1   **Individual Movement - Sequence Analysis Method (IM-SAM): Characterising**

2   **Spatio-Temporal Patterns of Animal Habitat Use across Landscapes**

3   De Groeve Johannes* [1,2], Cagnacci Francesca [1,3], Ranc Nathan [3,1], Bonnot

4   Nadège [4], Gehr Benedikt [5], Heurich Marco [6,7], Hewison Mark [8], Kroeschel

5   Max [7,9], Linnell John D.C.[10], Morellet Nicolas [8], Mysterud Atle [11], Sandfort

6   Robin [12], Van de Weghe Nico [1]

7   *1. Department of Geography, Ghent University, Krijgslaan 281, 9000 Ghent, Belgium*

8   *2. Department of Biodiversity and Molecular Ecology, Research and Innovation Centre,*

9   *Fondazione Edmund Mach, Via E. Mach 1, 38010 San Michele all' Adige, Trentino,*

10   *Italy*

11   *3.Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford*

12   *Street, Cambridge, MA 02138, USA*

13   *4.Office National de la Chasse et de la Faune Sauvage, Direction de la Recherche et de*

14   *l'Expertise, Unité ongulés sauvages, 5 allée de Bethléem, 38610 Gières, France*

15   *5.Department of Evolutionary Biology and Environmental Studies, University of Zurich,*

16   *Winterthurerstrasse 190, 8057 Zurich, Switzerland*

17   *6.Bavarian Forest National Park, Department of Conservation and Research,*

18   *Freyunger Str. 2, 94481 Grafenau Germany;*

19   *7.Chair of Wildlife Ecology and Wildlife Management, Faculty of Environment and*

20   *Natural Resources, University of Freiburg, Tennenbacher Straße 4 D-79106 Freiburg,*

21   *Germany*

22   *8.CEFS, Université de Toulouse, INRA, Castanet-Tolosan, France*

23   *9.Division of Wildlife Ecology, Forest Research Institute of Baden-Württemberg,*

24   *Wonnhaldestraße 4, 79100 Freiburg, Germany*

25   *10.Norwegian Institute for Nature Research, P.O. Box 5685 Torgard, NO-7485*

26   *Trondheim, Norway*

27   *11.Centre for Ecological and Evolutionary Synthesis, Department of Biosciences,*

28   *University of Oslo, P.O. Box 1066 Blindern, NO-0316 Oslo, Norway*

29   *12.Institute of Wildlife Biology and Game Management, Gregor-Mendel-Straße 33,*

30   *1180 Wien*

31

32

33   *Corresponding Author: Johannes De Groeve, degroevejohannes@gmail.com

34

Individual Movement - Sequence Analysis Method (IM-SAM): Characterising 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 characterising 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.

65 Similarly, by alternating between access to food and cover resources over the day,

66 several wild species have adapted to agro-ecosystems (Aulak and Babinska-Werka

67 1990; Cibien *et al.* 1989; Hewison *et al.* 2001, Podgórski *et al.* 2013) or even urbanised

68 areas (i.e., the phenomenon of 'urban wildlife'; Magle *et al.* 2012). Most animal

69 movement methods have predominantly focused on how to analyse the spatial

70 component so far, while the temporal dependence of habitat use is often analysed less

71 elegantly by pooling samples into classes, such as night/day and active/inactive,

72 typically ignoring the sequential nature of habitat selection. Consequently, we need

73 robust methodological approaches to understand the sequential temporal patterns in the

74 use of complementary habitats in order to take appropriate conservation and

75 management actions.

76   Habitats provide the resources (e.g. food, cover, thermal protection) that species

77 need for survival and reproduction (Manly *et al.* 2002). European-level mapping

78 products (e.g. Corine Landcover, Copernicus; see eea.europa.eu) are often used by

79 movement ecologists to quantify such resources (e.g. forest cover) by linking these

80 maps with GPS locations obtained from animal tracking projects. Improved spatio-

81 temporal resolution and range of both remote sensing products and animal tracking

82 datasets are allowing ecologists to derive ever more detailed animal trajectories

83 annotated with habitat information, and, hence, facilitate the study of the animal-habitat

84 relationship over time (Cagnacci *et al.* 2010, Demšar *et al.* 2015, Kays *et al.* 2015).

85   In geo-informatics, such habitat information is referred to as *biological and*

86 *environmental context* and the integrated analysis of movement trajectories in relation to

87 such contexts are termed *context-aware movement analysis* (Andrienko *et al.* 2011;

88 Dodge *et al.* 2013; Demšar *et al.* 2015). Geo-informaticians and ecologists have

89 together developed several context-aware methods to visualise and analyse movement

90   in relation to habitat type (see Demšar *et al.* 2015 for overview). The concept was first

91   proposed and applied to animal movement data by Andrienko *et al.* (2011). In a case

92   study using roe deer GPS movement data, the latter presented several methods to

93   visualise aggregated hourly use of open habitats for spatial clusters of locations. Xavier

94   and Dodge (2014) developed DYNAMO (Dynamic Multivariate Visualization of

95   Movement), a tool for animating trajectories annotated by habitat variables. Demšar *et*

96   *al.* (2015b) proposed a 3D-visualisation of a home range, where the x,y-plane is space,

97   and the z-axis is time, and aggregated information of used habitats could characterise

98   the space-time cube. Toor *et al.* (2016) developed a trajectory segmentation algorithm

99   based on temporal changes in habitat use using random forest models. All these context-

100 aware approaches, one way or another, investigate how the contextual information is

101 used through time.

102   One way of considering time dependency is to investigate sequentiality, which

103 takes into account the temporal order in which behavioral, environmental or movement

104 states occur. In the field of movement ecology, and especially for the study of recursions

105 (i.e., revisitations of the same places), several promising methods have been proposed,

106 often relying on approaches developed in different research fields or by combining

107 several concepts (Berger-Tal and Bar-David 2015). Fourier and wavelet transforms have

108 been used to simultaneously detect recursions at multiple temporal scales (Wittemyer *et*

109 *al.* 2008, Bar-David *et al.* 2009, Polansky *et al.* 2010, Riotte-Lambert *et al.* 2013, 2017).

110 Minimal conditional entropy was used to identify the temporal scale of repetitiveness in

111 resource patch visitation and to quantify the degree of predictability in movement

112 sequences (i.e, traplining, Riotte-Lambert *et al.* 2017). The latter study also presented an

113 algorithm based on time-lag matrices (Goto 2003) to mine for repeated movement sub-

114 sequences. Utilisation distribution in combination with residence time was also

115   proposed to identify areas of high recursion rate from movement data (Benhamou and

116   Riotte-Lambert 2012). Recently, model-based continuous-time movement metrics were

117   suggested to find recursion signatures (Péron *et al.* 2017), as well as periodograms

118   (Péron *et al.* 2016).

119   The Sequence Analysis Methods (SAM) approach was recently developed to

120   measure similarity between temporally ordered sequences of habitat use within

121   individuals or populations (De Groeve *et al.* 2016). SAM is a tree-based approach

122   developed in computer science to measure dissimilarity between multiple strings of

123   characters (Wagner and Fischer 1974) and has subsequently been used in different fields

124   of study. It was first adopted to measure dissimilarity between DNA and protein

125   sequences (Needleman and Wunsch 1970). The popularity of the technique in molecular

126   biology resulted in several applications in other fields such as in sociology to study life

127   courses (e.g. Abbott 1995, Wilson 2006, Gabadinho *et al.* 2011), in time geography for

128   transportation science (Wilson 2008), in tourism research (Shoval and Isaacson 2007),

129   in indoor navigation (Delafontaine *et al.* 2012), in choreography research (Chavoshi *et*

130   *al.* 2015), in human mobility (Brum-Bastos *et al.* 2018); and, recently, in the field of

131   animal movement ecology (De Groeve *et al.* 2016).

132   De Groeve *et al.* (2016) showed that, for a given proportion of habitat used,

133   animals can show very different sequential space use patterns. For example, while

134   animals may equally use open and closed habitats over a given time-window, their

135   sequential use patterns were markedly different (from random to day-night alternating

136   patterns). We describe here a methodological framework building on De Groeve *et al.*

137   (2016), where we analyse sequential patterns of habitat use of animals, while accounting

138   for individual-level variability in landscape heterogeneity, or IM-SAM (Individual

139   Movement - Sequence Analysis Method). In essence, the baseline SAM-framework uses

140   simulated sequential habitat use patterns to determine the classification of real habitat

141   use sequences, where simulated sequential habitat use patterns were generated based on

142   a biased and correlated random walk movement model. De Groeve *et al.* (2016)

143   generated the patterns of sequential habitat use in *artificially generated landscapes* that

144   *mimic* habitat composition and structure of an animal's movement context. Instead, in

145   IM-SAM we generated these sequential patterns in the *real landscape context for an*

146   *individual animal*, or individual reference area of use, and hence accounted for the *true*

147   variation in habitat composition and structure among individuals.

148   In this analysis we use IM-SAM to expand sequential habitat use analysis from

149   animal trajectories derived from a single local context (i.e. single-population) to a

150   continental scale (i.e. multiple-populations at the European level), specifically using

151   GPS trajectories of 404 individual European roe deer (*Capreolus capreolus*) from nine

152   populations with contrasting landscape structure across Europe. As roe deer are

153   generally described as a forest species, but often feed on rich resources available in

154   more open habitats (e.g. meadow, crop), we described sequential use of two simple

155   habitat classes, open and closed, converting regularised animal trajectories into multiple

156   character sequences, where each character in the sequence corresponds to the habitat

157   used at a given timestamp.

Material and methods – habitat use sequential analysis

159   To describe sequential use of open and closed habitats for individual animals, we

160   followed a workflow modified from De Groeve *et al.* (2016) that can be summarised in

161   four steps (Fig. 1). First, we produced an exploratory tree for *each individual roe deer*

162   based on biweekly trajectories annotated with habitat categories and used to formulate

163   hypotheses of expected patterns of sequential habitat use (Fig. 1a). The number of

164   sequences *per individual* exploratory tree depends on the monitoring period of the

165   individual. Next, we generated stochastic movement rules for such expected patterns of

166   sequential habitat use and ran the movement models within each individual's reference

167   area of use (here computed as the 100% MCP) in order to produce individual specific

168   simulated trajectories (Fig. 1b), and individual-level simulation trees (Fig. 1c).

169   Simulation trees were therefore based on the true landscape context where each

170   individual actually ranged. Finally, we combined real and simulated trajectories to

171   produce trees where real and simulated habitat use sequences with a high degree of

172   similarity were grouped together. After computing the proportions of simulated patterns

173   in each cluster, we could define sequential habitat use cluster types and assign these

174   identified tags to the real sequences included in that cluster (Fig. 1c/d). Finally, we

175   pruned the output tree and visualised only the classified real trajectories, to facilitate

176   interpretation (Fig. 1d). We now describe each step in detail. [Figure 1 near here]

177   ***Real trajectories – exploratory trees (Fig. 1a)***

178   We extracted roe deer trajectories from the EuroDEER database (Cagnacci *et al.* 2011,

179   euroungulates.org) and subsampled them into 16-day GPS trajectories with a fixed four-

180   hour relocation interval (0, 4, 8, 12, 16 and 20h) over a fixed yearly schedule starting on

181   January 1st (e.g. 01/01-16/01, 17/01-01/02, etc.). To increase the sample size, we also

182   included relocations within one hour from the above four-hour intervals, after a

183   sensitivity analysis verified that this did not affect exploratory tree clustering (see

184   Appendix S2). If multiple GPS-positions were within this time window, we selected the

185   closest one in time to the reference time stamp. Otherwise, if no fix was obtained for a

186   reference time stamp, we annotated the gap with a missing value (NA). The number of

187   sequences per individual ranged between 2 and 52 and depended on the monitoring time

188   and completeness of the individual trajectory. We annotated each 16-day trajectory with

189   the percentage tree cover (0-100%) extracted from the High-Resolution Layer-Tree

190   Cover Density 2012 (TCD, EEA 2012, 20m spatial resolution), thus obtaining the

191   *biweekly sequences* of habitat use. We reclassified TCD into two distinct classes, using a

192   cut-off point for pixel-level tree cover density of 50%, distinguishing closed (C, $\geq$50%)

193   and open (O, <50%) habitats. The final dataset consisted of 5402 biweekly habitat use

194   sequences of 404 animals (236 females and 168 males) from nine European roe deer

195   populations characterised by varying forest composition (see map in Appendix S1:

196   Southcentral Norway, NK5; Southwest France, FR8; Switzerland, CH25; Southern

197   Germany, DE15, DE31; Southeast Germany, DE1; Northern Italy, IT1, IT24; Eastern

198   Austria, AU17). After processing, the dataset consisted of 14,607 missing values

199   (2.82%) and 503,985 true GPS locations (97.18%), of which 273,230 (52.69%) were

200   classified as open and 230,755 (44.50%) as closed habitat. See Appendix S2 for the

201   complete data preparation procedure.

202   We generated an initial visualisation of the habitat use patterns by creating

203   exploratory trees separately for each individual (Fig. 1a). These trees were used to

204   describe sequential patterns and helped to build hypotheses for expected models of

205   sequential habitat use (see below). Sequence Analysis Methods use a dissimilarity

206   algorithm to compute the distance between all possible pairs of sequences. All these

207   pairwise distances are written into a dissimilarity matrix. Here, we computed the

208   pairwise distances using the Hamming dissimilarity algorithm (HD), which calculates

209   the minimum number of character substitutions (i.e., O and C) required to match a

210   number of sequences of equal length (Gabadinho *et al.* 2011). From the HD

211   dissimilarity matrix, we subsequently calculated dissimilarity trees using a hierarchical

212   clustering algorithm (Ward's method, Gabadinho *et al.* 2011, De Groeve *et al.* 2016).

213   The above described analytical steps were performed using the R-package TraMineR

214   (Gabadinho *et al.* 2011).

215   *From SAM to IM-SAM*

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

217   We computed individual reference areas of use as 100% Minimum Convex Polygons

218   (MCP) for each roe deer separately as a simple representation of the available space in

219   which movement of that individual could occur, including occasional excursions outside

220   the usual range. We then intersected each MCP with the TCD raster re-classified as

221   open/closed habitat as described above, and characterised all MCPs by their habitat

222   proportions. For illustrative purposes (i.e. Fig. 2), we also classified MCPs into 5

223   classes (0.1 to 0.5) of 'relative proportion of open/closed habitat'. So, for example, the

224   0.1 class indicates both 10% open - 90% closed, and 10% closed - 90% open habitat.

225   Within each MCP we simulated sequential habitat use patterns using a simple

226   spatially explicit stochastic movement model to express expected sequential habitat use

227   patterns that were determined by a memory-based movement model with parametrized

228   selection coefficients for open and closed habitats (see Appendix S3 for the full

229   description). In particular, according to the exploratory trees obtained from real

230   sequences with six locations per day (Fig. 1a), we identified six characteristic patterns

231   of sequential habitat use (Fig. 1b): homogeneous use of closed habitat, or pattern 'C', the

232   resulting day-night sequence (DS) being: DS=CCCCCC; homogeneous use of open

233   habitats, or pattern 'O' (DS=OOOOOO); and three patterns of day-night alternating use

234   of both open and closed habitats, or patterns 'A'. The alternating patterns were generated

235   on the assumption that roe deer use open and closed habitats in relation to the daylight

236   cycle, with use of open habitat more prevalent at night (Mysterud *et al.* 1999; Bonnot *et*

237   *al.* 2013). In addition, we accounted for variation in day length over the different sites

238   and seasons included in the study. In particular, to account for spatial and temporal

239   variation in day length, we distinguished the following three patterns: dominant use of

240 open habitat from 16:00 to 08:00 (pattern 'a24', DS=OOCCOO; winter condition in

241 most sites), equal use of both habitats - open from 20:00 to 08:00, closed from 08:00 to

242 20:00 (pattern 'a33', DS=OOCCCO; fall and spring in most sites), and dominant use of

243 closed habitat from 04:00 to 20:00 (pattern 'a42', DS=OCCCCO; summer in most sites).

244 In the pattern names, the character 'a' refers to daylight-night habitat alternation, where

245 the first number refers to the number of four-hour time periods spent in closed habitat

246 during daytime and the second to the number of four-hour time periods spent in open

247 habitat during the night. Finally, we defined a pattern of random use of the landscape

248 (hence reflecting its structure), or pattern 'U', as a control (i.e. constant selection

249 coefficient for open and closed, see Appendix S3). The seeds of stochastic simulations

250 were random release locations within each individual MCP, whereas successive steps

251 were based on six sets of habitat selection rules matching the aforementioned sequential

252 behaviours. For completeness, we ran the movement simulations with three selection

253 coefficients for each selection pattern to account for behavioral variability (selection

254 coefficient ratios: low, 1:0.2; intermediate, 1:0.1; high, 1:0.01). Each movement

255 simulation was repeated 50 times for each parameter set (i.e., six habitat selection rules

256 by three selection coefficients), varying release location between repetitions but holding

257 release location constant across parameter sets for any given repetition. We thus

258 obtained 900 simulated sequences per individual MCP (six habitat selection rules by

259 three selection coefficients by fifty repetitions).

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

261 We obtained 404 individual dissimilarities trees (see above) based on individual-specific

262 simulated sequences, thus illustrating the dissimilarity among expected habitat use

263 sequences for each individual separately. Note that HD can be customised by assigning

264 weights to character substitutions when comparing sequences. For example, in De

265   Groeve *et al.* (2016) substitution weights were derived from spatial autocorrelation of

266   habitat classes within simulated landscapes. However, in IM-SAM we considered

267   substitution weights to be redundant because simulated sequences were modelled within

268   each individual MCP and habitat was classified as a binary category. Hence, spatial

269   structure, and thus spatial autocorrelation, was directly accounted for by using this

270   individual-specific modelling approach.

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

273   In the obtained trees, the leaves are the sequences, and remaining nodes represent

274   clusters (groups of sequences) for which the branch lengths measure the extent of

275   dissimilarity. In other words, the longer the branch length, the higher the dissimilarity

276   between clusters. Hence, the number of clusters that are identified in a tree depends on a

277   cut-off value that should be selected to obtain the most robust dissimilarity tree (Hennig

278   *et al.* 2007). To assess robustness, an iterative procedure of sequence re-clustering such

279   as bootstrapping is generally used. Bootstrapping metrics express the consistency in

280   reclassification of sequences in the same clusters. The same procedure can be repeated

281   for different cut-off values (and therefore number of clusters) using the optimisation of

282   bootstrapping metrics as a criterion to decide upon the best cut-off point. In IM-SAM,

283   we propose a standardised procedure to identify the most robust and informative tree.

284   For the trees based on real trajectories only (Fig. 1a) we did not include any cut-

285   off, since we used them for exploratory purposes only. For the clustering of simulated

286   trajectories only, and both simulated and real trajectories in the final classification (Fig.

287   1c), instead, we allowed all cut-off values (that correspond to distances from the last

288   common node) that generated from a minimum of 2 to a maximum of 20 clusters per

289   tree. For each cut-off value, we computed the Jaccard bootstrapping index for each

290   cluster using bootstrap resampling of *n* number of sequences (BJ, or bootmean; see

291   Hennig 2007, R-package fpc, clusterboot) using 1000 iterations (De Groeve *et al.*

292   2016), and then we computed the median BJ of all clusters ($BJ_{MD} \in [0, 1]$), and the BJ

293   interquartile range ($BJ_{IQ} \in [0, 1]$). We finally defined a combined bootstrapping index

294   ($BJ_{IQMD} \in [0, 1]$) that we computed for each cut-off value:

295
$$BJ_{IQMD} = \frac{BJ_{MD} + [(1 - BJ_{IQ})]}{2}$$

296   To evaluate the optimised cut-off value, we used a semi-automated selection

297   procedure based on the optimisation of the BJ_{IQMD} index. Specifically, we plotted BJ_{IQMD}

298   in relation to the number of clusters for each individual (Fig. 2, top panel and Appendix

299   S4, Fig. S4.3). In most cases, the plot showed two peaks: a primary peak, corresponding

300   to a cut-off value that generated trees with two to three clusters, that separates

301   sequences based on general dissimilarity (for example: homogeneous sequences from

302   all others); and a secondary peak, corresponding to a cut-off value that generated trees

303   with five to eight clusters, catching the complexity of the sequences, i.e. the diversity of

304   sequential habitat use patterns generated by the simulations (see also De Groeve *et al.*

305   2016). Hence, we excluded the primary peak and defined the cut-off value based on

306   maximum values of BJ_{IQMD} within the secondary peak (Fig. 2, upper panel, light blue

307   region). Then, we did a visual check of all trees derived from the cut-off values selected

308   as above to identify potential inconsistencies, for example if some obvious clusters were

309   overlooked by the cut-off criterion, or if the pruned tree of real trajectories (see below)

310   showed a consistent structure.

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

312   Once we obtained the 'optimal' tree, we classified each cluster on the basis of the

313   sequential patterns that composed the cluster ('cluster tagging'). First, we calculated the

314   proportion of each sequential habitat use pattern in each cluster. Then, to identify cluster

315   types, we appended each pattern representing at least 10% of the cluster to a tag,

316   ordered by abundance (first criterion) and giving priority to homogeneous patterns

317   (second criterion). For example, if a cluster consisted of 40% a24 sequences, 30% a33

318   sequences, and 20% a42 sequences, the resulting tag was a24_a33_a42 (first criterion);

319   or, a composition of 30% of O sequences, 30% of a24, 20% of a33, and 15% of a42,

320   was tagged as o_a24_a33_a42 (second criterion). We also used a simplified

321   classification by retaining the most abundant sequential habitat use pattern only as the

322   cluster tag (e.g. a24 and O in the two examples above, respectively).

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

324   In a final step, we obtained the classification tree for each individual by re-running the

325   HD algorithm on both real and simulated sequences, and by using the cut-off distance as

326   defined above. This way, simulated sequences were used as a *guide for classification of*

327   *real sequences* to their most similar cluster type (real habitat use sequence tagging, Fig.

328   1c). For visualisation purposes, we pruned the classification tree by only retaining real

329   sequences (Fig. 1d).

330   **Results**

331   *Classification of individual-specific simulation trees*

332   The application of the HD algorithm to simulated biweekly sequences generated 404

333   trees, one per individual MCP. According to the simulation procedure (see above, From

334   SAM to IM-SAM, par. a; Fig 1b), we expected trees to be composed of six clusters with

335   150 sequences each. Instead, we found a substantial deviation from this expectation (see

336   Appendix S5 for a sample of 35 trees) that we attributed to individual differences in the

337 environmental context, and notably in the relative proportion of open vs. closed habitat

338 within the individual's MCP.

339   Using the IM-SAM cut-off criterion (i.e., the second peak in the plot of BJ_{IQMD}

340 for all cut-off values; Fig. 2, top panel), we automatically identified 394 simulation

341 trees, mainly composed of five (164 trees, 40.59%), six (192 trees, 47.52%) or seven

342 (38 trees, 9.41%) clusters, whereas trees with three or eight clusters were rare (7 and 3

343 respectively, or 1.73% and 0.74%). The overall average BJ_{IQMD} of these trees was

344 0.749±0.111 which corresponded to a high BJ_{MD} (0.850±0.066), indicating high inter-

345 cluster dissimilarity, and a low BJ_{IQ} (0.123±0.078), indicating low variability in inter-

346 cluster dissimilarity, thus confirming the robustness of the cluster classification. For 32

347 trees, the number of clusters was manually adjusted to distinguish clear and obvious

348 clusters (from five or six to six, seven, or eight clusters), but this resulted in negligible

349 change in average BJ_{IQMD} (decrease of 0.007, 0.742±0.115). Furthermore, we noted that

350 the relative proportion of open vs. closed habitat in the MCP affected cluster quality (i.e.

351 clustering robustness). Specifically, BJ_{IQMD} increased when the proportion of closed and

352 open habitats was more or less equal (Fig. 2, top panel: higher to lower values from

353 dark to light trend lines). The corresponding BJ_{MD} values also increased when the

354 proportions of open and closed habitats were similar (Fig. 2, top panel: purple shade on

355 the trend lines), whereas the corresponding BJ_{IQ} values decreased (Fig. 2, top panel:

356 orange to blue points along the trend lines). Despite these differences between classes of

357 habitat proportion, the trend in BJ_{IQMD} was consistent, with a second peak for values of

358 between five and seven clusters, with a rapid drop after that. [Figure 2 near here]

359   The proportion of open vs. closed habitat in the MCP also affected the cluster

360 size within simulation trees. When the proportion of open and closed habitat in the MCP

361 was similar, the trees for simulated sequences were more evenly classified between

362   different clusters. However, when a given habitat type was preponderant in the MCP,

363   some clusters were composed of a larger number of sequences than others (Fig. 2, lower

364   panel, e.g. median lower than 150, and more outliers for class 0.1).

365   *Identification of cluster types ('cluster tagging')*

366   We identified 16 main cluster types among the 404 simulated trees, each representing

367   more than 0.5% of all clusters (Fig. 3, circles in the lower panel). The 2300 clusters

368   consisted of one or more sequential habitat use patterns in differing proportions.

369   Specifically, 1613 clusters consisted of one sequential habitat use pattern (70% of total

370   number of clusters, Fig.3, top-left panel, and legend of cluster types in the lower part:

371   a24, a33, a42, c, o, u, present in 378, 272, 286, 290, 170, and 125 clusters, respectively).

372   513 (22%) and 139 (6%) clusters consisted of two or three sequential habitat use

373   patterns, respectively (o_u, a33_a42, c_u, a33_a24, a24_a33, a42_a33, u_a33, o_u_a24,

374   c_u_a42, present in 170, 125, 89, 52, 29, 21, 13, 66, and 27 clusters, respectively). The

375   remaining 33 cluster types represented less than 3% of all clusters in total and were also

376   used to classify real trajectories (these rare cluster types were omitted from the legend in

377   Fig. 3; see Appendix S6 for the full set of classified cluster types). [Figure 3 near here]

378   The number of sequential habitat use patterns occurring in clusters was affected

379   by the relative proportion of open vs. closed habitat in the MCP (Fig. 3, top-left panel).

380   That is, we observed clear-cut cluster identification (i.e., one pattern per cluster) for

381   trees derived from MCPs with similar relative proportions of open and closed habitat.

382   More 'unclear' cluster identification (i.e., with a mix of 3 to 4 sequential habitat use

383   patterns) was observed for trees derived from MCPs with a preponderance of one

384   habitat type only.

385   Similarly, the occurrence and relative importance of cluster types within trees

386   also depended on the relative proportion of open vs. closed habitat in the MCP (Fig. 3,

387   top-right panel). Indeed, sequences with random and alternating habitat use patterns

388   grouped together with homogeneous sequences when one habitat was prevalent in the

389   MCP (Fig. 3, top-right panel; the patterns are mirrored for high proportion of open or

390   closed habitat). Importantly, only a small proportion of alternating patterns clustered

391   together with a random pattern of use, indicating that alternating patterns rarely

392   occurred at random.

393   To sum up, we showed that sequences of habitat use patterns generated using

394   pre-defined habitat use processes within individual MCPs mostly clustered amongst

395   themselves, as expected, but there was some variation in the pattern. Clusters were

396   characterised by one or more sequential habitat use pattern, and this 'cross-pattern'

397   clustering was dependent on the relative composition of open and closed habitats in the

398   MCP.

399   *Classification of real animal trajectories into cluster types*

400   After classification of simulated sequences, we re-ran SAM also including real

401   trajectories so that these were grouped with the most similar cluster types. After pruning

402   (i.e., filtering out of the simulated sequences), we obtained the classification tree of the

403   real sequences for each individual (See Fig. 1d for an example, and Appendix S7 for a

404   sample of 35 trees). In total, 69.40% of the real habitat use sequences were classified

405   into only six cluster types (i.e., o_u, o_u_a24, c_u, a33, a42, a24, Table 1, in bold), and

406   97.55% of all real sequences into 17 cluster types (Table 1). The remaining 2.45% of all

407   real habitat use sequences matched another 21 cluster types. All six a-priori simulated

408   sequential habitat use patterns were represented amongst the real data sequences.

409   Specifically, 40.11% of the sequences were classified as one of the six cluster types

410   including a single sequential habitat use pattern (for example, 4.83% of the real

411   sequences were classified as homogeneous closed, c, and 11.85% were classified as

412   alternation a33, Table 1). Another 38.20% were classified into cluster types that

413   included a combination of two patterns, especially a combination of homogeneous

414   open/closed with random patterns (i.e., 29.40%, o_u, c_u, Table 1), or a combination of

415   two alternating patterns (i.e., 8.26%, a33_a42, a33_a24, a42_a33, a24_a33). Finally, the

416   remaining 21.69% of the real sequences were classified into cluster types that included a

417   combination of three or more sequential habitat use patterns. When these results were

418   considered with the most simplified classification (i.e. retaining the most abundant

419   pattern only), the majority of the sequences were classified as homogeneous open (o,

420   31.51%), and homogeneous closed (c, 24.38%). More than 40% of the sequences were

421   classified as one of the three types of habitat alternation (a33, 18.68%; a42, 10.90%;

422   a24, 10.07%), while the smallest proportion of habitat patterns corresponded to random

423   sequential use of habitat (u, 4.46%). [Table 1 near here]

424   **Discussion**

425   In this paper, we propose an ecological application of Sequence Analysis Methods, IM-

426   SAM to describe sequential habitat use of animal trajectories applied to European roe

427   deer across contrasting landscapes. Below, we first discuss the methodological advances

428   of IM-SAM. Second, we consider the ecological relevance of the observed spatio-

429   temporal patterns of roe deer sequential habitat use across Europe. Finally, we discuss

430   the broader applicability of IM-SAM for other ecological and geographical data.

431   *IM-SAM procedure*

432   IM-SAM provides a suitable method to detect similarity in sequential patterns in

433   movement data of animal species. The IM-SAM framework involves three

434   methodological steps. First, exploration trees are built using real sequences only (Fig.

435   1a). Then simulation trees are generated taking into account the individual spatial

436   context using simulated sequences only (Fig. 1b/c). Finally, classification trees are

437   produced based on real and simulated sequences combined (Fig. 1c/d). While this three-

438   step conceptual framework is identical to De Groeve *et al.* (2016), scaling up to a multi-

439   population approach involved several fundamental methodological adjustments which

440   we summarise in Table 2, and that improved the procedure. One of the most important

441   advances of IM-SAM compared to other previous ecological applications of the SAM

442   framework (De Groeve *et al.* 2016) is to account for the *individual* spatial context in

443   which an animal moves, by generating individual-specific, spatially-explicit simulated

444   sequences. In this way, individual sequential patterns of habitat use can be identified in

445   a comparable manner across a diversity of landscapes, as done here, facilitating multi-

446   population comparisons. Moreover, in this application, we generated simulations based

447   on expected day-night habitat use patterns. While many natural processes follow

448   alternating and repetitive rhythms (e.g. Wittemyer *et al.* 2008, Bar-David *et al.* 2009,

449   Benhamou and Riotte-Lambert 2012, Polansky *et al.* 2010, Péron *et al.* 2016, 2017,

450   Riotte-Lambert *et al.* 2013, 2017), this might not always be a pattern of interest for

451   other studies. In general, the simulation rules must be based on the question addressed,

452   on the behavioral traits of the species, and the spatio-temporal resolution of the study.

453   For example, when studying migration-timing and the use of summer vs winter ranges,

454   simulations might be better based on a weekly timeframe. [Table 2 near here]

455   The IM-SAM procedure only detects sequential patterns that are coded within

456   the simulation rules. While this appears to be a constraint at first sight, the approach

457   based on simulated movement rules and exploratory trees enables classifying real

458   trajectories within a hypothetical-deductive framework, i.e. based on reproducible

459   expectations. In this sense, exploratory trees represent the empirical observations on

460   which to build the set of hypotheses.

461   The underlying behavioral choice mechanism of our movement model used to

462 simulate sequential patterns within the individual reference area of use (MCP) is

463 simultaneous (*sensu* Van Moorter *et al.* 2013), therefore a given habitat type will be

464 used more with increasing availability. It is important to underline that the preference

465 for the habitat is *fixed*, hence variable use with availability does not correspond to a

466 functional response. An alternative behavioral choice mechanism, for instance

467 hierarchical, could be integrated in the movement model. In such an approach, the use

468 of a given habitat type would be independent of its availability (Van Moorter *et al.*

469 2013). As of today, movement models have overwhelmingly used a simultaneous

470 behavioral choice mechanism (Van Moorter *et al.* 2013). Although further research is

471 required, empirical evidence suggests that simultaneous choice is appropriate (Schuck-

472 Paim and Kacelnik 2007). In terms of IM-SAM, the above indicates that it is easier to

473 distinguish different sequential habitat use patterns when the relative proportion of

474 habitats available to the individual is similar.

475   In comparison with the simplified model used in De Groeve *et al.* (2016), we

476 here simulated sequences within the true landscape context, giving more complex and

477 variable patterns, and hence resulting in more complex dissimilarity trees, requiring

478 methodological refinements to obtain robust classifications. In classification trees, the

479 optimal number of clusters can be derived using many different approaches. Several

480 SAM-applications (e.g. Shoval and Isaacson 2007) define the cut-off value by visual

481 exploration of clusters in dissimilarity trees without considering the robustness of the

482 clusters. More objective methods use within- and between-cluster quality assessments,

483 such as silhouette plots (Rousseeuw *et al.* 1987) and the Calinski-Harabaz Index

484 (Caliński and Harabasz 1974), or, as often used in DNA-analysis, cluster stability

485 procedures based on bootstrapping (e.g. Jaccard bootstrapping, BJ). In our case, the

486   sequences corresponded to simulated behaviors (i.e., discrete trajectories) that were

487   obtained through a set of stochastic rules applied to real and highly heterogeneous

488   environments. Hence, some variability in the output sequences, and so in the clustering,

489   can be expected, especially when the proportion of alternative habitats is highly unequal

490   within individual MCPs (see Fig. 3). For this reason, we extended the approach of De

491   Groeve *et al.* (2016), based on median values in BJ, by combining a central tendency

492   (BJ$_{MD}$), and a dispersion measure (BJ$_{IQ}$) of cluster quality into a unique index. We think

493   that this procedure could be appropriate for other SAM applications, as it represents a

494   semi-automated standardised approach.

495   One of the most important advantages of IM-SAM is the possibility to express

496   sequences as multi-level habitat categories, as showed in other studies (De Groeve *et al.*

497   2016, Brum-Bastos *et al.* 2018). De Groeve *et al.* (2016) annotated trajectories with

498   contextual information derived from two habitat variables (elevation and habitat

499   openness) expressed as single character codes (i.e., high-open, high-closed, low-open,

500   low-closed), and Brum-Bastos *et al.* (2018) instead generated character codes for each

501   context variable which were then analysed as multi-channel sequences. Here, we used a

502   simple case of two alternative habitat types (open vs. closed) that showed promising

503   sequential pattern variability in a single roe deer population (De Groeve *et al.* 2016).

504   Note that for continuous or discretised habitat variables, which are expected to be

505   spatially correlated, substitution weights are essential to correct for classes that are more

506   similar to each other. For example, in the case of four habitat classes with different

507   forest cover density (<25%, 25-50%, 50-75%, >75%), a forest cover density of <25% is

508   more similar to the 25-50% category than to the >75% one. In the R package TraMineR

509   such substitution weights can be based on automatic computation of transition rates,

510   probability or user-defined (Gabadinho *et al.* 2011). While automatic computation of

511   substitution weights is sufficient for exploration trees, we recommend assessing them

512   directly by measuring spatial correlation at the relevant scale (i.e., median step length;

513   see De Groeve *et al.* 2016) for simulation and classification trees. However, while the

514   exploration phase can handle a large number of classes, the complexity of simulated

515   sequential habitat use patterns increases with the number of habitat categories, hence we

516   suggest using the exploration phase to identify the most essential for simulation and

517   classification. Multi-channel sequence analysis, as proposed by Brum-Bastos *et al.*

518   (2018), offers interesting future avenues for more complex combinations of habitat (or

519   contextual) variables.

520   In IM-SAM, simulations of sequential habitat use patterns are performed in the

521   individual reference area of use that were obtained with a simple geometric method

522   (MCP 100 %). However, there is no limitation on using other methods to assess the area

523   of use, for example, to overcome the sample size dependence of MCP (Spencer *et al.*

524   1990, Powell *et al.* 2000).

525   *Ecological insights and geographical applications*

526   In our study, 40% of the real sequences from all roe deer populations were classified as

527   alternating patterns between open and closed habitats. This suggests that not only the

528   proportion, but also the sequential order in which open and closed habitats are used, is

529   an important metric for characterising the space use strategy of individual roe deer.

530   Activity and physiological circadian cycles, such as feeding-rumination, may explain

531   the observed alternation between open and closed habitats. Indeed, roe deer are known

532   mainly to select for forest and cover habitats during rumination and resting (Cederlund

533   1981, Mysterud and Østbye 1995), and to favour edges and open areas during peak

534   foraging activity, at twilight and during the night (Pagon *et al.* 2013). This pattern may

535   be less pronounced in areas with less human disturbance, such as for a Canadian elk

536   population (Ensing *et al.* 2014). Indeed, in human-dominated European landscapes,

537   habitat alternation is likely a behavioral response of ungulates to both landscape

538   heterogeneity and temporal variation in human activities. Because rich open landscapes

539   are often associated with higher risk of predation or disturbance, in such human-

540   dominated environments, prey species must generally trade their acquisition of high-

541   quality resources against risk avoidance (Godvik *et al.* 2009). By alternating between

542   rich open areas and more closed forest habitats, with less forage but a higher degree of

543   shelter, prey may hence resolve the risk-resource trade-off (Fraser and Huntingford

544   1986). In particular, wild ungulate species, including roe deer, generally use closed

545   refuge habitats during daytime, when human disturbance is greater, and rich open

546   habitats during night time (e.g. Bonnot *et al.* 2013; Padié *et al.* 2015).

547   A large proportion of real habitat use sequences of roe deer were also classified

548   as homogeneous open, meaning that a non-negligible number of deer intensively used

549   human-exploited agricultural lands during both day and night (i.e., crops: South-France,

550   Southern Germany; husbandry: Switzerland; Aulak and Babinska-Werka 1990). Indeed,

551   agricultural areas may simultaneously provide both high-quality food and cover

552   resources for roe deer, at least during certain parts of the year (Hewison *et al.* 2001,

553   Cimino and Lovari 2006, Bjørneraas *et al.* 2011). Homogeneous closed sequences, on

554   the other hand, were more common in forest landscapes, but also occurred in

555   agricultural landscapes, suggesting that some individuals are strictly associated with a

556   given habitat type. Finally, our results indicate that most of the time, the sequential use

557   of open and closed habitats by roe deer was not random, since only a small proportion

558   of real sequences were tagged with a random pattern of habitat use.

559   A further step towards understanding the ecological significance of sequential

560   habitat use would involve accounting for seasonal and individual variability, and to

561   include further covariates, such as specific landscape features (e.g., fragmentation), or

562   individual life-history traits (sex, age). IM-SAM 'tags' can be used with levels of a

563   categorical variable in established statistical modelling frameworks, such as multivariate

564   statistics (e.g. Jongman *et al.* 1995) or Generalised Linear Modelling (e.g. Pinheiro and

565   Bates 2000; Zuur *et al.* 2009). SAM was originally applied to the ecological context as a

566   spatio-temporal exploratory tool (De Groeve *et al.* 2016). IM-SAM takes this a step

567   forward, opening up the potential to use spatio-temporal patterns as a variable in spatial

568   ecological modelling.

569   ***IM-SAM applicability***

570   While IM-SAM was applied here on animal habitat use sequences, human

571   geography may also benefit from this novel framework. From a technical point of view,

572   while several Sequence Analysis studies exist in time geography, to the best of our

573   knowledge, robust classifications such as those obtained by bootstrapping and the use of

574   BJ_{IQMD} have rarely been explored. Moreover, the routine applicability of IM-SAM could

575   be supported by the use of a popular data analysis software (R version 3.4.1., R Core

576   Team 2017; package TraMineR, Gabadinho *et al.* 2011). Conceptually, with the ongoing

577   advances in human and animal tracking techniques, IM-SAM could ultimately be used

578   as a tool to simultaneously compare patterns of space use in animals and humans. For

579   example, mapping sequential animal and human space use in the same area could help

580   understand if and how they differ or conflict. Alternatively, potential effects of traffic,

581   recreation, hiking, cycling and other human activities could be assessed by modelling

582   them as environmental drivers of sequential habitat use. Furthermore, after

583   characterising animal and human sequential space use, one could explore the sequential

584   pattern of non-movement related metrics obtained through biologging, such as activity,

585   heart rate, body temperature, or food intake (see Ropert-Coudert *et al.* 2005).

605

606   **References**:

607

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

610

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

613

614   Andrienko, G., Andrienko, N. and Heurich, M., 2011. An event-based conceptual model

615   for context-aware movement analysis. *International Journal of Geographical*

616   *Information Science*, **25** (9), 1347-1370.

617

618   Aulak, W. and Babińska-Werka, J., 1990. Use of agricultural habitats by roe deer

619   inhabiting a small forest area. *Acta Theriologica*, 35 (1-2), 121-127.

620

621   Bar-David, S., Bar-David, I., Cross, P.C., Ryan, S.J., Knechtel, C.U. and Getz, W.M.,

622   2009. Methods for assessing movement path recursion with application to

623   African buffalo in South Africa. *Ecology*, 90 (9), 2467-2479.

624

625   Benhamou, S. and Riotte-Lambert, L., 2012. Beyond the Utilization Distribution:

626   Identifying home range areas that are intensively exploited or repeatedly visited.

627   *Ecological Modelling*, 227, 112-116.

628

629   Berger-Tal, O. and Bar-David, S., 2015. Recursive movement patterns: review and

630   synthesis across species. *Ecosphere*, 6 (9), 1-12.

631

632   Bjørneraas, K., Solberg, E.J., Herfindal, I., Moorter, B.V., Rolandsen, C.M., Tremblay,

633   J.P., Skarpe, C., Sæther, B.E., Eriksen, R. and Astrup, R., 2011. Moose Alces

634   alces habitat use at multiple temporal scales in a human-altered landscape.

635   *Wildlife Biology*, 17 (1), 44-54.

636

637   Bonnot, N., Morellet, N., Verheyden, H., Cargnelutti, B., Lourtet, B., Klein, F. and

638   Hewison, A.J.M., 2013. Habitat use under predation risk, hunting, roads and

639   human dwellings influence the spatial behaviour of roe deer. *European Journal*

640   *of Wildlife Research*, 59, 185-193.

641

642   Brum-Bastos, V.S., Long, J.A. and Demšar, U., 2018. Weather effects on human

643   mobility: a study using multi-channel sequence analysis. *Computers,*

644   *Environment and Urban Systems*.

645

646   Cagnacci, F., Boitani, L., Powell, R.A. and Boyce, M.S., 2010. Animal ecology meets

647   GPS-based radiotelemetry: a perfect storm of opportunities and challenges.

648   *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365,
649   2157-2162.

651   Cagnacci, F., Focardi, S., Heurich, M., Stache, A., Hewison, A.J., Morellet, N.,
652   Kjellander, P., Linnell, J.D., Mysterud, A., Neteler, M. and Delucchi, L., 2011.
653   Partial migration in roe deer: migratory and resident tactics are end points of a
654   behavioural gradient determined by ecological factors. *Oikos*, 120 (12), 1790-
655   1802.

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

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

663   Chavoshi, S.H., De Baets, B., Neutens, T., De Tré, G. and Van de Weghe, N., 2015.
664   Exploring dance movement data using sequence alignment methods. *PloS one*,
665   10 (7), e0132452.

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

671   Cimino, L. and Lovari, S., 2003. The effects of food or cover removal on spacing
672   patterns and habitat use in roe deer (Capreolus capreolus). *Journal of Zoology*,
673   261 (3), 299-305.

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

680   Delafontaine, M., Versichele, M., Neutens, T. and Van de Weghe, N., 2012. Analysing
681   spatiotemporal sequences in Bluetooth tracking data. *Applied Geography*, 34,
682   659-668.

683

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

687

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

691

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

696

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

701

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

704

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

707

708   Gabadinho, A., Ritschard, G., Mueller, N.S. and Studer, M., 2011. Analyzing and
709   visualizing state sequences in R with TraMineR. *Journal of Statistical Software*,
710   40 (4), 1-37.

711

712   Godvik, I.M.R., Loe, L.E., Vik, J.O., Veiberg, V., Langvatn, R., and Mysterud, A. 2009.
713   Temporal scales, trade-offs and functional responses in habitat selection of red
714   deer. *Ecology* 90, 699-710.
715

716   Goto, M. 2003. A chorus-section detecting method for musical audio signals. IEEE
717   International Conference on Acoustics, Speech, and Signal Processing (ICASSP)
718   Proceedings; April 2003. 437-440.
719

720   Hennig, C., 2007. Cluster-wise assessment of cluster stability. *Computational Statistics*
721   *and Data Analysis*, 52, 258-271.
722

723   Hewison, A.J., Vincent, J.P., Joachim, J., Angibault, J.M., Cargnelutti, B. and Cibien, C.,
724   2001. The effects of woodland fragmentation and human activity on roe deer
725   distribution in agricultural landscapes. *Canadian journal of zoology*, 79 (4), 679-
726   689.
727

728   Kämmerle, J.L., Brieger, F., Kröschel, M., Hagen, R., Storch, I. and Suchant, R., 2017.
729   Temporal patterns in road crossing behaviour in roe deer (Capreolus capreolus)
730   at sites with wildlife warning reflectors. *PLoS One*, 12 (9), 1-18.
731

732   Kays, R., Crofoot, M.C., Jetz, W. and Wikelski, M., 2015. Terrestrial animal tracking as
733   an eye on life and planet. *Science* 348, aaa2478.
734

735   Magle, S.B., Hunt, V.M., Vernon, M. and Crooks, K.R., 2012. Urban wildlife research:
736   past, present, and future. *Biological Conservation*, 155, 23-32.
737

738   Manly, B.F.J., McDonald, L.L., Thomas, D.L., McDonald, T.L. and Erickson, W.P.,
739   2002. Introduction to Resource Selection Studies. In: *Resource selection by*
740   *animals: statistical design and analysis for field studies*. Dordrecht: Kluwer
741   Academic Publishers, 1-15.
742

743   Mysterud, A. and Østbye, E. 1995. Bed-site selection by European roe deer (*Capreolus*
744   *capreolus*) in southeastern Norway during winter. *Canadian Journal of Zoology*
745   73, 924-932.

746

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

750

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

754

755   Padié, S., Morellet, N., Hewison, A.J.M., Martin, J.L., Bonnot, N., Cargnelutti, B. and
756   Chamaillé-Jammes, S., 2015. Roe deer at risk: teasing apart habitat selection and
757   landscape constraints in risk exposure at multiple scales. *Oikos*, 124 (11), 1536-
758   1546.

759

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

763

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

768

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

773

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

776

777   Podgórski, T., Baś, G., Jędrzejewska, B., Sönnichsen, L., Śnieżko, S., Jędrzejewski, W.
778   and Okarma, H., 2013. Spatiotemporal behavioral plasticity of wild boar (Sus

779   scrofa) under contrasting conditions of human pressure: primeval forest and
780   metropolitan area. *Journal of Mammalogy*, 94 (1), 109-119.
781
782 Polansky, L., Wittemyer, G., Cross, P.C., Tambling, C.J. and Getz, W.M., 2010. From
783   moonlight to movement and synchronized randomness: Fourier and wavelet
784   analyses of animal location time series data. *Ecology*, 91, 1506-1518.
785
786 Powell, R. A. 2000. Animal home ranges and territories and home range estimators. - In:
787   Boitani, L. and Fuller, T. K. (eds), Research techniques in animal ecology:
788   controversies and consequences. Columbia University Press, New York, 65–103.
789
790 R Core Team, 2017. R: A language and environment for statistical computing. R
791   Foundation for Statistical Computing, Vienna, Austria. https://www.R-
792   project.org/.
793
794 Riotte-Lambert, L., Benhamou, S. and Chamaillé-Jammes, S., 2013. Periodicity
795   analysis of movement recursions. *Journal of theoretical biology*, 317, 238-243.
796
797 Riotte-Lambert, L., Benhamou, S. and Chamaillé-Jammes, S., 2017. From randomness
798   to traplining: a framework for the study of routine movement behavior.
799   *Behavioral Ecology*, 28 (1), 280-287.
800
801 Ropert-Coudert, Y. and Wilson, R.P., 2005. Trends and perspectives in animal-attached
802   remote sensing. *Frontiers in Ecology and the Environment*, 3 (8), 437-444.
803
804 Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of
805   cluster analysis. *Journal of computational and applied mathematics*, 20, 53-65.
806
807 Schuck-Paim, C. and Kacelnik, A., 2007. Choice processes in multialternative decision
808   making. *Behavioral Ecology*, **18** (3), 541-550.
809
810 Shoval, N. and Isaacson, M., 2007. Sequence alignment as a method for human activity
811   analysis in space and time. *Annals of the Association of American geographers*,
812   97 (2), 282-297.

813

814   Spencer, S.R., Cameron, G.N. and Swihart, R.K., 1990. Operationally defining home
815   range: temporal dependence exhibited by hispid cotton rats. *Ecology*, *71*(5),
816   1817-1822.

817

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

821

822   Van Moorter, B., Visscher, D., Herfindal, I., Basille, M. and Mysterud, A., 2013.
823   Inferring behavioural mechanisms in habitat selection studies getting the
824   null-hypothesis right for functional and familiarity responses. *Ecography*, **36** (3),
825   323-330.

826

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

829

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

832

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

835

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

840

841   Xavier, G. and Dodge, S., 2014. An exploratory visualization tool for mapping the
842   relationships between animal movement and the environment. *In: MapInteract*
843   *2014 – Proceedings of the 2nd ACM SIGSPATIAL International Workshop on*
844   *MapInteraction,* 36-42.

845

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

849

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

852

853

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				c					a33			a42		a24		u	
31.51				24.38					18.68			10.90		10.07		4.46	

857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875

876 Table 2. Comparison between SAM and IM-SAM, distinguishing the general procedure (a), the three-step

877 framework (exploration, b; simulation, c; classification, d) and the pivotal summarised advantage of IM-

878 SAM (e). The steps that were identical between SAM and IM-SAM are in *italics*.

		SAM	IM-SAM
	Distance metric	*Hamming Distance*	*Hamming Distance*
	substitution weights	Yes	no
a	**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
	movement simulation context	Nine simulated arenas *	Real movement context (MCP)
	Movement model	*Biased and correlated random walk*	*Biased and correlated random walk*
c		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

879 * *Nine simulated arenas that represent the distribution and composition of real home ranges.*
880 *** Real sequences are matched to simulated arenas by measuring the proportion of available habitat at sequence level.*
881 *§ Under the hypothesis that ungulates maintain a disturbance and predator avoidance strategy, using mainly open habitat during*
882 *the night and closed habitat during the day, three types of alternation were simulated in relation to temporal and spatial variation in*
883 *sunrise and sunset.*
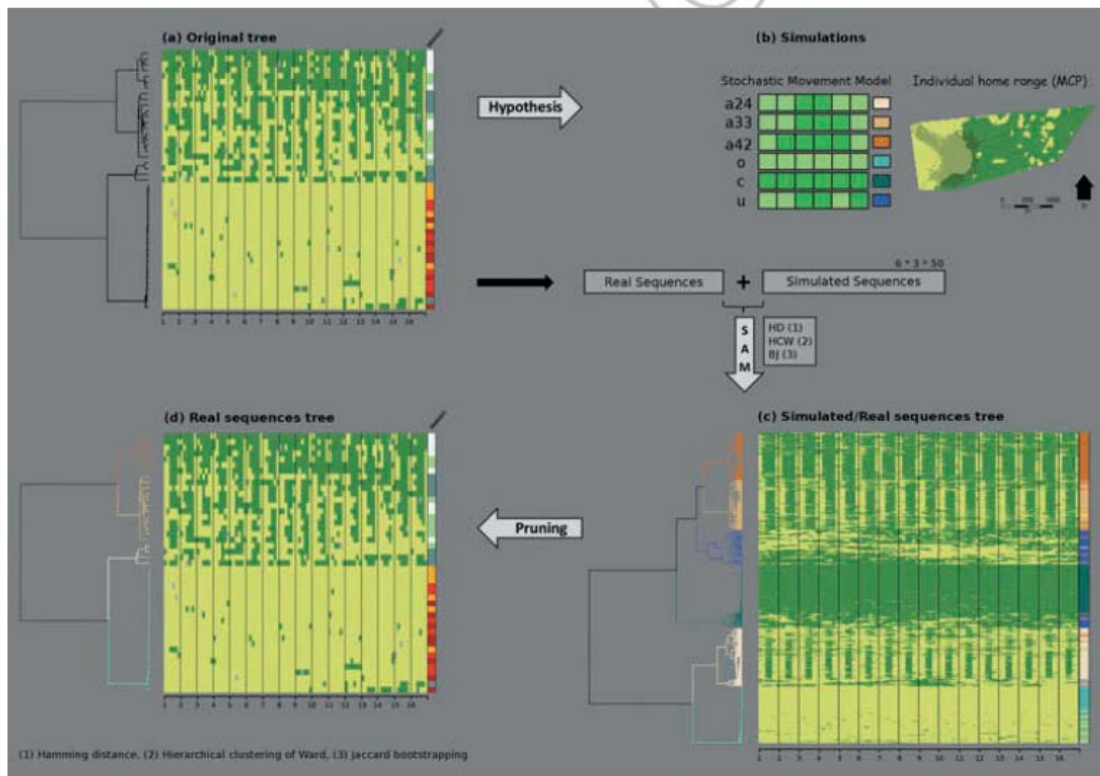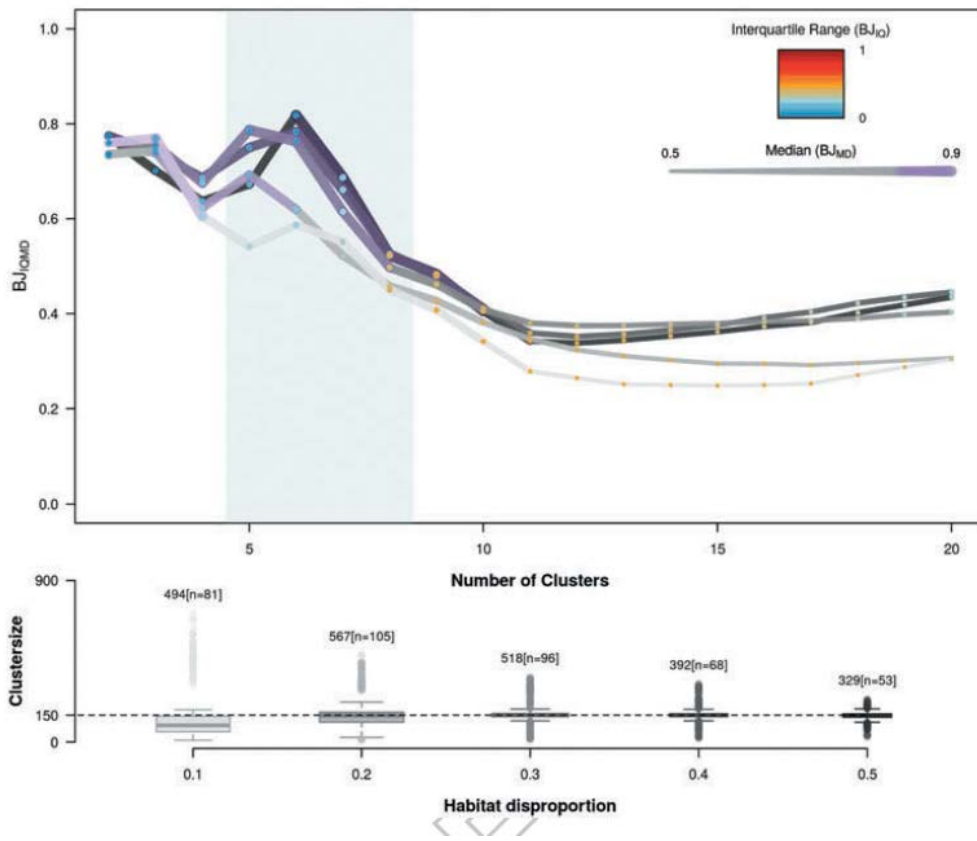884

885

886

887

888

889

890

891

892

893

894

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

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

915 Figure 3. **Bottom panel - Cluster types:** coloured circles represent all sixteen main cluster types
916 identified in simulation trees, scaled by the number of clusters of that type (actual number of clusters in
917 brackets). The colour is the legend for the top-right figure (see below). **Top panel left – Cluster**
918 **composition (a)**: frequency of occurrence of cluster types composed by up to five sequential habitat use
919 patterns. The coloured gradient indicates the proportion of open habitat in MCPs on which simulation
920 trees were based. **Top panel right – Cluster composition (b)**: relative proportion of cluster types in
921 simulated trees (main y-axis), plotted against the proportion of open habitat in MCPs (main x-axis).
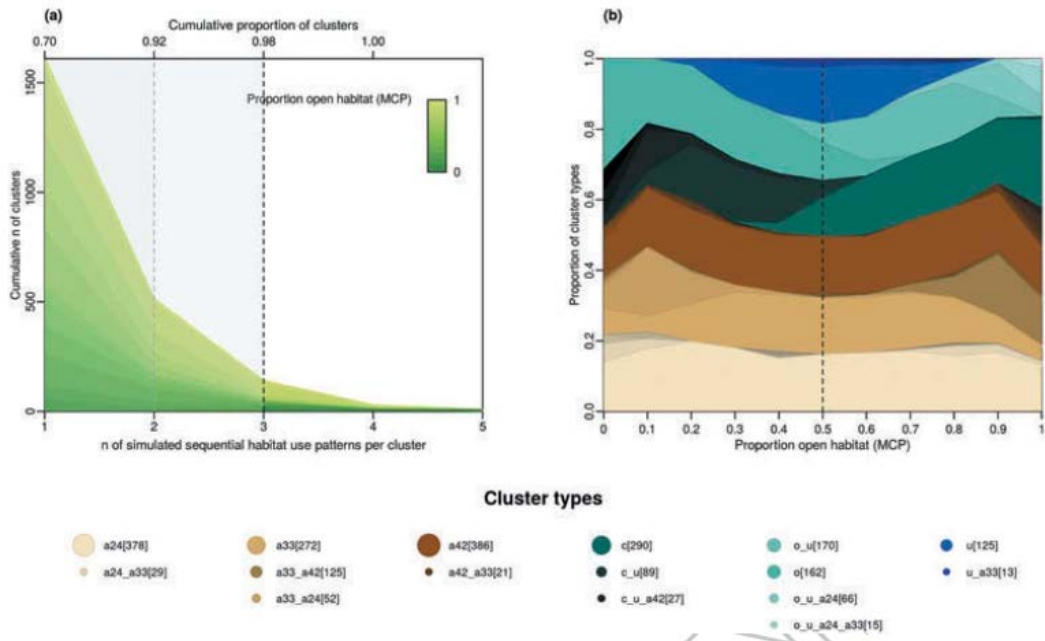922

923

924

925   Fig 1

926

927

928 Fig 2

929

930

931   Fig 3