Temporal segmentation of animal trajectories informed by habitat use

Mariëlle L. van Toor,[1,2],† Scott H. Newman,[3] John Y. Takekawa,[4,5] Martin Wegmann,[6] and Kamran Safi[1,2]

[1]*Department of Migration and Immuno-Ecology, Max Planck Institute for Ornithology, Am Obstberg 1, 78315 Radolfzell, Germany*
[2]*Department of Biology, University of Konstanz, Universitätsstraße 10, 78464 Konstanz, Germany*
[3]*Food and Agriculture Organization of the United Nations, Emergency Centre for Transboundary Animal Diseases,*
No. 3 Nguyen Gia Thieu Street, Hanoi, Vietnam
[4]*U.S. Geological Survey, Western Ecological Research Center, San Francisco Bay Estuary Field Station, 505 Azuar Drive,*
Vallejo, California 94592 USA
[5]*Science Division, National Audubon Society, 220 Montgomery Street, San Francisco, California 94104 USA*
[6]*Department of Remote Sensing, Institute for Geography and Geology, University of Würzburg, Campus Hubland Nord 86,*
97074 Würzburg, Germany

Abstract.   Most animals live in seasonal environments and experience very different conditions throughout the year. Behavioral strategies like migration, hibernation, and a life cycle adapted to the local seasonality help to cope with fluctuations in environmental conditions. Thus, how an individual utilizes the environment depends both on the current availability of habitat and the behavioral prerequisites of the individual at that time. While the increasing availability and richness of animal movement data has facilitated the development of algorithms that classify behavior by movement geometry, changes in the environmental correlates of animal movement have so far not been exploited for a behavioral annotation. Here, we suggest a method that uses these changes in individual–environment associations to divide animal location data into segments of higher ecological coherence, which we term niche segmentation. We use time series of random forest models to evaluate the transferability of habitat use over time to cluster observational data accordingly. We show that our method is able to identify relevant changes in habitat use corresponding to both changes in the availability of habitat and how it was used using simulated data, and apply our method to a tracking data set of common teal (*Anas crecca*). The niche segmentation proved to be robust, and segmented habitat suitability outperformed models neglecting the temporal dynamics of habitat use. Overall, we show that it is possible to classify animal trajectories based on changes of habitat use similar to geometric segmentation algorithms. We conclude that such an environmentally informed classification of animal trajectories can provide new insights into an individuals' behavior and enables us to make sensible predictions of how suitable areas might be connected by movement in space and time.

Key words: *Anas crecca;* animal movement; common teal; habitat use; life history; migration; niche dynamics; random forest models; segmentation; simulation; species distribution model; transferability.

† **E-mail:** mvantoor@orn.mpg.de

INTRODUCTION

The technological advances that allow us to follow animals in the wild have revolutionized the field of movement ecology (Cagnacci et al. 2010, Hussey et al. 2015, Kays et al. 2015). Since the invention of simple tags such as bird bands a centennial ago, the miniaturization and increased efficiency in power consumption have given rise to modern tags which transmit or record locations

of the tagged animals at an unprecedented rate. Animal location data have become ever more accurate in space and time, and the duration over which a single individual can be observed steadily increased. Thus, animal movement data have become not only more accurate, but also much more abundant. With the increased spatiotemporal resolution owed to the technological developments, the amount of details that can be gleaned from movement data has also increased. It is now possible to not only know where animals were and how much space they used, but also what they were doing during the time of observation. The contextualization of locations, that is, the ability to put locations in a behavioral context, allows us to address important questions like how an individual allocates its time and energy to specific behaviors. Contextualization also is the basis on which we can associate resource distribution with a more detailed perspective of space use, as well as study the interactions between tagged individuals or even species, particularly in predator–prey dyads. The identification of behavior from animal trajectories thus provides a unique and important perspective on ecology in high detail in the wild.

While in the past mainly expert opinion was used for the behavioral classification of animal trajectories, the exponential growth of collected movement data as well as the necessity for reproducibility poses logistical limits on expert-based contextualization. Hence, behavioral classification is being increasingly based on algorithms often referred to as segmentation algorithms. These algorithms subset a behaviorally heterogeneous trajectory into a discrete number of segments that characterize distinct patterns representing coherent behavior (Gurarie et al. 2016). Current segmentation algorithms often rely on metrics such as speed and tortuosity of the trajectory and are in general based on the geometry of the movement alone (e.g., Gurarie et al. 2009, Garriga et al. 2016).

Trajectories can, however, be characterized not only by their geometry, but also by the environmental conditions an individual was observed in. Certain behaviors like foraging and resting are often tied to a specific habitat, as observed in, for example, Spanish stone martens (*Martes foina*) who use pastures for foraging and orchards for resting (Santos and Santos-Reis 2009). Similarly,

the association between an animal and its environment can change with changing life-history stages. For example, some species of migratory birds use very different habitat in their temperate breeding grounds compared to what they use in the tropical wintering areas (Martínez-Meyer et al. 2004, Nakazawa et al. 2004, Batalden et al. 2007). Finally, habitat segregation can occur for different age classes of the same species, for example, in cave salamanders (*Hydromantes* (*Speleomantes*) *strinatii*, Ficetola et al. 2012).

Thus, behavior across multiple temporal scales can be linked to the environment an individual is observed in. In the reverse conclusion, changes in the association between an individual and its environment could thus indicate changes in behavior. These changes in the association of individuals with their surrounding environment can be the consequence of different processes: The individual moved to a different habitat, the environment itself changed over time, or both happened in parallel, all of which can be indicative of a behavioral change.

We argue that the changes in the relationship between an individual and its environment can, similar to a segmentation based on movement geometry, be utilized for a behavioral segmentation of trajectories. In this study, we suggest a new class of segmentation algorithms that uses environmental correlates of a trajectory, rather than movement geometry, for behavioral classification. We will utilize changes in the realized ecological niche of tracked individuals and, by comparing snapshots of their habitat use, introduce niche segmentation for movement trajectories. Besides being important for a better understanding of how individuals change their behavior in relation to changes in the environment, niche segmentation is also a key to finding a minimum adequate number of time-explicit niche models when modeling habitat use of animals that undergo niche shifts in their life cycle. The identification of distinct realized niche volumes is pivotal to accurately predict where animals will eventually be for a specific life-history stage, or under specific environmental conditions.

Segmentation of animal trajectories by changes in habitat use

Our approach to niche segmentation is based on the classic habitat suitability, or ecological

niche model (also termed species distribution models, SDM), and uses a measure of transferability to cluster subsets of a trajectory into segments based on their similarity in habitat use. Species distribution models are derived from the Grinnellian and Hutchinsonian niche concepts (Grinnell 1917, Hutchinson 1957, Soberón 2007) and capture the environment species preferentially occur in to understand and approximate the potential distribution and abundance of individuals, populations, and species in space (e.g., Sattler et al. 2007). While originally developed for the estimation of a species' ecological niche, SDMs are now also being used on the level of animal groups or individuals to identify intraspecies variation in habitat use, for example, between age classes (Ficetola et al. 2012) or sexes (van Toor et al. 2011). One of the main advantages of SDMs is that they can describe a complex environment in terms of a single metric—habitat suitability—which is based on where a species or an individual occurs, and which environmental conditions are available. In contrast to a multivariate description of the environmental conditions at a given location, habitat suitability at a specific location can be easily compared across, for example, individuals or for different points in time.

To evaluate changes in habitat use along a series of animal locations, we divide an environmentally annotated trajectory into non-overlapping windows of equal window size S (Fig. 1, step 1). We model habitat use within each of these windows using an SDM and subsequently compare habitat use between windows. To do so, we estimate the pairwise similarity in habitat use between all windows using a measure of transferability, which we refer to as the Discriminatory Index (DI, see Section 1 in Appendix S1). The DI quantifies how well a SDM can discriminate between presences and pseudoabsences, taking values between 1 (perfectly correct discrimination) and −1 (complete opposite prediction). By calculating the DI of all pairwise comparisons of windows, we obtain a matrix of transferability that estimates similarity of habitat use across the SDMs of all windows (Fig. 1, step 2). Finally, we group the windows into niche segments based on the similarity of habitat use through a clustering of the transferability matrix (Fig. 1, step 3). To achieve this, we ordinate the transferability matrix such that windows for which DI is higher become placed closer

to each other, and windows for which DI is lower become placed further apart in two-dimensional space. A clustering is then applied to the ordination axes derived from the transferability matrix. In contrast to assuming a priori a number of clusters, we determine the number of clusters that produces the most compliant clustering using the respective cluster silhouettes (Rousseeuw 1987). The resulting clustering of windows is finally used to annotate the original data with an environmentally informed segmentation, which we refer to as niche segmentation (Fig. 1, step 4).

Testing the niche segmentation with simulated data

We test our niche segmentation concept first on simulated environmental and movement data. This allows us to evaluate whether known changes in habitat use can be detected using our approach, an assessment which would be impossible to make in empirical animal trajectories. We consider the two different processes that can lead to shifts in the relationship between an individual and its environment, and their combination: (1) changes in the environment available to the individuals without changes in habitat preference (niche following), (2) changes in the habitat preference of individuals without changes in the environment (niche switching), or (3) both processes in parallel (simultaneous change). By integrating all three of these processes into the simulated data, we can investigate whether changes in habitat preference or changes in the surrounding environment are more likely to be detected by the niche segmentation.

For our simulations we simulate movement trajectories using correlated random walks (CRW, e.g., Codling et al. 2008) which we biased by modeled preferences for the surrounding environmental conditions. Different from an unbiased CRW, the environmental conditions at every possible location are taken into account in an iterative stepwise simulation of the trajectories. This preference for a certain environmental condition is thus incorporated in the movement trajectory, and results in a realized niche reflecting this preference. We introduce changes in habitat use by switching this preference at specific known points in time mimicking niche switching, and/or gradually changing the environmental conditions to simulate niche following and the simultaneous change of both. We then
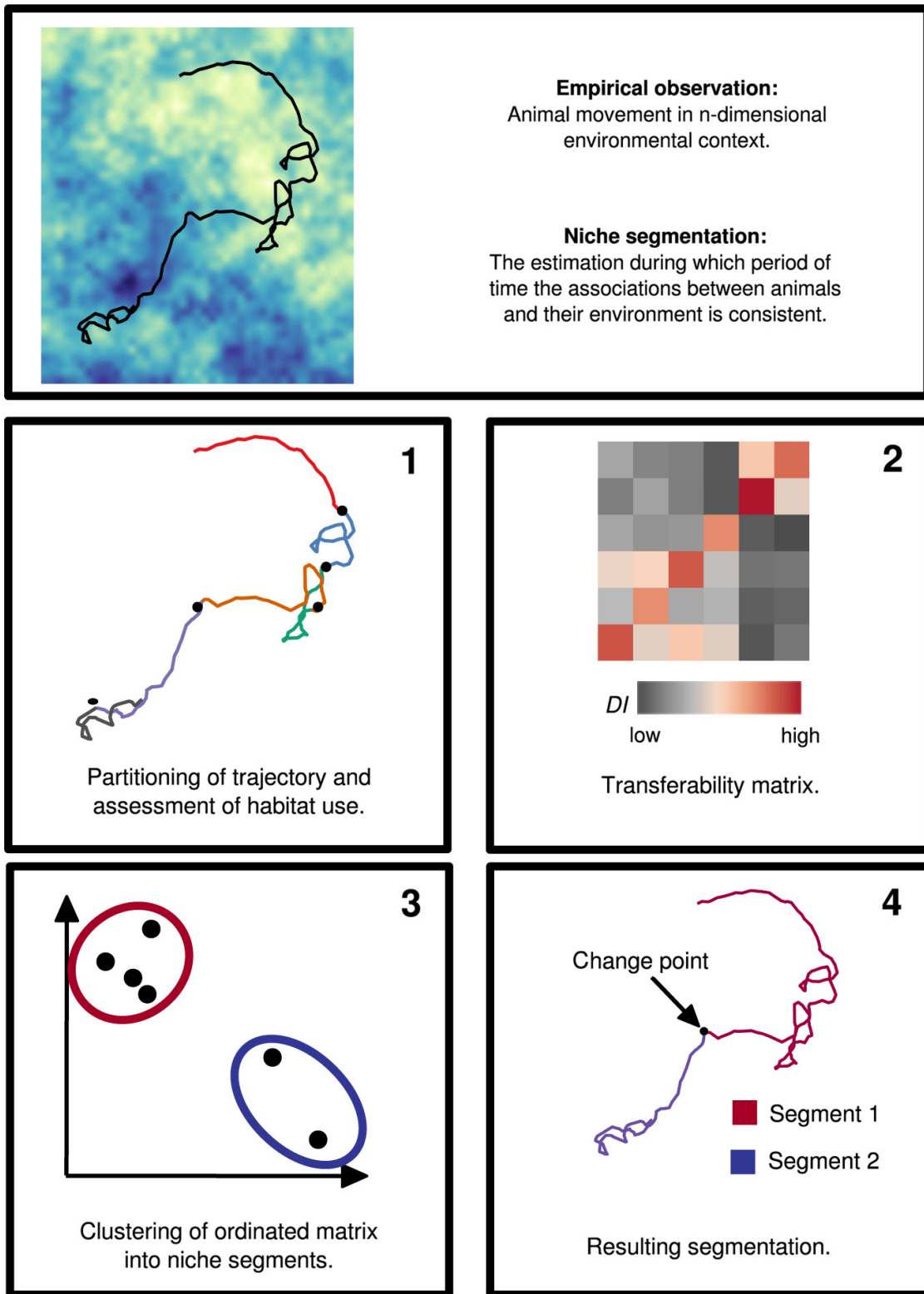
Fig. 1.   Flow chart outlining the process of the niche segmentation.

use the niche segmentation concept to identify break points in the observed habitat use of the simulated trajectories, and test the ability of the method to reconstruct the underlying, simulated process.

Application to an empirical tracking data set of common teal

Finally, we apply the niche segmentation approach to empirical tracking data of the common teal, a small species of dabbling duck with a wide distribution in the Northern hemisphere (*Anas crecca*, L. 1758). We make no a priori assumption about the timing of changes in habitat use over the annual cycle of the common teal, but instead reconstruct it from the data using the niche segmentation. While this species is resident in some parts of its range, most of the populations are considered migratory. We therefore expect to find changes in the temporal signature of habitat use at least as an effect of the migration between breeding and wintering ranges. Typical for empirical animal movement data sets is that they contain many idiosyncrasies making the analysis more challenging compared to simulated trajectories. We show how to address typically occurring characteristics of empirical data in the segmentation approach, such as irregular sampling or the inclusion of location error.

In this study, we focus our main interest on seasonal changes of habitat use. We repeat the niche segmentation using four different window sizes *S* to investigate how the choice of *S* influences the resulting segmentation. Subsequently, we evaluate whether the niche segments detected by the segmentation reflect relevant changes in habitat use known for the species. We derive a set of SDMs specific to the niche segments identified by the segmentation to compare the spatial predictions of habitat suitability to the life

history of common teal as published in the literature. Furthermore, we test how migration and catching site of individuals contributed to the resulting segmentation of the data set. Finally, we compare the performance of SDMs derived for niche segments to SDMs based on the complete data set.

MATERIALS AND METHODS

Data preparation

Simulated data.—We used simulations to test the capability of our segmentation method to detect changes in habitat use under three different scenarios: (1) niche following (constant habitat preference in a changing environment), (2) niche switching (changing habitat preference in a constant environment), and (3) a simultaneous change (changing habitat preference and changing environment). We also allowed other parameters to vary, namely the number of niche segments introduced to the simulated data, and the size of windows the data were partitioned by (*S*). For each scenario, we varied all these parameters to estimate their effect on the accuracy of the method (Table 1), and computed 1000 replicates per scenario and number of niche segments while *S* was chosen randomly. To simulate the movements of individuals in artificial landscapes, we used random fields (Schlather et al. 2015, R-package *RandomFields*, version 3.1.8) and CRWs biased by habitat preference. We created landscapes using a Whittle–Matern covariance model on a grid of 250 by 250 cells of arbitrary size. This size was chosen because it was sufficiently large such that no simulated individual ever encountered an edge, thus avoiding edge effects. The habitat preference of individuals was sampled from the range between the 5%- and 95%-quantiles of the available

Table 1. Setup for the different simulation experiments.

Parameter	Niche following	Niche switching	Simultaneous change
Total number of replicates	4000	4000	4000
Habitat preference	Constant	Variable	Variable
Environment	Variable	Constant	Variable
Number of niche segments	Variable (3–5)	Variable (3–5)	Variable (3–5)
Window size *S*	Random (50–500)	Random (50–500)	Random (50–500)

Notes: Listed are all parameters used and how they were handled for each of the different scenarios. We computed 1000 replicates for every combination of parameters, except for window size which was chosen randomly.

environmental conditions. We calculated a bias layer from the environment by normalizing the absolute difference between the individuals' preference and the environmental value for each grid cell. For each of the simulations, we used a group of five individuals. The starting locations for the individuals were sampled from a circle around the grid center with a radius of 50 cells, and weighted by the corresponding bias layer values. For the biased CRWs, we created a two-dimensional kernel density of step lengths (Weibull distribution with $k = 2$ and $\lambda = 1$) and turning angles (wrapped Cauchy distribution with $\mu = 0$ rad and $\gamma = 0.9$) with 100 by 100 grid cells, representing 1000 potential steps of varying probability.

At each point in time (t), we calculated the putative end locations of all 1000 steps relative to the individual's previous position x_{t-1}. The realized step with the new position x_t was sampled from all possible steps weighted by the product of their probability (the kernel density) and the environmental bias of the corresponding locations. For the three different simulation scenarios, the data for the segmentation were prepared as follows: (1) Niche following; We simulated a gradual change in the environment by shifting the values of the initial environmental layer by an arbitrary amount, and interpolated the number of desired segments between these two layers. Then, all individuals were allowed to take 100 steps on each of the layers, with the starting location on each layer corresponding to the last position on the previous layer. (2) Niche switching; For each niche segment, all individuals were allowed to take 100 steps, resulting in a total of 500 presence locations for all five individuals. After reaching the last location, a new environmental preference was sampled and a new bias layer computed, and the process repeated.

(3) Simultaneous change; In this case, we first created the environmental time series as in (1), then sampled the number of species segments (either smaller or equal to the number of environmental segments). Again, for each of the segments, the individuals were allowed to take 100 steps in the corresponding environment and biased by the corresponding preference. We sampled 500 locations per segment as pseudoabsences to achieve a 1:1-ratio of presences vs. absences. Both presences and pseudoabsences were subsequently annotated with the environmental information in space and time. To prepare the data for segmentation, the complete data set with information on position, presence or absence, environmental conditions, and the true niche segment for both environment and the individuals' habitat preference was partitioned into windows with the predefined size S. We provide a commented R-script that provides all necessary details to repeat the simulations in the Data S1.

Tracking data of common teal

In addition to the simulation study, we tested our method on a tracking data set of the common teal. Between 2007 and 2010, 34 individuals of common teal were caught at five different study sites (China, India, Kazakhstan, Egypt, and Turkey) and equipped with ARGOS tags before release (PTT-100; Microwave Telemetry, Columbia, Maryland, USA). These tracking data are part of a broader disease and migration ecology study implemented by the Food and Agricultural Organization of the United Nations (FAO) and the U.S. Geological Survey (USGS). Locations were taken throughout the day, and 6448 positions for 22 individuals were obtained in total (Table 2). The median sampling frequency for all individuals was 0.83 fixes per hour (25%-quantile: 0.24 fixes per hour, 75%-quantile: 2.40 fixes per

Table 2. A summary over the catching sites and corresponding sample sizes.

Catching site	Year	Number of individuals	First fix taken	Tracking days (mean ± SD)	Locations (mean ± SD)
China/Lake Poyang	2007	3	March 18–20	336 ± 243	289 ± 157
Egypt	2009	8	January 18–November 26	161 ± 126	182 ± 293
India	2008	5	December 9–18	133 ± 81	211 ± 112
Kazakhstan	2008	2	September 15–17	89 ± 47	96 ± 50
Turkey	2010	4	February 10–15	232 ± 99	720 ± 383

Note: The number of tracking days and locations are listed as a mean per individual.

hour). These tracking data show characteristics that are typical of empirical data, and we addressed these using different approaches as described below:

Variation in location error.—A common confounder of tracking data is the use of different tags or geopositioning sensors for the collection of animal movement data. The effect is a variation of location error in the data set, and consequently the corresponding environmental conditions. To account for this, we corrected for location error using temporally explicit estimates of the individuals' space use rather than the actual locations. From these utilization distributions (UD), we derived pseudopresences that better reflect the actual distribution of individuals in space. We used the dynamic Brownian Bridge movement model (dBBMM, Kranstauber et al. 2012, R-package *move*, version 1.2.475) which estimates the UD of an individual from its movement path while also accounting for temporal autocorrelation and the spatial error of locations. Moreover, the dBBMM is time explicit, allowing us to estimate an individual's UD at any given point in time. We applied the dBBMM to each of the individual tracks, using the estimates from Douglas et al. (2012) to associate each location with the respective spatial error according to its ARGOS quality class. Prior to the sampling of pseudopresence locations, we split the dBBMM by day to obtain estimates of the UD during each day of tracking. We sampled 24 locations from the daily 99%-UDs weighted by the likelihood of the individual having used that location as indicated by the UD.

Individuals from multiple populations.—Our data set was comprised of individuals from multiple populations and study years. While such a setup is generally desirable, here individuals from for example China were caught and tagged in 2007, whereas Kazakhstan served as a study site only in 2008. Temporal changes in habitat use in this data set might thus have been caused by changes in the realized niche of individuals, but also by the change in study site over the years. For this reason, we pooled the available location data for all study years and applied the segmentation to the data using Julian days.

Irregular sampling.—Irregular sampling caused by intentionally irregular sampling schemes, missed fixes, or fluctuations in battery power can cause problems with methods that expect a regular time series of locations. For the niche segmentation, irregular sampling is especially relevant to the choice of a suitable window size S. The choice of S influences both the temporal resolution with which changes in habitat can be detected, but also the sample size available for the SDMs. We chose the position of windows so that larger temporal gaps fall between, rather than within windows. To accommodate the resulting differences in temporal spacing of windows, we incorporated this temporal information during the ordination of the transferability matrix using a constraining matrix that reflects the difference between the last date of one window and the first location of the subsequent window in days.

Lack of true absence data.—As true absences were not available, we sampled pseudoabsence data. We used the tracking data to estimate the area available to the individuals instead of drawing random samples from a previously specified spatial area, which has been reported to affect sensitivity and specificity of SDMs (Václavík et al. 2012). To achieve this, we randomized the steps of all individual trajectories. We kept the start and end locations of trajectories constant, and randomized the order of steps in-between these locations. We then used the corresponding step lengths and turning angles to calculate the trajectories in the randomized order. This leads to a wide spread of random tracks around the actual movement of ducks both in space and in time, which we took as a reflection of where the animals could have been during the same time of year. We generated 100 alternative routes for each individual, calculated dBBMMs for these, and sampled pseudoabsence locations in the same way as we did for the presence locations.

We annotated all locations with environmental information using the Movebank Env-DATA-Tool (Dodge et al. 2013). We initially chose 57 variables containing information on climate, land use, elevation, human impact and vegetation features. Variables available as time series were interpolated linearly from the closest available measurements in time weighted by the inverse temporal distance. We eliminated all environmental variables for which more than 10% of the data were missing, so that only 19 variables remained (Appendix S1: Table S1).

We repeated the segmentation procedure (see section *Segmentation*) using four different window sizes for the initial partitioning of the data (*S*: 500, 1000, 1500, 2000 locations). First we divided the data into subsets so that no subset contained large temporal gaps (>10 d). We then divided the data subsets into the smallest possible temporal units that met a minimum data criterion, for which we used the window size *S*.

Segmentation

Step 1: Assessing habitat use with random forest models.—Prior to modeling habitat use, all windows of the respective data set were divided into a training and a test data set, each containing 50% of the data. We then built random forest models for all windows (simulated data: one model per window; empirical tracking data: 100 replicates per window) based on the training data sets (Breiman 2001, R-package *randomForest*, version 4.6.12). To prevent overfitting of the empirical tracking data models, we optimized the number of environmental variables as determined by fivefold cross-validations (random forest cross-validation for feature selection) on the corresponding test data.

Step 2: Transferability matrix.—We then assessed the transferability of the resulting random forest models, each based on a specific window, for (1) the test data of the corresponding window and (2) the test data of the other windows using the DI (see Section 1 in Appendix S1, Rubner et al. 2000, R-package *emdist*, version 0.3.1). In doing so, we assessed the ability of the window models to predict the habitat use of all (other) windows. For the empirical tracking data, we used multiple model replicates for the same subset. We first chose the best ten model replicates as determined by their DI on their respective test data. We then used these ten best models to calculate the DI on their respective test data and also for the test data of the other windows. The resulting DI measures were then averaged per window and used to derive the transferability matrix (see Fig. 1, step 2). For the simulated data, we used the DI as calculated from the single model replicates.

Step 3: Ordination and clustering of windows by transferability.—We ordinated the transferability matrix using a canonical correspondence analysis (CCA, Ter Braak 1986, R-package *vegan*, version

2.3.3) to make it suitable for clustering. For the empirical tracking data, this was applied with a constraining time distance matrix representing the temporal distance between the last location of window *k* and the first location of window *k* + 1 to correct for the temporal irregularities of the tracking data. Assuming that similar habitat use between windows would translate to high values of DI, we sought to detect coherent model ensembles using a fuzzy clustering algorithm (Kaufman and Rousseeuw 1990, R-package *cluster*, version 2.0.3). We replicated the clustering on the ordinated transferability matrix using $i = 2,…,n$, where n is the total number of locations divided by the number of windows, of desired clusters. Post hoc, we then determined the optimum number of clusters using the cluster silhouettes (Rousseeuw 1987, Kaufman and Rousseeuw 1990). Using the number of clusters *i* that produced the most compliant clustering, we merged the windows according to the clustering (termed niche segments) and annotated the complete data set accordingly.

Analysis of segmentation results

Segmentation of simulated data.—We evaluated the performance by testing the clustering derived from the segmentation against the known niche segments that were introduced during the data simulation using the Adjusted Rand Index (ARI, Rand 1971, Gordon 1999, R-package *fpc*, version 2.1.10). The ARI is a measure of similarity between two clustering alternatives that ranges from ARI = 0 (no better than random) to ARI = 1 (the clustering alternatives are identical). We calculated the ARI between the resulting segmentation and the simulated changes in the environment (for the scenarios of niche following and simultaneous change), as well as between the resulting segmentation and the simulated changes in habitat preference (for the scenarios of niche switching and simultaneous change). We tested the effects of changes in the different simulation parameters (Table 1) on the performance of the niche segmentation with a linear model, using the ARI as dependent variable. We standardized all numeric parameters using $\hat{x} = x - \bar{x}/\sigma_x$, where x is the actual observation, $\bar{x}$ the mean of all observations, and σ_x their standard deviation. For the window size *S* and the number of niche segments, we also included

quadratic and cubic transformations in the linear model as suggested by a preliminary generalized additive model (GAM). We determined the 95% confidence intervals on the mean ARI of all simulations using bootstrapping with 10,000 replicates.

Segmentation of the empirical tracking data set.—
Validation. We derived random forest models for all niche segments detected by the segmentation using the respective locations corresponding to the segments (=niche segment models). We randomly sampled 2000 locations as training data set from each niche segment and modeled habitat use as described above, repeating the process to obtain 100 model replicates. We calculated the DI of all model replicates on the data not used for training and kept only the ten models with the highest resulting DI. We compared the spatial predictions made from niche segment models, assuming that those niche segments with similar habitat use should result in similar spatial predictions for the same environmental conditions. We obtained the environmental conditions for every day of the study period with a resolution of 0.25 degrees while restricting the analysis to the spatial range of the species in Eurasia (BLI NS 2013). We used the niche segment models to predict habitat suitability for the complete study period and averaged the predictions over the 10 model replicates. We extracted the predictions for those Julian days corresponding to the respective niche segments and calculated the volumes of intersections for each unique combination of these daily predictions to generate an empirical distribution of intrasegment variation. We calculated the volumes of intersection as the sum of the absolute per-pixel difference between any two spatial predictions. To estimate the differences between niche segments, we calculated daily volumes of intersection using (1) the reference prediction from the respective niche segment model and (2) the prediction for the same day but derived from the other niche segment model. This resulted in two distributions: one representing intrasegment variation and one representing intersegment differences for each pairwise comparison of two segments, which allowed us to do two-way comparisons using Kolmogorov–Smirnov tests. Only those segment pairs for which Kolmogorov–Smirnov tests suggested significant differences between

intrasegment and intersegment variation in both directions were considered sufficiently different. Else, we combined the respective niche segment data sets. We then compared the validated segmentations for the different window sizes used.

Contributions of migration and study site to the segmentation. Variation in habitat use might not only arise through changes in the environment or habitat preference of individuals, but also from environmental differences between study sites and the migration of individuals between their breeding and wintering areas. To estimate how this contributed to the final segmentation, we computed a GAM for ordered categorical data (Wood 2006, R-package *mgcv*, version 1.8.6). We used the segmentation for the window size of $S = 2000$ locations as response variable, the catching site as a linear predictor, and two smoothing parameters: Julian day (using a cyclic cubic regression spline) as well as longitude and latitude (using a spherical spline).

Comparison with full models. In addition to the niche segment models, we also built models for the complete data set following the same procedure as previously described, only that we used data covering the complete study period for the training of random forest models (100 replicates with 2000 randomly sampled locations each). We calculated the DI of each model replicate for the complete data set and only kept the ten model replicates with the highest DI. Subsequently, we used the full model replicates to predict both the full and the segmented data sets and estimated their respective performances using DI. We also calculated the predictive performance of the niche segment model replicates on both the full and segmented data sets. We used *t* tests and adjusted the *P*-values for repeated testing using Bonferroni corrections. All analyses were conducted using the software R (R Development Core Team 2012). The code used for the simulation and segmentation of the simulated data is available in the Supporting Information (Data S1).

Results

Simulation experiments

The 95% confidence intervals of the mean overall ARI were (0.476; 0.484), while the modal ARI was 0.71 (Fig. 2). This suggests that our
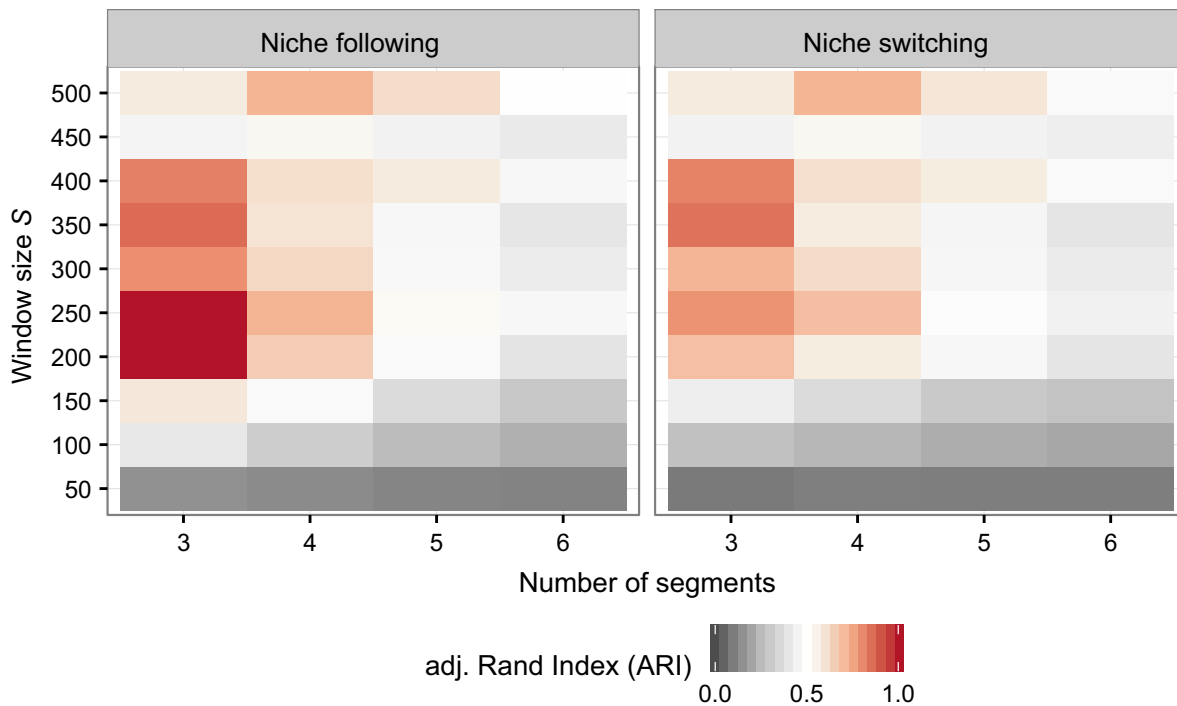
Fig. 2. Performance of the segmentation method according to the Adjusted Rand Index. The results for segmentation performance are shown separately for the niche following (change in the environment) and niche switching scenario (change in habitat preference). The results for the simultaneous change scenario are included within these results.

segmentation method performed better than random in all three scenarios (niche following, niche switching, simultaneous change). Overall, the performance was slightly higher when we calculated ARI based on the simulated environmental change than for changes in habitat preference. We found that segmentation performance also differed with both S and the true number of niche segments, as well as with the ratio between window size and the total number of locations (Table 3). Overall, we found that the performance tended to be higher for lower numbers of introduced segments and for small to medium window sizes.

Segmentation of waterbird tracking data

The validation process that we applied after the segmentation of the empirical tracking data suggested that all detected niche segments differed significantly in their spatial predictions of habitat suitability (Kolmogorov–Smirnov tests, $P < 0.01$ in all cases). Thus, the SDMs derived from the detected niche segments represented

different habitat use (see also Appendix S1: Fig. S2). Moreover, we found that different environmental variables were retained in the different niche segment models (Appendix S1: Fig. S3).

Table 3. Influences of the parameters on the method performance as evaluated by the ARI.

Coefficient	Estimate	SE	t-Value
Intercept	0.500	0.002	258.401
Changes in species preference	−0.044	0.003	−15.114
Window size (S)	0.0877	0.019	45.987
S^2	−1.542	0.043	−35.987
S^3	0.696	0.026	26.623
Number of niche segments (n)	0.352	0.140	2.511
n^2	−0.808	0.288	−2.802
n^3	0.419	0.150	2.794
Ratio $\frac{n}{S}$	0.094	0.005	20.031

Notes: ARI, Adjusted Rand Index.

Results from a linear model with ARI as response variable, and the number of true segments, the simulation scenario, number of data points, window size, and the ratio of window size/number of locations as predictors.
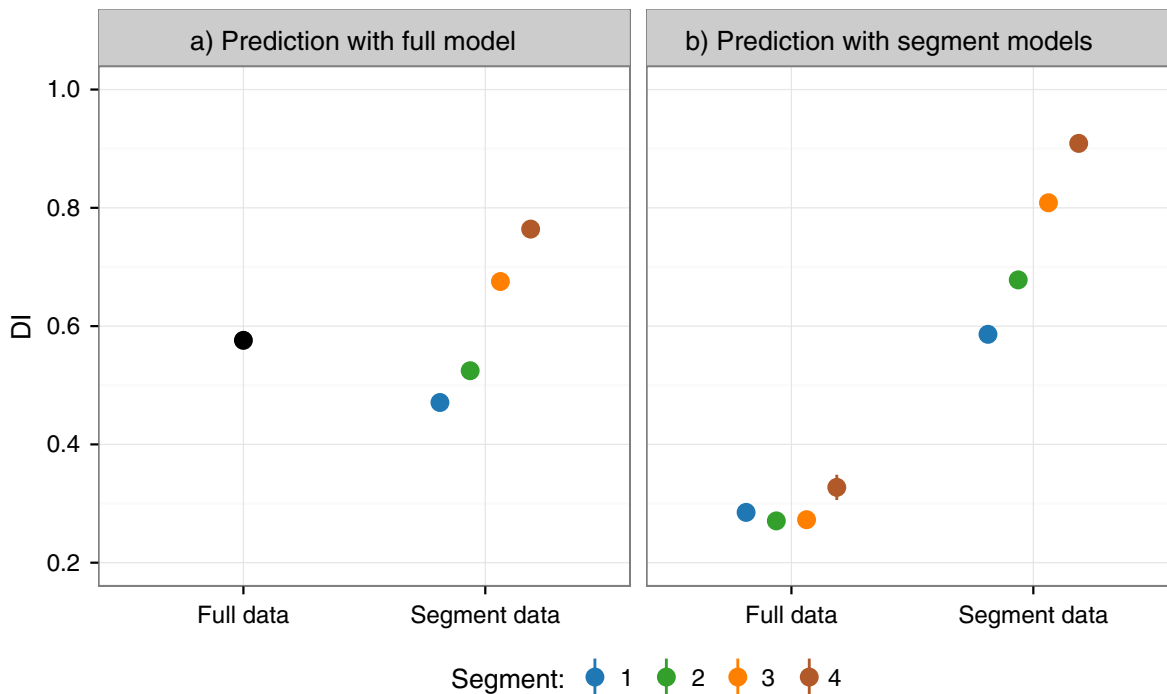
Fig. 3. Comparison between full and niche segment random forest models. We predicted both the full and niche segment data sets using the full model (a) and niche segment models (b), and estimated the models' ability to accurately predict the data with the Discriminatory Index. Shown here are the mean ± SD for the window size $S = 2000$.

When comparing the temporal structure of the segmentations resulting from different window sizes, we found that three niche segments were detected for the smaller three window sizes (S: 500, 1000, 1500 locations) and four for the largest window size (S: 2000 locations). The minimum temporal duration between change points of niche segments was naturally shorter for the smaller windows, and so the number of change points was lower for larger windows. However, we found that the timing of change points corresponded across window sizes (Appendix S1: Fig. S4) and also seemed to roughly match the timing of life-history stages of the species (Scott and Rose 1996, Kear 2005; see also Fig. 4). Migratory populations tend to arrive at their breeding grounds from late February onwards, corresponding to the first set of change points (beginning of March to mid-March for all S). Common teal usually start breeding in May, where we found a second change point in the largest window size. With the start of incubation in June, males usually migrate to molt at suitable sites, whereas the females stay at the breeding grounds until fall migration. For all window sizes, we found change points around the beginning of June as well as in mid-June, which is followed by a period of consistent habitat use until November, when the birds usually arrive at the wintering grounds. The last change point occurred in the beginning of December.

We tested how the use of individuals caught at different study sites as well as the migration of individuals might have influenced the segmentation using a GAM for ordered categorical data. The results suggest that the segmentation could be best explained by Julian day (Appendix S1: Table S2). Study site except for Kazakhstan, and the spatial position of locations used for the window models did only have minor effects on the temporal sequence of niche segments. When comparing the predictive performance of SDMs that were derived from (1) the segmented data and (2) the full data set, we found that niche segment models outperformed the full models on the respective niche segment data sets (two-sample t tests, $P < 0.01$ in all cases, Fig. 3). Niche

segment models were, however, outperformed by the full models when applied to the full data set. This pattern was observed for the segmentations based on all the window sizes used.

Discussion

Under the assumption that changes in observed habitat use are indicative of behavioral changes, they can be used for a segmentation of animal movement data, similar to a segmentation by geometric features of a trajectory (Gurarie et al. 2016). In the simulated data, the clustering of windows based on a single measure of transferability allowed us to successfully detect the inbuilt temporal structure. When we applied the niche segmentation to an empirical data set of common teal, change points roughly corresponded to the species' life history, and the results from the segmentation were robust against changes in window size. We thus think that the niche segmentation can detect relevant changes in habitat use across multiple scales. Overall, the niche segmentation performed better than random on the simulated data. Several of the simulation parameters, however, showed an effect on the success of the segmentation as indicated by the ARI. The most crucial step in the segmentation was the choice of the window size S, which affects the resolution in determining changes in habitat use as well as the number of locations available to the window-specific models. We thus think that the overall higher values of ARI for smaller window sizes were likely due to the increased temporal resolution, while smaller window sizes inherently lead to less generalizable models in terms of habitat use. Whereas it has been shown that SDMs for specialized species can provide sensible predictions of species occurrence using just a few presence locations (Pearson et al. 2007), this does not apply to non-independent animal observations from tracking data, especially for a limited number of individuals. But despite this trade-off, the simulations showed that the niche segmentation can successfully detect changes in habitat use.

Using temporally explicit models of the space use of individuals and a temporally weighted partitioning of data into windows, we could apply the segmentation to an empirical tracking data set of the common teal. In contrast to simulated data,

however, empirical movement data are often characterized by sampling irregularity and location errors. So while we could show that the segmentation worked well in the case of regular data with exact positions, it was not clear whether this ability of the method transfers to empirical movement data. We incorporated the temporal structure of the irregular tracking data using a constraining time distance matrix during the ordination of the transferability matrix. We also think that smaller temporal irregularities will not greatly affect the outcome of the segmentation, because the partitioning of the data into windows will inherently result in a fuzzy segmentation. Much of the temporal irregularity of empirical data can be redeemed by how the data are partitioned into windows, especially when the temporal scale of interest is greater than the scale of temporal inaccuracies. The impact of location error, however, we deem to be rather small for mainly two reasons: (1) In many cases, the spatial resolution of the environmental data is lower than the spatial resolution of movement data and will become more so over time as the use of highly accurate GPS devices increases, and (2) the spatial autocorrelation of the environment, that is, neighboring pixels of environmental data have highly correlated values. In our case study on the common teal, however, the location error was highly variable across the trajectories which could have introduced a bias in the accuracy of the environmental information corresponding to the locations over time. With the use of time explicit, we could reduce this bias by sampling pseudopresence points that reflect the environmental conditions within the 99% UDs rather than just at a single erroneous location.

The detected change points of the niche segmentation bear similarities to the life-history cycle of the species (Fig. 4). What is also important to note is that the predictions from niche segment models also differed considerably in the spatial distribution of suitable habitat (Appendix S1: Fig. S2), corresponding to the changes in the spatial distribution of the species in the wild. Finally, niche segment models outperformed full models when predicting the respective niche segment despite using the same amount of samples for training the models, which indicates that the changing interactions between individuals and environment over time could not be completely captured by the full models.
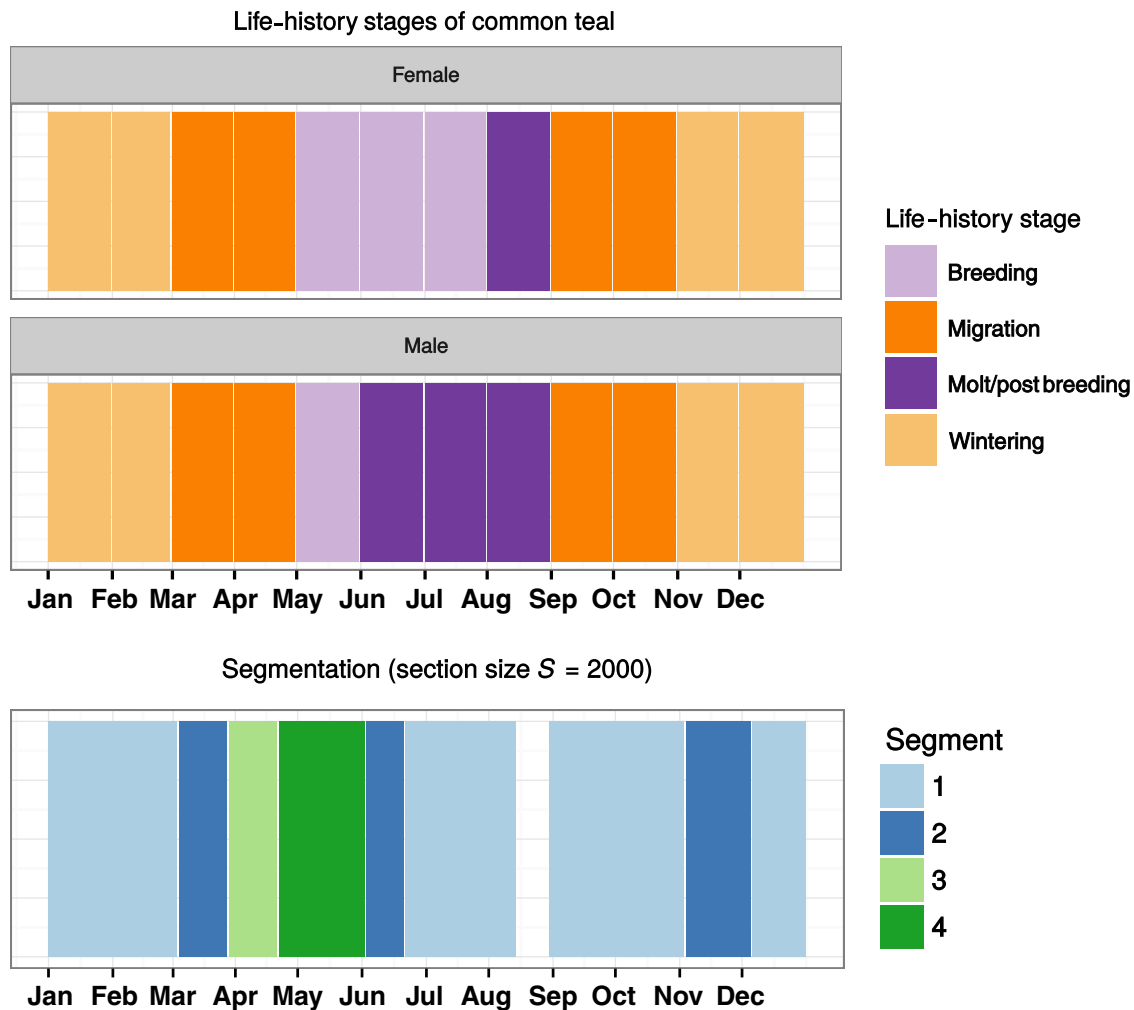
Fig. 4.   Comparison of the life-history stages of common teal for both males and females (Scott and Rose 1996, Kear 2005) with the segmentation derived with a window size of $S = 2000$. The coloring is kept differently to not imply any untested similarities.

Although changes in habitat use are likely to be driving the segmentation pattern, there are alternative explanations that could result in the segmentation of the data into multiple niche segments. One of them is the use of location data from several different populations. Habitat composition at these study sites might be so different that changes in habitat use could more easily be explained by catching site, especially because individuals at the different study sites were caught during different times of the year (Table 2). This can result in an unbalanced distribution of location data from the different populations across the year. However, we found that the series of niche segments could not be explained by study site, and only the individuals from Kazakhstan seemed to use different habitats than the individuals from the other populations. Consequently, the niche segmentation was likely to reflect consistent differences in habitat use of common teal throughout the year.

In general, there is a growing tendency to use the framework of SDMs for fine-scaled analyses of variation of habitat use to illuminate the dynamic interface between individuals and their environment (e.g., van Toor et al. 2011, Pikesley et al. 2015), as well as an increasing availability of long-term and high-resolution animal

movement data sets. The fusion of both provides a promising approach to extract an environmentally informed behavioral signature from the trajectories of single individuals as well as groups of animals. As movement and habitat use are an expression of individual decisions, because individuals are driven by physiological necessities like energy requirements and optimal breeding sites (e.g., Boone et al. 2006, Nathan et al. 2008, Suárez-Seoane et al. 2008, Bischof et al. 2012), the behavioral annotation of trajectories will allow us to derive ecologically meaningful models of animal movement. In addition to increasing the understanding of animal movement, SDMs used for the spatial prediction of habitat suitability on larger scales can profit from the incorporation of changes in both habitat preferences and the environment. This might especially apply to migratory species which are currently underrepresented in studies applying SDMs (Web of Science search, 12 March 2015, 358 and 15,698 publications using the keywords "species distribution model" with and without "migrat*"). This is of special concern, as the areas used by migratory species are currently underrepresented by protected areas (Runge et al. 2014). SDMs could therefore incorporate the underlying temporal dynamics of habitat use and habitat availability by applying a niche segmentation and deriving separate models for segments. This might help to improve predictions of habitat suitability for these cases or for species that require different types of habitat or resources during their life cycle (e.g., Werner and Gilliam 1984, Wilbur 1988, Hatase et al. 2002). Thus, it could be possible to identify restrictions of the fluctuations in the availability of suitable habitat, and how suitable areas might be connected in space and time. Overall, using a niche-based segmentation can contribute to the understanding of animal behavior from using remotely tracked animal movement data. The niche segmentation will determine change points in habitat use rather fuzzily and with only approximate times, owed to the fact that they are found by comparing windows of a discrete size and thus sudden changes can only be found if they coincide with falling at the very end or start of a window. Yet, the fuzziness in finding changes also can be an indication of gradual changes involved in species–environment interactions. A segmentation based on habitat use adds a component of information to models that cannot necessarily be covered by a segmentation based on movement geometry alone. By integrating information like age or sex, smaller and more subtle changes in habitat use will potentially become traceable. Overall, a niche segmentation as presented here can not only contribute to the understanding of processes that are mediated by how individuals interact with a changing environment, but also provide new opportunities to integrate changing landscapes into the study of the spatial dimension of animal behavior.

Acknowledgments

Literature Cited

Batalden, R. V., K. Oberhauser, and A. T. Peterson. 2007. Ecological niches in sequential generations of eastern North American Monarch butterflies (Lepidoptera: Danaidae): the ecology of migration and likely climate change implications. Environmental Entomology 36:1365–1373.

BirdLife International and NatureServe (BLI NS). 2013. Bird species distribution maps of the world. Version 3.0. http://www.birdlife.org/datazone/info/spcdownload

Bischof, R., L. E. Loe, E. L. Meisingset, B. Zimmermann, B. V. Moorter, and A. Mysterud. 2012. A migratory northern ungulate in the pursuit of spring: Jumping or surfing the green wave? American Naturalist 180:407–424.

Boone, R. B., S. J. Thirgood, and J. G. C. Hopcraft. 2006. Serengeti wildebeest migratory patterns modeled from rainfall and new vegetation growth. Ecology 87:1987–1994.

Breiman, L. 2001. Random forests. Machine Learning 45:5–32.

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

Codling, E. A., M. J. Plank, and S. Benhamou. 2008. Random walk models in biology. Journal of the Royal Society Interface 5:813–834.

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

Douglas, D. C., R. Weinzierl, S. C. Davidson, R. Kays, M. Wikelski, and G. Bohrer. 2012. Moderating Argos location errors in animal tracking data. Methods in Ecology and Evolution 3:999–1007.

Ficetola, G. F., R. Pennati, and R. Manenti. 2012. Spatial segregation among age classes in cave salamanders: Habitat selection or social interactions? Population Ecology 55:217–226.

Garriga, J., J. R. B. Palmer, A. Oltra, and F. Bartumeus. 2016. Expectation-maximization binary clustering for behavioural annotation. PLoS ONE 11:e0151984.

Gordon, A. 1999. Classification. Chapman and Hall/CRC, Boca Raton, Florida, USA.

Grinnell, J. 1917. The niche-relationships of the California thrasher. Auk 34:427–433.

Gurarie, E., R. D. Andrews, and K. L. Laidre. 2009. A novel method for identifying behavioural changes in animal movement data. Ecology Letters 12:395–408.

Gurarie, E., C. Bracis, M. Delgado, T. D. Meckley, I. Kojola, and C. M. Wagner. 2016. What is the animal doing? Tools for exploring behavioural structure in animal movements. Journal of Animal Ecology 85:69–84.

Hatase, H., N. Takai, Y. Matsuzawa, W. Sakamoto, K. Omuta, K. Goto, N. Arai, and T. Fujiwara. 2002. Size-related differences in feeding habitat use of adult female loggerhead turtles *Caretta caretta* around Japan determined by stable isotope analyses and satellite telemetry. Marine Ecology. Progress Series 233:273–281.

Hussey, N. E., et al. 2015. Aquatic animal telemetry: a panoramic window into the underwater world. Science 348:1255642.

Hutchinson, G. E. 1957. Population studies – animal ecology and demography – concluding remarks. Cold Spring Harbor Symposia on Quantitative Biology 22:415–427.

Kaufman, L., and P. J. Rousseeuw. 1990. Finding groups in data: an introduction to cluster analysis. John Wiley & Sons, New York, New York, USA.

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

Kear, J. 2005. Ducks geese and swans: general chapters. Special accounts (Anhima to Salvadorina). Volume 1. Oxford University Press, Oxford, UK.

Kranstauber, B., R. Kays, S. D. LaPoint, M. Wikelski, and K. Safi. 2012. A dynamic Brownian bridge movement model to estimate utilization distributions for heterogeneous animal movement. Journal of Animal Ecology 81:738–746.

Martínez-Meyer, E., A. T. Peterson, and A. G. Navarro-Sigüenza. 2004. Evolution of seasonal ecological niches in the *Passerina* buntings (Aves: Cardinalidae). Proceedings of the Royal Society of London. Series B: Biological Sciences 271: 1151–1157.

Nakazawa, Y., A. T. Peterson, E. Martínez-Meyer, and A. G. Navarro-Sigüenza. 2004. Seasonal niches of narctic-neotropical migratory birds: implications for the evolution of migration. Auk 121:610–618.

Nathan, R., W. M. Getz, E. Revilla, M. Holyoak, R. Kadmon, D. Saltz, and P. E. Smouse. 2008. A movement ecology paradigm for unifying organismal movement research. Proceedings of the National Academy of Sciences of the United States of America 105:19052–19059.

Pearson, R. G., C. J. Raxworthy, M. Nakamura, and A. Townsend Peterson. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. Journal of Biogeography 34: 102–117.

Pikesley, S. K., et al. 2015. Modelling the niche for a marine vertebrate: a case study incorporating behavioural plasticity, proximate threats and climate change. Ecography 38:803–812.

R Development Core Team. 2012. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rand, W. M. 1971. Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association 66:846–850.

Rousseeuw, P. J. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20:53–65.

Rubner, Y., C. Tomasi, and L. J. Guibas. 2000. The earth mover's distance as a metric for image retrieval. International Journal of Computer Vision 40:99–121.

Runge, C. A., T. G. Martin, H. P. Possingham, S. G. Willis, and R. A. Fuller. 2014. Conserving mobile species. Frontiers in Ecology and the Environment 12:395–402.

Santos, M. J., and M. Santos-Reis. 2009. Stone marten (*Martes foina*) habitat in a Mediterranean ecosystem: effects of scale, sex, and interspecific interactions. European Journal of Wildlife Research 56:275–286.

Sattler, T., F. Bontadina, A. H. Hirzel, and R. Arlettaz. 2007. Ecological niche modelling of two cryptic bat species calls for a reassessment of their conservation status. Journal of Applied Ecology 44:1188–1199.

Schlather, M., A. Malinowski, P. J. Menck, M. Oesting, and K. Strokorb. 2015. Analysis, simulation and prediction of multivariate random fields with package RandomFields. Journal of Statistical Software 63:1–25.

Scott, D. A., and P. M. Rose. 1996. Atlas of Anatidae populations in Africa and Western Eurasia | Clc. Wetlands International Publication, Wetlands International, Wageningen, The Netherlands.

Soberón, J. M. 2007. Grinnellian and Eltonian niches and geographic distributions of species. Ecology Letters 10:1115–1123.

Suárez-Seoane, S., E. L. García de la Morena, M. B. Morales Prieto, P. E. Osborne, and E. de Juana. 2008. Maximum entropy niche-based modelling of seasonal changes in little bustard (*Tetrax tetrax*) distribution. Ecological Modelling 219:17–29.

Ter Braak, C. J. F. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology 67:1167–1179.

van Toor, M. L., C. Jaberg, and K. Safi. 2011. Integrating sex-specific habitat use for conservation using habitat suitability models. Animal Conservation 14:512–520.

Václavík, T., J. A. Kupfer, and R. K. Meentemeyer. 2012. Accounting for multi-scale spatial autocorrelation improves performance of invasive species distribution modelling (iSDM). Journal of Biogeography 39:42–55.

Werner, E. E., and J. F. Gilliam. 1984. The ontogenetic niche and species interactions in size-structured populations. Annual Review of Ecology and Systematics 15:393–425.

Wilbur, H. M. 1988. Interactions between growing predators and growing prey. Pages 157–172 *in* B. Ebenman and L. Persson, editors. Size-structured populations. Springer, Berlin, Germany.

Wood, S.. 2006. Generalized additive models: an introduction with R. Chapman & Hall/CRC, Boca Raton, Florida, USA.

SUPPORTING INFORMATION

Additional Supporting Information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/ecs2.1498/full