguish very well high- from low-elevation sequences (Fig. 4, Appendix S5 right vs. left panels).

More interesting is the classification of sequences in different sequential patterns of open/closed habitat use (second order). Within our roe deer population, we found evidence of the four different patterns of open/closed sequential use that have been addressed in our simulations. Specifically, 98 sequences out of 111 were classified according to those patterns (91.5%). Moreover, some interesting differences emerged in sequential use of open/closed habitats for different elevations (Fig. 4, Appendix S5). At high elevation, half of the sequences have an alternating ( $C_1A$ : 32%, 18 sequences) or homogeneous closed pattern ( $C_1C$ : 25%, 14 sequences), whereas only 11% of the sequences are random ( $C_1R$ : 6 sequences). At low elevation, 63% of the sequences are homogeneous closed ( $C_2C$ : 41%, 22 sequences) or random ( $C_2R$ : 22%, 12 sequences), and conversely alternating sequences are uncommon ( $C_2A$ : 13%, 7 sequences). Surprisingly, both at high and low elevations, there are also homogeneous open sequences ( $C_1O$ : 18%, 10 sequences;  $C_2O$ : 14%, 8 sequences). At the study area scale, different sequential patterns of open/closed habitat use according to elevation are possibly linked to those habitats availability in the study area (Fig. 2: open more available at high elevation). Indeed, individual habitat use sequences are obviously related to the availability of habitat classes within the home range (Fig. 4), in particular sequences classified as homoge-

neous sequences (C and O) derived from home ranges with a large proportion of one specific habitat type (Fig. 4.1–4.3, <30% open; Fig. 4.7–4.9, >70% open). Instead, heterogeneous sequences showed an alternating or random pattern (A and R) and derived from home ranges where both habitat types were available (Fig. 4.4–4.6, >40% open and <60% open). Interestingly, though, both homogeneous and heterogeneous trajectories lay next to each other in the study area (Fig. 5). Also, several sequences with the same home range availability were assigned to different sequential patterns. For example, 776\_W was classified in  $C_2A$  and 787\_S in  $C_2C$ , when they both have 30% open in home ranges (Fig. 4.3). Or, 771\_S was classified in  $C_1C$  and 789\_S in  $C_1R$ , when they both have 40% open in home ranges (Fig. 4.4).

Finally, sequences switching between high and low were classified in the mixed-alternating cluster (only one sequence) or in the undefined clusters (i.e. as an outlier) (Fig. 4,  $C_1U$ : 14%, 8 sequences;  $C_2U$ : 9%, 5 sequences). Notably,  $C_1U$  mainly consists of sequences changing from high to low elevation (e.g. Fig. 4.4, 797\_W; Fig. 4.6, 784\_S, 795\_W), whereas  $C_2U$  shows the opposite pattern (e.g. Fig. 4.5, 773\_S), as a result of the migration between winter and summer ranges (Fig. 5). Alternatively, undefined sequences may correspond to sequential behaviours we did not simulate, such as the shift from homogeneous closed to homogeneous open sequential habitat use within the bimonthly period (e.g. Fig. 4.5, 767\_W).

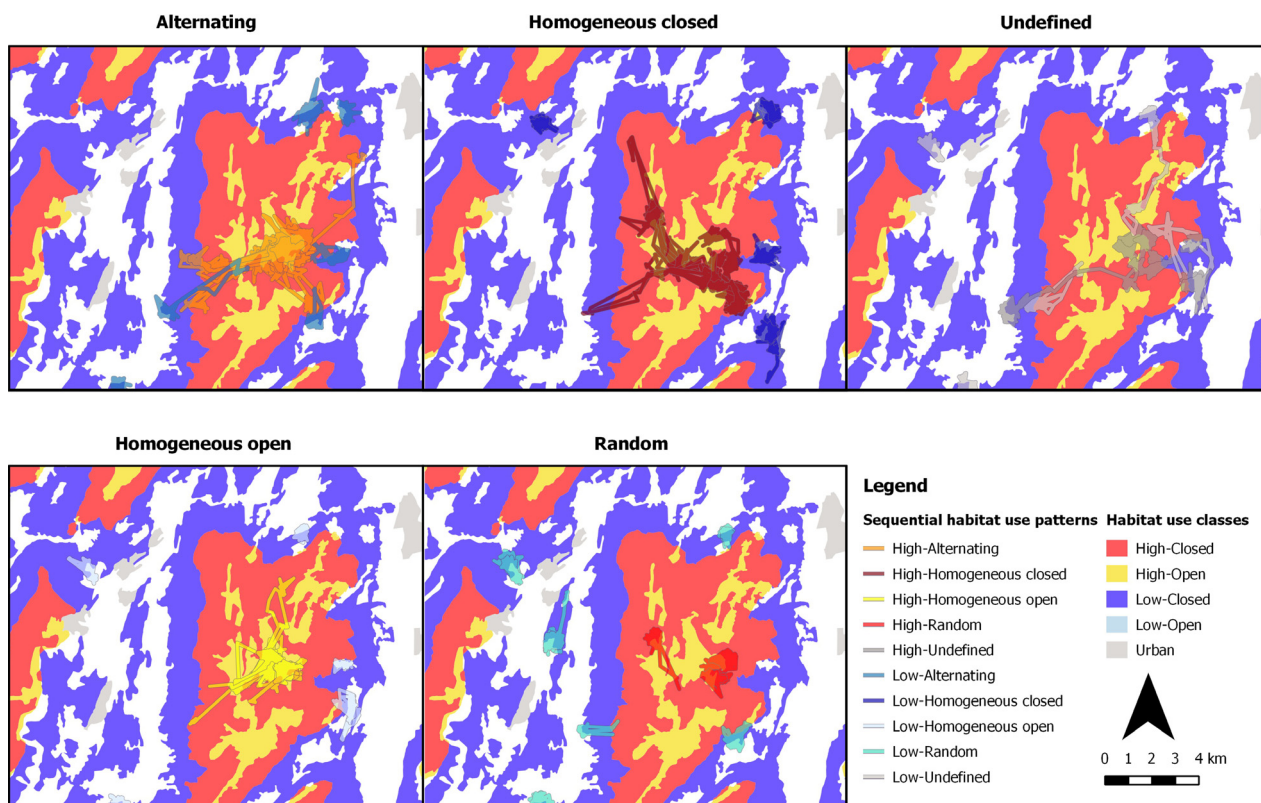


Fig. 5. Clusters of real bimonthly trajectories with similar patterns of sequential habitat use in their spatial context. Legend in bottom-right panel.

## SEQUENTIAL HABITAT USE VS. DESCRIPTIVE VARIABLES

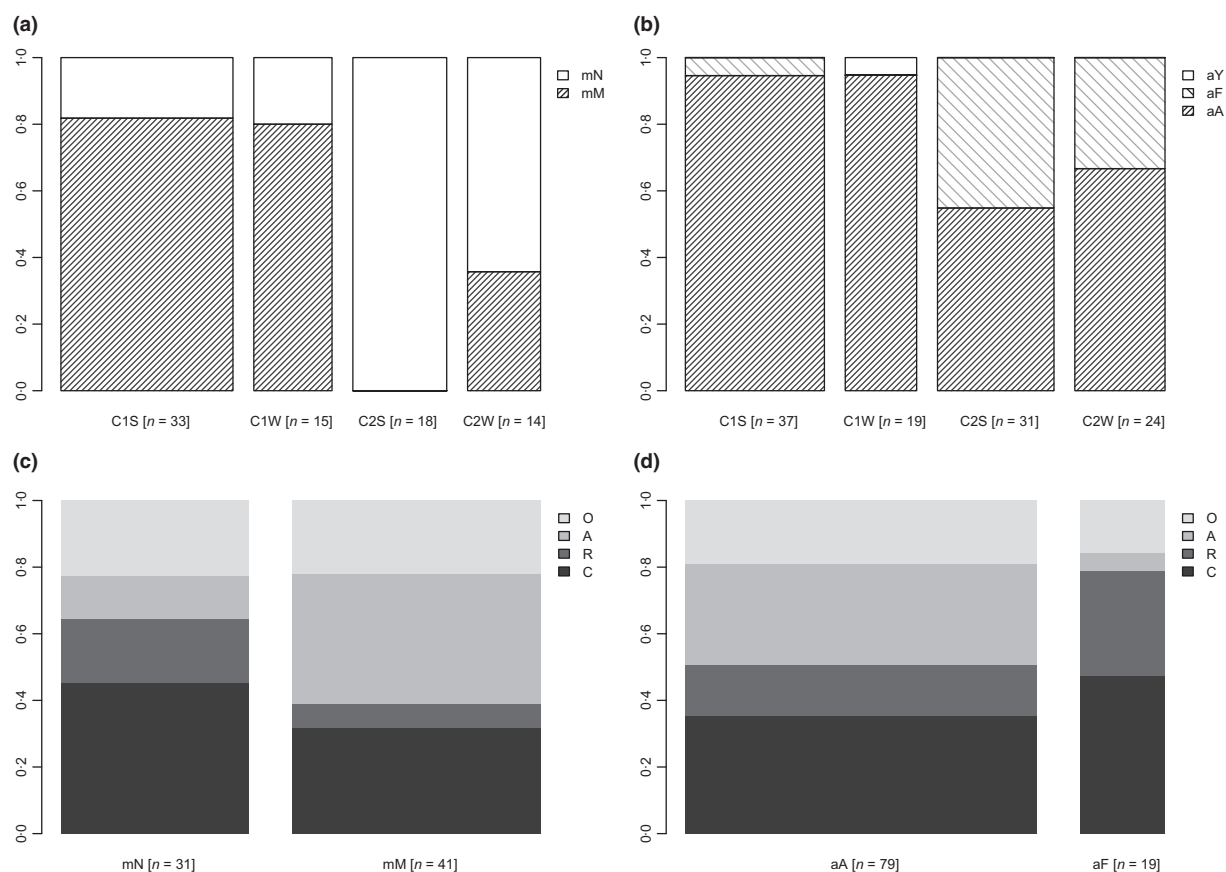
Descriptive variables of individual sequences, in particular migration occurrence, sex and age, enabled to further interpret the classification of trajectories in first- and second-order clusters. In winter, individual trajectories were classified both at high and at low elevation, regardless of the space-use strategy (i.e. both migrants and non-migrants; see Figs 4 and 6a). Conversely, in summer migrants' trajectories were always classified at high elevation and non-migrants' mainly at low elevation. Sequences that were classified as alternating mainly belonged to migrants, whereas homogeneous open sequences equally corresponded to migrant and non-migrant individuals (Figs 4 and 6c). Finally, more than 60% of the non-migrant sequences were either classified as homogeneous closed or random.

Interestingly, trajectories of fawns were classified at low elevation, both in summer and in winter, except one sequence (Figs 4 and 6b). In terms of second-order clusters, trajectories of fawns were classified in the great majority as homogeneous closed or random (Figs 4 and 6d). Adults' sequences were more equally distributed between different patterns of habitat use.

## ASSESSMENT OF PREDICTIONS AND CASE STUDY DISCUSSION

The first clear spatio-temporal pattern that emerged in the study of roe deer population through exploratory analysis with SAM is a differential use of altitude. This can be linked to *migration* for two pieces of evidence: first, some individuals with sequences classified at high elevation in some trees also showed sequences classified at low elevation in other trees (Fig. 4, e.g. individual 799). Secondly, some sequences from the same individuals included both high and low altitudes in the migration periods, when animals move between seasonal ranges (Fig. 4, e.g. individual 797, Fig. 5 – undefined). Moreover, the classification of sequences in high- and low-elevation clusters can be attributed to *partial migration* for a further evidence; that is, sequences of some other individuals were always included in the low-elevation cluster (Fig. 4, e.g. individual 783), or always across both (Fig. 4, e.g. individual 784); that is, they were resident. As such, prediction P1a is supported.

The property of SAM as a valid exploratory tool to identify spatio-temporal patterns of individual movements was also highlighted when looking at descriptive variables. Sequential use of high and low elevations was same among sexes (Fig. 4: sequences of both sexes were included in both clusters; P1b),



**Fig. 6.** Individual descriptive characteristics are summarized by the clusters to which real individual sequences were assigned. Migration (%; panel a; migrant-mM, non-migrant-mN) and age (%; panel b; adult-aA, yearling-aY, fawn-aF) in relation to first-order clusters (high, C<sub>1</sub>; low, C<sub>2</sub>), both in summer and in winter. Second-order clusters (%; C, R, A, O; see legend, Fig. 4 for meaning of codes) in relation to migration (panel c) and age (panel d).



but not across age classes (Figs 4 and 6; P1c). This is a novel result for roe deer that opens up new directions of research. For example, age class could be included as an explanatory variable to assess partial migratory behaviour of roe deer.

The most innovative outcome of our analysis was the spatio-temporal analysis of open/closed habitat use by individual roe deer. The most obvious, and yet relevant, result was that roe deer did not always use open/closed habitat at random (P2a). As predicted, sequences of individuals were clearly distinguished between homogeneous and heterogeneous use or shifted between the two categories. We intend to underline that by SAM exploratory analysis, we always looked at sequential *use* of habitat; that is, habitat selection was not considered here. The term 'random' use, therefore, does not refer to habitat use according to total habitat availability, as in classic habitat selection analysis (Manly *et al.* 2002), but to sequential use of available habitats as it would happen by chance. Indeed, we found that sequential use of habitats was heavily influenced by availability in the home range. For example, home ranges including a high percentage of closed habitat likely resulted in a 'random' homogeneous use of closed (Fig. 4.1–4.3, Fig. 5 – homogeneous closed). Conversely, home ranges including both open and closed habitats may lead to 'random' use of both (Figs 4.4–4.6 and 5 – random). Here, we have two observations. On the one side, the fact that trajectories with a different spatio-temporal pattern laid next one to each other in the study area (Fig. 5) indicates that sequential habitat use reflects the selection operated by animals to define their home ranges within the study area (i.e. second-order habitat selection, *sensu* Johnson 1980). On the other side, though, we had clear evidence of *alternative* tactics of temporal use of habitats available within the home range (several examples of sequences with same total habitat availability, but different sequential patterns, see above and Fig. 4). When both open and closed were available in home ranges, one pattern stuck out as particularly well represented, that is the alternating use of both habitats on a daily basis (P2c; Figs 4 and 5). To our knowledge, this is among the first systematic exploratory analyses of a behavioural pattern previously empirically described, and linked to individual personalities (Bonnot *et al.* 2015). According to our results, the propensity of roe deer to expose to open habitats can result in a systematic spatio-temporal pattern of habitat use, which is evident only in some individuals (Fig. 4). Even more, our explorations indicated that migrant individuals used this pattern much more than resident, at equal habitat availability in the home range (Figs 4.4–4.6 and 5 – alternating). Migration attitude in partial migratory populations has been previously attributed to a 'boldness' syndrome (Chapman *et al.* 2011). Investigating the effect of personalities at both seasonal and daily habitat use scales is a very exciting research direction that our exploratory results suggest.

A further result that supports the individual differences in sequential habitat use is the presence of homogenous open

sequences, an unexpected and yet observed pattern both at high and at low elevation (P2d; Fig. 4).

## Discussion: Application of SAM to animal habitat use

In this paper, we showed that SAM is a useful and powerful tool to explore and compare sequences of habitat use by animals, and extract common spatio-temporal patterns. Importantly, we took into consideration an aspect of animal ecology largely overlooked in the literature. Indeed, one of the most interesting and captivating outcomes of SAM is that different topological levels of trees are associated with hierarchical similarities between the individuals' sequential use of environmental features. Repetitive patterns of sequential habitat use are informative with respect to the interaction between animals and their environment. In fact, well-known phenomena, such as migration, or feeding-resting cycles emerge as repetitive patterns of habitat use at different temporal scales, as shown by our study case.

To understand the potential usefulness of SAM for movement ecology, we shall first clarify what SAM *is not*. SAM *is not* a spatial explicit method and *does not* provide predictive models of habitat use, nor of animal distribution. Instead, SAM embeds the temporal component of habitat use, in the form of real ordered sequences of used habitat classes. SAM can provide information at the *population and individual level*, by clustering individuals in trees through robust algorithms that search for dissimilarities in spatio-temporal patterns of habitat use. Based on all these considerations, we see SAM as an approach to *explore* temporal patterns in habitat use across an animal population. As such, it differentiates from and it complements current exploratory methods in habitat analysis. The R-package AdehabitatHS provides a suite of niche-based methods that are primarily used to express the realized niche (Calenge & Basille 2008). However, niche-based methods are also meaningful to identify the most relevant explanatory variables, and their relationships, for predictive habitat selection models, such as RSF (Boyce *et al.* 2002). RSF model the probability of disproportional habitat use in a hypothetical-deductive framework, by means of selection of competing models. A preliminary investigation of most relevant variables thus allows to better express hypotheses, and focuses the analysis. SAM may complement the aforementioned approaches by also exploring the temporal component of habitat use patterns. A big advantage of niche-based methods or recursive partitioning trees is their ability to provide a graphical representation of the importance and relation between variables. Similarly, SAM summarizes common patterns of sequential habitat use across the population, by clustering. Importantly, we suggest that the way to handle the length of the sequences (i.e. time resolution and total duration) should depend on the research questions. In the study case, we decided to split the individuals' sequences in bimonthly periods, since we were interested in intra-annual patterns. This design implies that each individual is present more than once in the trees, and caution must be

taken when interpreting the results (i.e. pseudo-replication Tukey 1977).

In our exercise, we considered a simple combination of habitat classes, although the R-package TraMineR allows to define more complex combinations of environmental parameters. All the same, sequences based on many habitat classes, would lead to a very articulated dissimilarity tree, thus likely difficult to interpret. We therefore suggest a rationale in the combined and complementary use of the aforementioned methods, when exploring spatio-temporal patterns in animal movement data. First, niche-based or recursive partitioning methods or other simpler multivariate approaches (e.g. PCA) can be applied to identify the most important and least related environmental parameters. Then, ordered sequences of locations can be matched to those variables by spatial join with relevant geographical layers, and sequences of habitat use thus obtained. Sequence analysis can then be performed, and provide a representation of existing spatio-temporal patterns, or be used to formulate new hypotheses evaluated through a classic model selection approach.

Technically, we suggest to simulate expected sequential behaviours (e.g. based on preliminary classifications or previous knowledge, Fig. 1a), and use them as a guide to extract the sequential habitat use pattern of real sequences, while accounting for spatial correlation (through the substitution matrix) and habitat availability. As our study case shows, real-world sequences can exhibit more complex behaviours than those represented by simulations. We suggest that if the great majority of real trajectories are classified in cluster types derived from simulations, then the real sequential habitat use is well represented. Otherwise, one may want to change the simulation rules, which can be easily modified in the movement model.

Sequence analysis methods is a well-suited method for data acquired by animal-borne tracking technologies, since the method can account for two main limitations related to animal movement data: missing locations (Frair *et al.* 2010) and spatial correlation (Dray, Royer-Carenzi & Calenge 2010). We explicitly remarked for the first time that the weight matrix used by SAM can be used to deal with spatial correlation. Specifically, we offer a quantifiable and repeatable assessment of the spatial correlation between different habitat classes (i.e. substitution matrix). Notwithstanding the novelty, future studies may assess the sensitivity of SAM output to changes in the substitution matrix.

While the framework presented in this paper is relevant and innovative in ecological studies, in practice there are also some limitations to SAM. First, temporal correlation is not directly accounted, whereas there is a clear temporal association between consecutive observations. To account for temporal correlation, a variant of the Hamming distance known as the fuzzy Hamming distance (Bookstein, Klein & Raita 2001) could be used, but is only developed for binary data. Secondly, several researchers in evolutionary biology (e.g. Morrison 2010) and social sciences (e.g. Wu 2000) have argued that many steps in SAM are based on subjective decisions (e.g. definition of classes, parameter settings, interpretation of results).

These limitations notwithstanding, we believe that SAM offers great advantages and new insights into movement ecology studies. For instance, the method can be promptly extended to multiple or other species than roe deer as well as to other spatial (e.g. home range) and temporal resolutions (e.g. hours) and reveal yet underappreciated or overlooked patterns.

## Acknowledgements

We are grateful to MOVE (COST Action IC0903) to provide funding for two Short-Term Scientific Missions (STSMs) to conduct this research. Also, we thank Fondazione Edmund Mach and Special Research Fund (BOF) of Ghent University for their financial support to JDG. Finally, we wish to thank the collaborative EURODEER project to provide geospatial and environmental data. We are particularly grateful to Ferdinando Urbano for his support with data management, to Matteo Marcantonio for precious suggestions in R programming and to Matthias Studer for advice on TraMineR.

## Data accessibility

We provide the R script (Appendix S6) to repeat the sequence analysis as here presented. The full data set is made available in the Dryad digital repository (De Groeve *et al.* 2015).

## References

- Abbott, A. (1995) Sequence analysis: new methods for old ideas. *Annual Review of Sociology*, **21**, 93–113.
- Andrienko, G., Andrienko, N. & Heurich, M. (2011) An event-based conceptual model for context-aware movement analysis. *International Journal of Geographical Information Science*, **25**, 1347–1370.
- Benhamou, S. & Riote-Lambert, L. (2012) Beyond the utilization distribution: identifying home range areas that are intensively exploited or repeatedly visited. *Ecological Modelling*, **227**, 112–116.
- Bonnot, N., Verheyden, H., Blanchard, P., Cote, J., Debeffe, L., Cargnelutti, B., Klein, F., Hewison, A.J.M. & Morellet, N. (2015) Interindividual variability in habitat use: evidence for a risk management syndrome in roe deer? *Behavioral Ecology*, **26**, 105–114.
- Bookstein, A., Klein, S.T. & Raita, T. (2001) Fuzzy Hamming distance: a new dissimilarity measure. *Combinatorial Pattern Matching*, **2089**, 86–97.
- Boyce, M.S., Vernier, P.R., Nielsen, S.E. & Schmiegelow, F.K.A. (2002) Evaluating resource selection functions. *Ecological Modelling*, **157**, 281–300.
- Buchin, M., Dodge, S. & Speckmann, B. (2012) Context-aware similarity of trajectories. *Proceedings of the 6th International Conference on Geographic Information Science (GIScience)*. Springer, Berlin Heidelberg.
- Cagnacci, F., Boitani, L., Powell, R.A. & 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.M., Morellet, N. *et al.* (2011) Partial migration in roe deer: migratory and resident tactics are end points of a behavioural gradient determined by ecological factors. *Oikos*, **120**, 1790–1802.
- Calenge, C. (2006) The package adehabitat for the R software: a tool for the analysis of space and habitat use by animals. *Ecological Modelling*, **197**, 516–519.
- Calenge, C. & Basille, M. (2008) A general framework for the statistical exploration of the ecological niche. *Journal of Theoretical Biology*, **252**, 674–685.
- Chapman, B.B., Hulthén, K., Blomqvist, D.R., Hansson, L.-Å., Nilsson, J.-Å., Brodersen, J., Nilsson, P.A., Skov, C. & Brönmark, C. (2011) To boldly go: individual differences in boldness influence migratory tendency. *Ecology Letters*, **14**, 871–876.
- De Groeve, J., Van de Weghe, N., Ranc, N., Ometto, L., Rota-Stabelli, O. & Cagnacci, F. (2015) Data from: Extracting spatio-temporal patterns in animal trajectories: an ecological application of sequence analysis methods (SAM). *Dryad Digital Repository*. doi:10.5061/dryad.h4f7p.
- Delafontaine, M., Versichele, M., Neutens, T. & Van de Weghe, N. (2012) Analysing spatio-temporal sequences in Bluetooth tracking data. *Applied Geography*, **34**, 659–668.

- Dray, S., Royer-Carenzi, M. & Calenge, C. (2010) The exploratory analysis of autocorrelation in animal-movement studies. *Ecological Research*, **4**, 34–41.
- European Environment Agency, EEA (2010) CORINE Land Cover 2006. URL <http://www.eea.europa.eu/dataand-maps/data/corine-land-cover-2000-2006> (accessed 20 August 2015).
- Fortin, D., Hawthorne, L.B., Boyce, M.S., Smith, D.W., Duchesne, T. & Mao, J.S. (2005) Wolves influence elk movements: behaviour shapes a trophic cascade in Yellowstone national park. *Ecology*, **86**, 1320–1330.
- Frair, J.L., Fieberg, J., Hebblewhite, M., Cagnacci, F., DeCesare, N.J. & Pedrotti, L. (2010) Resolving issues of imprecise and habitat-biased locations in ecological analyses using GPS telemetry data. *Philosophical Transactions of the Royal Society B, Biological Sciences*, **365**, 2187–2200.
- Gabardin, A., Ritschard, G., Müller, N.S. & Studer, M. (2011) Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*, **40**, 1–37.
- Hennig, C. (2007) Cluster-wise assessment of cluster stability. *Computational Statistics and Data Analysis*, **52**, 258–271.
- Hewison, A.J.M., Vincent, J.P. & Reby, D. (1998) Social organisation of European roe deer. *The European roe Deer: The Biology of Success* (eds R. Andersen, P. Duncan & J.D.C. Linnell), pp. 189–219. Scandinavian University Press, Oslo.
- Johnson, D.H. (1980) The comparison of usage and availability measurements for evaluating resource preference. *Ecology*, **61**, 65–71.
- Manly, B., McDonald, L., Thomas, D., MacDonald, T. & Erickson, W. (2002) *Resource Selection by Animals. Statistical Design and Analysis for Field Studies*. Kluwer Academic Publisher, London.
- Ministry of Economy, Trade, and Industry of Japan, METI, National Aeronautics and Space Administration of the United States, NASA (2012) Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model Version 1 (GDEM V1). URL <http://asterweb.jpl.nasa.gov/gdem.asp> (accessed 20 August 2015).
- Morrison, D.A. (2010) Sequence alignment: methods, models, concepts, and strategies. *Systematic Biology*, **59**, 363–365.
- Ramanzin, M., Sturaro, E. & Zanon, D. (2007) Seasonal migration and home range of roe deer (*Capreolus capreolus*) in the Italian eastern Alps. *Canadian Journal of Zoology*, **85**, 280–289.
- Rodgers, A.R. & Carr, A.P. (1998) *HRE: The Home Range Extension for Arc-View*. Ontario Ministry of Natural Resources, Centre for Northern Forest Ecosystem Research, Thunder Bay, Ontario, Canada. URL <http://flash.lakeheadu.ca/~arodgers/hre/> (accessed 4 March 2014).
- Saïd, S. & Servanty, S. (2005) The influence of landscape structure on female roe deer home-range size. *Landscape Ecology*, **20**, 1003–1012.
- Shoval, N. & Isaacson, M. (2007) Sequence alignment as a method for human activity analysis in space and time. *Annals of the Association of American Geographers*, **97**, 282–297.
- Therneau, T., Atkinson, B. & Ripley, B. (2014) Recursive Partitioning and Regression Trees. URL <http://cran.r-project.org/web/packages/rpart/rpart.pdf> (accessed 10 October 2014).
- Tufto, J., Andersen, R. & Linnell, J.D.C. (1996) Habitat use and ecological correlates of home range size in a small cervid: the roe deer. *Journal of Animal Ecology*, **65**, 715–724.
- Tukey, J.W. (1977) *Exploratory Data Analysis*. Addison-Wesley, Reading, Massachusetts.
- Urbano, F., Cagnacci, F., Calenge, C., Dettli, H., Cameron, A. & Neteler, M. (2010) Wildlife tracking data management: a new vision. *Philosophical transactions of the Royal Society of London Series B, Biological Sciences*, **365**, 2177–2185.
- Wilson, C. (2008) Activity patterns in space and time: calculation representative Hagerstrand ClustalG software. *Transportation*, **35**, 485–499.
- Wittemyer, G., Polansky, L., Douglas-Hamilton, I. & 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 of the United States of America*, **105**, 19108–19113.
- Worton, B.J. (1995) Using Monte-Carlo simulation to evaluate kernel-based home-range estimators. *Journal of Wildlife Management*, **59**, 794–800.
- Wu, L.L. (2000) Some comments on sequence analysis and optimal matching methods in sociology: review and prospect. *Sociological Methods and Research*, **29**, 41–64.

Received 2 May 2015; accepted 23 July 2015  
Handling Editor: Stéphane Dray

## Supporting Information

Additional Supporting Information may be found in the online version of this article.

### Appendix S1. Simulated arenas.

**Figure S1.1.** Nine square gridded arenas, each composed of 10 000 squared cells of 100 × 100 m, with varying proportions of open (light colors) and closed (dark colors) habitats (10–90%) at high (blue) and low elevation (red/orange).

### Appendix S2. The movement model.

### Appendix S3. Substitution matrix to account for spatial correlation.

**Figure S3.1.** Simulated dispersed (a), random (b) and clustered (c) spatial distribution of habitat patches based on the shape of the study area.

**Figure S3.2.** Spatial correlation (Z-scores) in relation to habitat availability (x-axis, % open habitat) for HCxHC (red), HOxHO (yellow), LCxLC (blue), LOxLO (light blue).

### Table S3.1 Join-Count Statistics for simulated scenarios (a, b, c).

**Table S3.2** Substitution matrices for simulated scenarios (a, b, c), expressed as dissimilarity measured using Join-Counts, i.e. the proportion of the total number of adjacent polygons where the specified habitat neighbour relationship is not observed.

**Table S3.3** Join-Count Statistics for 9 arenas with variable proportions of open and closed habitats (expressed as the percentage of open habitat). See table S3.1 for explanation of codes.

### Table S3.4 Substitution matrices for the nine arenas.

### Appendix S4. Trees.

**Figure S4.1–4.9.** Simulated and roe deer bimonthly sequences of habitat use classes (right portion of the panel; daily scale reported bottom-right), and resulting dissimilarity tree, based on the Hamming distance algorithm (left portion of the panel).

### Appendix S5. Clusters of real bimonthly sequences.

**Figure S5.1** Clusters of real bimonthly sequences with similar patterns of sequential habitat use, extracted from the corresponding dissimilarity trees.

### Appendix S6. R-script of the complete analysis.