Lowe *et al. Anim Biotelemetry* (2020) 8:10
https://doi.org/10.1186/s40317-020-00195-y

Animal Biotelemetry

RESEARCH

Open Access

Detecting commonality in multidimensional fish movement histories using sequence analysis

Michael R. Lowe[1*], Christopher M. Holbrook[1] and Darryl W. Hondorp[2]

Abstract

Background: Acoustic telemetry, for tracking fish movement histories, is multidimensional capturing both spatial and temporal domains. Oftentimes, analyses of such data are limited to a single domain, one domain nested within the other, or ad hoc approaches that simultaneously consider both domains. Sequence analysis, on the other hand, offers a repeatable statistical framework that uses a sequence alignment algorithm to calculate pairwise dissimilarities among individual movement histories and then hierarchical agglomerative clustering to identify groups of fish with similar movement histories. The objective of this paper is to explore how acoustic telemetry data can be fit to this statistical framework and used to identify commonalities in the movement histories of acoustic-tagged sea lamprey during upstream migration through the St. Clair-Detroit River System.

Results: Five significant clusters were identified among individual fish. Clusters represented differences in timing of movements (short vs long duration in the Detroit R. and Lake St. Clair); extent of upstream migration (ceased migration in Lake St. Clair, lower St. Clair R., or upper St. Clair R.), and occurrence of fallback (return to Lake St. Clair after ceasing migration in the St. Clair R.). Inferences about sea lamprey distribution and behavior from these results were similar to those reached in a previous analysis using ad-hoc analysis methods.

Conclusions: The repeatable statistical framework outlined here can be used to group sea lamprey movement histories based on shared sequence characteristics (i.e., chronological order of "states" occupied). Further, this framework is flexible and allows researchers to define a priori the movement aspect (e.g., order, timing, duration) that is important for identifying both common or previously undetected movement histories. As such, we do not view sequence analysis as a panacea but as a useful complement to other modelling approaches (i.e., exploratory tool for informing hypothesis development) or a stand-alone semi-quantitative method for generating a simplified, temporally and spatially structured view of complex acoustic telemetry data and hypothesis testing when observed patterns warrant further investigation.

Keywords: Acoustic telemetry, Fish movement, Hierarchical clustering, Optimal matching

Background

Understanding fish movements is integral to fisheries management [1], species and habitat conservation [2–5], and mitigating the impacts of invasive species [6, 7]. In recent decades, passive acoustic telemetry in the aquatic environment (hereafter, 'acoustic telemetry') has become the principal tool for monitoring fish movements [8, 9] and allows for detailed insight into fish migration

*Correspondence: mlowe@usgs.gov
[1] Hammond Bay Biological Station, Great Lakes Science Center, United States Geological Survey, 11188 Ray Rd., Millersburg, MI 49759, USA
Full list of author information is available at the end of the article

route selection and timing [10–12], spawning behavior [13], factors that affect population demographics, and interactions with other species [14] and their environment [15, 16]. While objectives and goals vary amongst projects, acoustic telemetry research generally involves (1) attachment of an electronic tag that emits a unique identification code to individual fish, (2) using an array of stationary or mobile receivers to detect telemetered fish as they move through areas or regions of interest, (3) generating a set of geo-referenced, time-stamped detection records of each fish at each receiver location, and (4) interpretation of detection data to make inferences about individual fish movements and habitat use patterns at the individual and population levels [1, 17]. More recently, telemetry has been used to identify geographical organization and spatial structure in fish migration patterns at the population level that may be significant to conservation or management efforts [2, 3].

Acoustic telemetry data are time-indexed records of fish location, but statistical methods that enable joint consideration of temporal and spatial domains are rarely used in the analysis of acoustic telemetry detection. In the aquatic environment, this challenge is exacerbated by incomplete or patchy spatial coverage and variation in space use and movement timing among individuals. For those reasons, fish movements are often displayed graphically in both domains but analyzed independently [3, 12, 18] or with one domain nested within the other [19]. Ad hoc approaches (e.g., analyses developed for a specific task) are frequently used to simultaneously incorporate space and time in models of individual movements [20, 21], and survival [4], but those methods are not easily repeated and have not been used to identify movement structure at the population level. Few studies have used cluster analysis to identify structure (i.e., commonalities) within populations [2, 22]. For example, Kessel et al. used supervised agglomerative clustering (i.e., detection histories were manually arranged along a similarity gradient) to identify migratory contingents within a population of lake sturgeon (*Acipenser fulvescens*). Their classification was based on an intuitively derived dissimilarity metric and showed that the St. Clair Detroit River System (SCDRS) lake sturgeon population contained multiple divergent migration behaviors. We outline a statistical framework that expands on those approaches by using sequence analysis [23] and cluster analysis to identify common movement structures among a group of acoustic-tagged fish (i.e., population structure).

Sequence analysis is a reproducible statistical framework for identifying patterns in temporally and spatially ordered lists of objects (e.g., amino acids), states (e.g., employed vs unemployed), or events (e.g., marriage, divorce, childbirth). Originally used for sequence matching in bioinformatics in the 1970s [24] and further developed for studying life course trajectories in the social sciences [23], sequence analysis is a multivariate statistical approach that uses (1) a suite of well-studied metrics for estimating dissimilarity between every pair of ordered lists/sequences and (2) statistical separation into groups of common membership using multivariate tools (e.g., discriminant analysis, cluster analysis and multidimensional scaling). Though sequence analysis methods have been used to address spatial questions [25], they have only recently been applied to animal movement [26] or fish telemetry data [27]. Given the ability to simultaneously consider multidimensional information contained across the entire movement history, sequence analysis appears to be a viable approach for identifying common or previously undetected movement structures at the population level and providing an improved understanding of important aspects of fish movement ecology.

The goal of this paper is to evaluate the combined use of sequence and cluster analysis to identify common movement structures among individual fish movement histories. We outline a multistep process that first converts detections for each individual fish into temporally ordered movement histories (Fig. 1a) that contain both the spatial and temporal aspects of the original data. We specifically examine the impact of (1) the temporal resolution of the input data on movement sequence interpretation and (2) the cost structure used to calculate the distance measures (i.e., how the metric for determining the difference between two sequences is calculated). Lastly, a statistical framework for clustering movement histories among acoustic-tagged sea lamprey (*Petromyzon marinus*) is presented (Fig. 1b).

Sea lamprey are invasive in the Laurentian Great Lakes and have been the subject of a bi-national, basin-wide population control program since the 1950s [26, 27]. The control program has focused largely on reproductive aspects of sea lamprey biology and, as such, the spawning behaviors of adult sea lamprey are well documented in the Great Lakes [28]. Following an extended parasitic phase, adult sea lamprey detach from their host and migrate, sometimes 100 s of kilometers, to spawning tributaries during the spring [29]; though there is no evidence of population-level natal philopatry. Peak spawning occurs when water temperatures reach 17.0–19.0 C [30, 31]. During the spawning cycle, sea lamprey stop feeding, their internal organs degenerate [32] and, as a result, both spawning and non-spawning adults die.

Despite extensive monitoring and control efforts throughout the Great Lakes, sea lamprey abundance in Lake Erie has remained above targets set by fishery managers. It was hypothesized that unrecognized recruitment in the SCDRS was responsible for recent increases

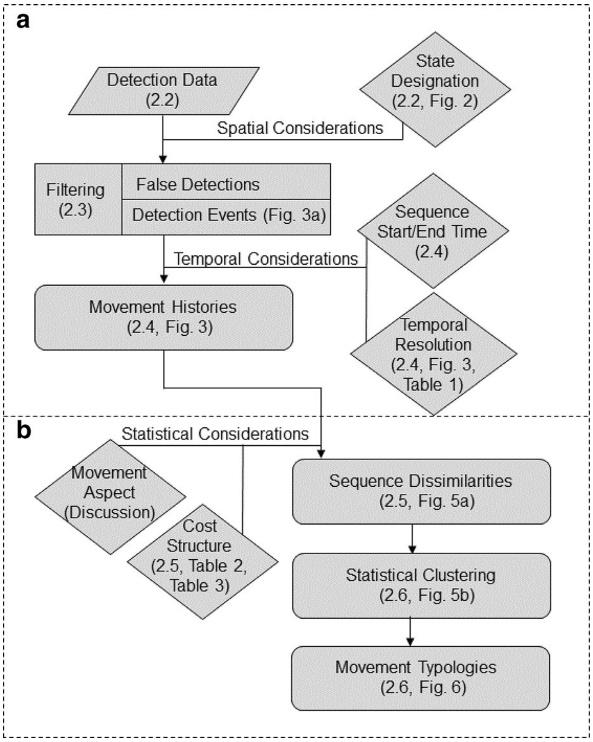

Fig. 1 Generalized workflow. Workflow outlining the process of identifying commonalities in fish movement histories from acoustic telemetry data. The diagram is separated into two main components; data processing (**a**) and statistical framework (**b**). Trapezoids represent data inputs, rectangles are processes or filters, diamonds are user decisions/considerations, and rounded rectangles are derived outputs. The corresponding section in manuscript (e.g., 2.2), figures, and appendices (e.g., S1) are shown within each structure. 'Movement aspect' is discussed at length in the penultimate paragraph in the discussion

in sea lamprey abundance throughout Lake Erie. However, monitoring and control efforts in the SCDRS have been complicated by a lack of barriers to migration and a discharge that exceeds other rivers in the region by an order of magnitude; both factors effect trapping efficiency for sea lamprey assessment and monitoring. In order to better understand the movement ecology of invasive sea lamprey, improve estimates of population size for control purposes, and identify novel spawning habitats in the SCDRS 27 acoustic-tagged adult sea lamprey were released in the lower Detroit River (Fig. 2) during the spring of 2014. Individual movements were recorded throughout the SCDRS using an array of acoustic receivers. An ad hoc model, that assumed the final spawning locations approximated a multinomial process, was used to conclude that spawning most likely occurred in the St. Clair River [33]. That study also elucidated a "fallback" behavior (i.e., movement downstream after cessation of upstream migration) in 10 individuals that coincided with water temperatures commensurate with peak spawning activity [34] and viewed as evidence that a spawning event had occurred. Those same data are used in this paper with the explicit goal of assessing the applicability of sequence analysis methods to fish movement histories. As such, our purpose is not to revisit those 27 adult sea lamprey movement histories within a different analytical framework in search of new ecological insights but rather use those data to provide a contextual comparison; if sequence analysis is to be considered a viable tool for analyzing acoustic telemetry data then the method should, at a minimum, provide results that recapitulate those of Holbrook et al. [33].

Methods

Study system

The SCDRS is a 150 km long river corridor that contains (from upstream to downstream) the St. Clair River, Lake St. Clair, and the Detroit River and connects southern Lake Huron with the western basin of Lake Erie. Discharge averages 5200 m^3 s^{-1} [35, 36], is seasonally consistent, and mostly derived from Lake Huron [37]. The waters of the SCDRS are oligotrophic with temperatures ranging from < 2 C in the winter to 19–25 C in July [38].

Fish movement data

Twelve female and 15 male adult, spawning condition sea lamprey (43.0–58.0 cm total length) were collected from the Grand River, Ohio between 11 and 13 May 2014 and surgically implanted with acoustic transmitters (model V8-4H, Vemco; Halifax, Nova Scotia, Canada) before being released in the lower Detroit River (Fig. 2) on 16 May 2014 at 1343 GMT. Transmitters had an expected tag life of 112 days and were transmitting through the end of August. Each transmitter emitted a burst of coded acoustic pulses every 60–180 s (120 s nominal delay) and timestamped detections (i.e., when an acoustic pulse was detected) were recorded as individual fish moved through an acoustic telemetry array that consisted of 72 receivers (model VR2W; Vemco) distributed among 12 locations within the SCDRS (Fig. 2). Additional detections from 462 receivers located outside of the SCDRS were accessed via the Great Lakes Acoustic Telemetry Observation System (https:\\glatos.glos.us). Each receiver location was assigned to one of seven discrete spatial units (hereafter 'states') in the SCDRS from downstream to upstream (Fig. 2); Lake Erie (included all receivers downstream of the SCDRS), lower Detroit River, upper Detroit River, Lake St. Clair, lower St. Clair River, upper St. Clair River, and Lake Huron (included all receivers upstream of the SCDRS). Five tributaries within the SCDRS (e.g., Belle, Black, Clinton, Pine, and Thames Rivers; Fig. 2),

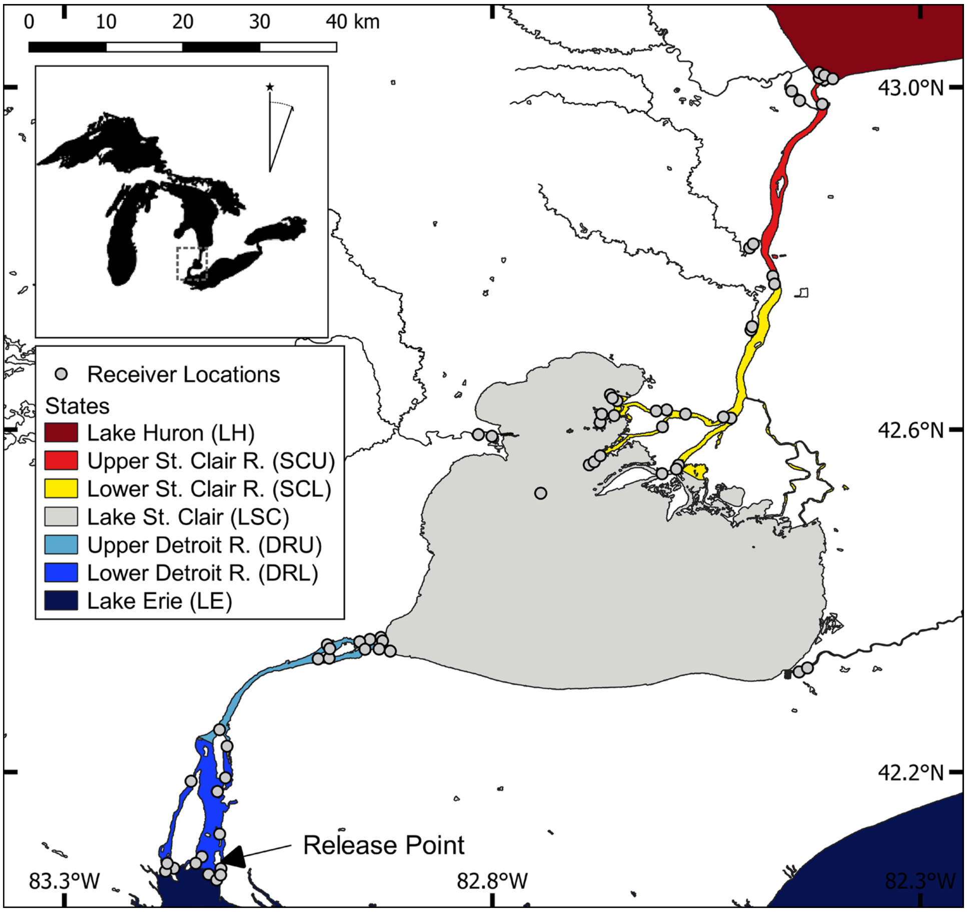

Fig. 2 Map of study area. St. Clair River Detroit River system with the acoustic telemetry receiver configuration, release location, and associated river states. Inset map indicates the location of the St. Clair River Detroit River System within the Great Lakes basin

which also contained receivers, were assigned to the state that contained the tributary mouth for each. However, only one fish moved into a tributary during the study [33].

Filtering detection data

Potentially false detections that resulted from signal code collisions [39, 40] were filtered from the dataset by omitting all detections that were not within 3600 s (i.e., 30 times the nominal delay) of another detection of the same tag code on the same receiver [41]. False detections can occur when two or more fish pass within the detection range of the same receiver and their acoustic tags transmit at the same time (i.e., tag collisions) and the receiver deciphers a "false" code instead of the two codes that collided. Such events depend on the number of tagged fish within detection range of a receiver and are generally rare; of the 7005 total detections in our study, only 101 (1.4%) were identified as potential false detections. Filtered detection data were further distilled into detection events representing time intervals in which each fish occupied each state. Each detection event was comprised of only the first and last detections of an uninterrupted series of detections for each fish within a state. In this case, an interruption only occurred when an individual was detected in a different state. Thus, events were separated by periods of transition between states when the state was not known.

Converting detection data to movement history sequences

Filtered detection events were converted to movement history sequences containing the state (e.g., Lower Detroit R., Lake St. Clair, etc.; see *2.2 Fish Movement Data* for list of possible states) of each fish in each 1-h time interval throughout the study period. Each sequence started on 16 May at 1300 GMT coincident with the release of acoustic-tagged fish into the lower Detroit River and ended 1 July 1300 GMT. The 1 July cutoff for all movement histories was based on the observed final detection events for all fish that ranged from 22 May to 30 June with 20 of the 27 final detections occurring after 15 June. Each fish was assigned to a dominant state occupied during that time interval based on the proportion of time spent in each state (i.e., state-specific residence time/total time for that interval). This was necessary when an individual transitioned from one state to another. During periods when fish were not detected (i.e., in portions of the SCDRS between states not covered by receivers), that last state occupied was carried forward until the next detection and transitions into a new state were never imputed. Movement history sequences were stored in a matrix containing the chronologically ordered state occupation for each individual fish (i.e., one row for each fish and one column for each time interval).

The time resolution used in movement history sequences is a critical decision because overly coarse resolutions can mask ecologically significant state changes and overly fine resolutions add unnecessary computational and interpretational complexity. The time resolution used in this analysis (1 h) was determined by comparing sequences constructed at 1-, 6-, 12-, 24-, and 96-h intervals to identify the temporal resolution that best preserved the multidimensional information contained in the filtered detection events. This process resulted in five movement history matrices from 27 × 1104, 27 × 184, 27 × 92, 27 × 46, and 27 × 12 for the 1-, 6-, 12-, 24-, and 96-h intervals, respectively (Fig. 3). Resulting movement histories showed marked differences in the range of habitats occupied by individual sea lamprey. Though the St. Clair River was a prominent feature at all resolutions, the 1-h resolution showed the greatest diversity in movement histories (Fig. 3b) and there was less apparent information at the coarsest temporal resolution (Fig. 3f). These results were corroborated by hierarchical agglomerative clustering which was used to evaluate the degree of similarity among the five movement sequences for each fish, individually (Additional File 1; Fig. 1). Further, the proportion of movement histories that required imputation ranged from 8 to 88% with higher resolution data (i.e., 1 h intervals) requiring more imputation than coarser resolutions (Table 1). However, a multisample equality of proportions test, with continuity correction, indicated that the mean proportion of imputed movement histories did not differ among the five time intervals ($\chi^2 = 1.107$, $df = 4$, $p = 0.89$). As a result, all analyses are based on the sea lamprey movement histories constructed at 1-h intervals (Fig. 3b).

Calculating dissimilarity

Sequence analysis methods are predicated on quantifying the extent to which each pair of movement histories are dissimilar. Dissimilarity, as defined in this paper, is the "cost" needed to convert one movement history sequence into another movement history sequence (i.e., edit distance). Conversion can be accomplished through two operations using optimal matching (OM) within the edit distance framework: substitutions (i.e., changing the observed state in one sequence to match the observed state at the same position in the other sequence) and insertion-deletion (indel; i.e., inserting a new observation into one sequence or deleting an observation from the other sequence). There are numerous cost regimes under the umbrella of the edit distance framework that differ by the way in which substitution and indel costs are calculated, but generally, for a given cost structure (i.e., dissimilarity measure) an algorithm is used to identify the lowest-cost set of operations needed to produce a match from two sequences. Studer and Ritschard [42] provide an extensive review of the most commonly used cost regimes for calculating dissimilarity measures (including Euclidean and Chi squared distances) and the scenarios in which each approach is best suited. For this analysis, we sought a cost regime that met triangle inequality (i.e., ensured coherence between computed dissimilarities) and reflected ecological reality (i.e., did not allow 2nd order or higher movements (skipped states) and did not allow changes to the length of the sequences).

The cost regime used to calculate dissimilarities among movement history sequences in this analysis (custom cost regime, described below) was selected among five candidate cost regimes (Table 2):: (1) substitutions and indel operations had the same cost (i.e., Levenshtein distance), (2) only indel operations were allowed (i.e., Levenshtein II distance), (3) only substitutions were allowed (i.e., Hamming distance), (4) a data driven cost regime, and (5) a custom cost regime based on state attributes (i.e., connectivity). All of the cost regimes are variations of the 'optimal matching' method in the 'seqdist' function of the R package "TraMineR". For Levenshtein distances, the costs of substitutions and indels were both equal to one. Thus, Levenshtein distances were equivalent to the minimum number of operations required to transform one sequence into another. For Levenshtein II distances, substitutions were effectively disallowed by setting the cost of each substitution eight times larger than the cost of an indel. Similarly, for Hamming distances, indels were effectively disallowed by setting the cost of each indel three times larger than the cost of a substitution. The data driven cost regime was based on the observed probabilities of all sea lamprey transitioning from one state to another (i.e., transition rates; Table 3a). Data driven substitution costs (SC) were calculated as follows

$$SC_{(i,j)} = cval - TR_{(i,j)}$$

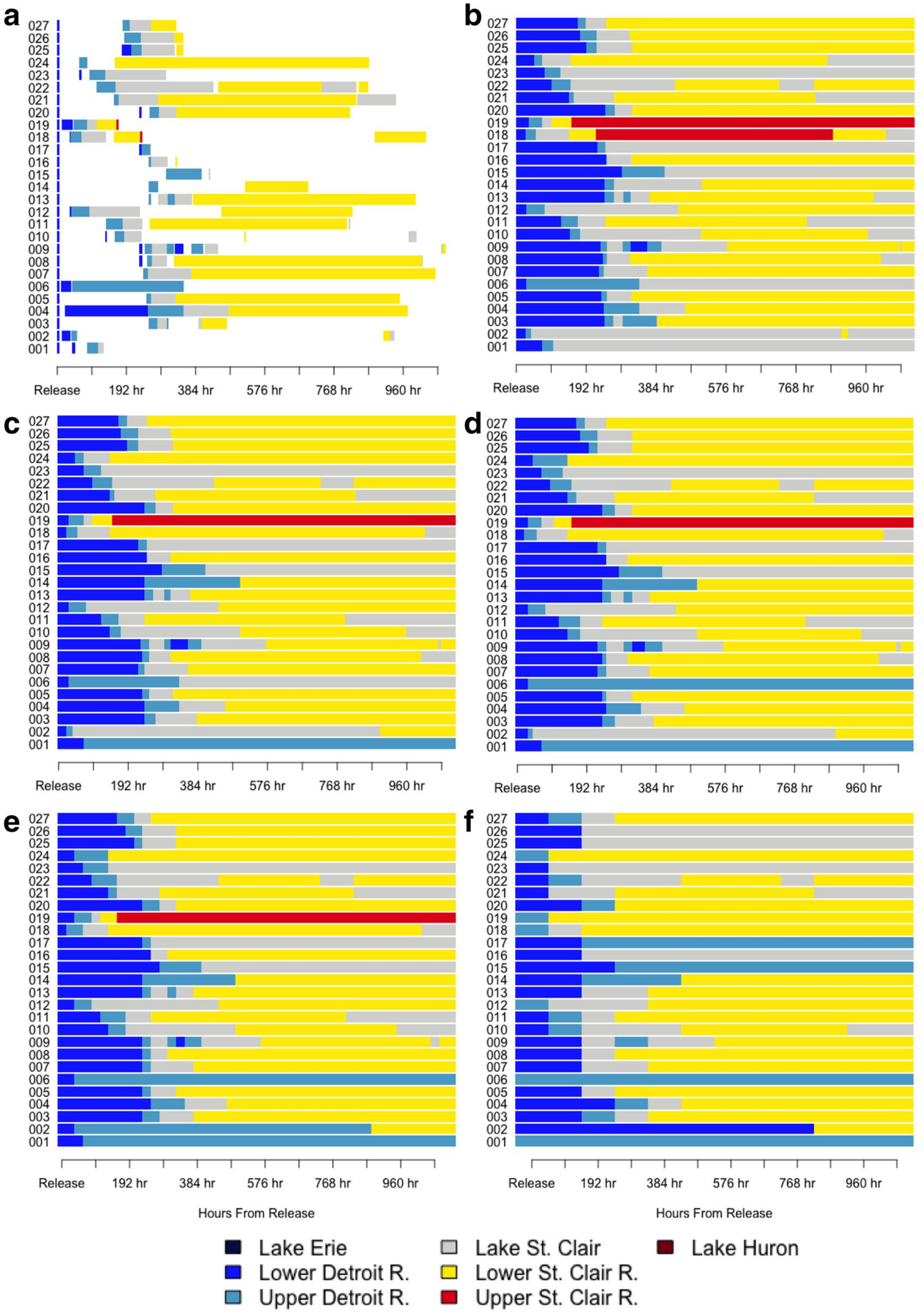

Fig. 3 Movement sequences at different resolutions. Detection events displaying the **a** raw detection data for all acoustic-tagged sea lamprey (n = 27) in the St. Clair River Detroit River System. Movement histories constructed at **b** 1-, **c** 6-, **d** 12-, **e** 24-, and **f** 96-h intervals from the release date (16 May 2014) in the Lower Detroit River (DRL) until 1 July 2014. Each tick on the x-axis represents 4 days

Table 1 Proportion of movement histories imputed

Fish	1 h	6 h	12 h	24 h	96 h
001	0.88	0.86	0.82	0.81	0.83
002	0.86	0.84	0.80	0.79	0.83
003	0.87	0.88	0.85	0.85	0.67
004	0.11	0.14	0.12	0.13	0.08
005	0.34	0.36	0.34	0.35	0.25
006	0.68	0.69	0.68	0.67	0.67
007	0.24	0.26	0.25	0.26	0.17
008	0.30	0.31	0.28	0.26	0.17
009	0.81	0.80	0.75	0.72	0.58
010	0.88	0.88	0.84	0.80	0.58
011	0.39	0.39	0.35	0.35	0.25
012	0.48	0.49	0.46	0.46	0.33
013	0.33	0.35	0.31	0.30	0.17
014	0.80	0.81	0.77	0.76	0.58
015	0.81	0.80	0.79	0.77	0.75
016	0.84	0.85	0.81	0.83	0.75
017	0.86	0.87	0.84	0.83	0.83
018	0.69	0.69	0.66	0.63	0.58
019	0.85	0.86	0.84	0.83	0.83
020	0.47	0.49	0.45	0.43	0.33
021	0.27	0.29	0.27	0.26	0.17
022	0.31	0.32	0.29	0.28	0.17
023	0.79	0.79	0.76	0.76	0.67
024	0.32	0.33	0.30	0.28	0.17
025	0.84	0.84	0.82	0.83	0.67
026	0.84	0.85	0.83	0.85	0.75
027	0.86	0.86	0.84	0.85	0.67
Mean =	0.62	0.63	0.60	0.59	0.50
S.E. =	0.05	0.05	0.05	0.05	0.05
n =	1104	184	92	46	12

Proportion of individual movement histories that were imputed at 1-, 6-, 12-, 24-, and 96-h intervals. Total number of time intervals indicated by n values

where *TR* is the observed transition rate from the origin state *i* to the arrival state *j* for all fish combined (Table 3a) and cval is a scalar that sets the base value for all calculations equal to 2 [59]. Substitution costs ranged from 1.001 to 1.023 (when fish remained in the same state) to 2 when fish were not observed transitioning between two states (Table 3b). Data driven indel operations were assigned a value of 1.05 [59]. Lastly, we created a custom cost regime based on the likelihood of 2nd order or higher movements occurring in the SCDRS. Despite being observed in the data due to missed detections (Fig. 4, Table 2), second order or higher order movements (i.e., movements between non-adjacent states) were physically impossible. Substitutions between adjacent states (e.g., Lower Detroit R. and Upper Detroit R.) were assigned a cost of 1 while substitutions between non-adjacent states (e.g., Upper Detroit R. to Lower St. Clair R.) were assigned a cost of 2. Individual indel operations had a cost of 0.95. However, to maintain equal lengths among the 27 movement histories (i.e., 1104 hourly observations), any indel operation was necessarily accompanied by another indel; thus the cumulative cost of an indel was 1.90.

To compare cost structures, we calculated operation summaries for the alignment of each movement history sequence to a reference sequence. The last sequence in the data set (Fish ID = "027") was arbitrarily selected as the reference sequence. The operation summaries included number of substitutions, number of indels, the total number of operations, number of 2nd order or higher movements needed to align sequences, and change in sequence length [as a proportion of the original length (n = 1104)]. The Levenshtein and Levenshtein II cost regimes resulted in the fewest and most total operations, respectively (Table 3). The Hamming and custom cost regimes were the only approaches that resulted in no 2nd order or higher movements; though the former did result in a 35% (378 h or 16 days) increase in the length of movement histories. The data driven approach and the custom cost regime performed similarly with the primary difference being a single alignment that required a 2nd order or higher movement using the data driven approach (Table 3). The custom cost regime was used in analyses because it minimized substitutions corresponding to 2nd order or higher movements and minimized changes to the length of the movement histories through indel operations.

Identifying common movement histories

Hierarchical agglomerative clustering, based on dissimilarities among movement history sequences, was used to identify common movement histories representative of groups of fish. Clusters were identified using Ward's D^2 clustering criterion and uncertainty was evaluated using multiscale bootstrap resampling (nboot = 1000) which provided approximately unbiased p-values [43]. Significant clusters ($\alpha = 0.05$) were further examined by extracting the representative set of movement history sequences from each cluster (i.e., identifying the movement history sequences that best defined each cluster) [44]. Each representative set of sequences was identified using a two-step process. In the first step, a first-order Markov model is used to estimate the sequence likelihood (i.e., the product of the probability that each successive state is expected to occur at a given time step) and the resulting probability was used to order all sequences within a cluster. Second, redundant (i.e., similar) sequences were identified as those (1) within a neighborhood radius of 25% of the theoretical maximum dissimilarity (i.e., the dissimilarity value of the two sequences in each cluster group that are maximally

Table 2 Summary of all five cost regimes

Arrival state	LE	DRL	DRU	LSC	SCL	SCL	LH
Origin state							
a							
LH	0	0	0	0	0	0	0
SCU	0	0	0	0	0.003	0.999	0
SCL	0	0	0.001	0.030	0.989	0.001	0
LSC	0	0.001	0.021	0.977	0.008	0	0
DRU	0	0.018	0.977	0.003	0	0	0
DRL	0	0.981	0.001	0	0	0	0
LE	0	0	0	0	0	0	0
b							
LH	2	2	2	2	2	2	2
SCU	2	2	2	2	1.997	1.001	2
SCL	2	2	1.999	1.97	1.011	1.999	2
LSC	2	1.999	1.979	1.023	1.992	2	2
DRU	2	1.982	1.023	1.997	2	2	2
DRL	2	1.019	1.999	2	2	2	2
LE	2	2	2	2	2	2	2

Cost structure and number of operation (mean ± standard error) summary of each of the five cost regimes. Histories with 2nd order is the number of movement histories that required second order substitutions for alignment. Proportion increase is the proportional length change for each sequence due to an indel operation

Table 3 Observed transition rates and state specific substitution costs for data-driven cost regime

Regime	Cost structure		Operations			2nd order	Percent increase
	Substitution	Indel	Substitution	Indel	Total		
Levenshtein	1	1	378.5 ± 338.1	78.5 ± 104.1	457.1 ± 360.7	12	0.04 ± 0.04
Levenshtein II	2	0.25	0.0	779.6 + 672.3	779.6 + 672.3		0.4 + 0.3
Hamming	1	3	428.0 + 347.5	0.0	428.0 + 347.5	23	0.01 + 0.04
Data driven	S2a	1.05	302.8 + 302.8	178.3 + 118.3	481.0 + 371.5	1	0.08 + 0.05
Theory driven	Custom Matrix	0.95	300.6 + 305.1	180.8 + 128.8	481.9 + 372.4		0.08 + 0.06

Observed transition rates (a) for 27 sea lamprey in the Saint Clair Detroit River system. Transition rates reflect movement from the origin state to the arrival state and occur in both upstream and downstream directions. Substitution costs (b) based on the observed transition rates for 27 sea lamprey in the Saint Clair Detroit River system. From downstream to upstream, *LE* Lake Erie, *DRL* Lower Detroit River, *DRU* Upper Detroit River, *LSC* Lake St. Clair, *SCL* Lower St. Clair River, *SCU* Upper St. Clair River, and *LH* Lake Huron

different) and (2) that cover a minimum of 50% of the of the sequences in the cluster [44]). This process progressed iteratively through every candidate sequence, starting with the first sequence (i.e., highest probability of occurrence from previous step; centroid of the cluster). Two measures of quality were used to indicate the amount of spread among sequences within each cluster (i.e., 'within representative sequence spread') and the mean distance of the representative sequence to the cluster centroid (i.e., 'mean distance') [44]. All analyses were conducted in the R-environment (version 3.4.3; [45]). Detection data were processed using the 'glatos' package in R. The R package 'TraMineR' was used for developing dissimilarity measures among movement sequences [46] and cluster analyses were done using the 'pvclust' function in the 'pvclust' package [43].

Results

All 27 acoustic-tagged sea lamprey were detected in the SCDRS receiver array resulting in 6904 individual detections. Detections were further collated into 1072 discrete detection events (Fig. 3a) that ultimately formed 27 detection histories ranging in length from short, discontinuous sequences that required numerous imputations (e.g., individuals '001', '002', and '016'; Fig. 3a), to longer, discontinuous sequences that required moderate

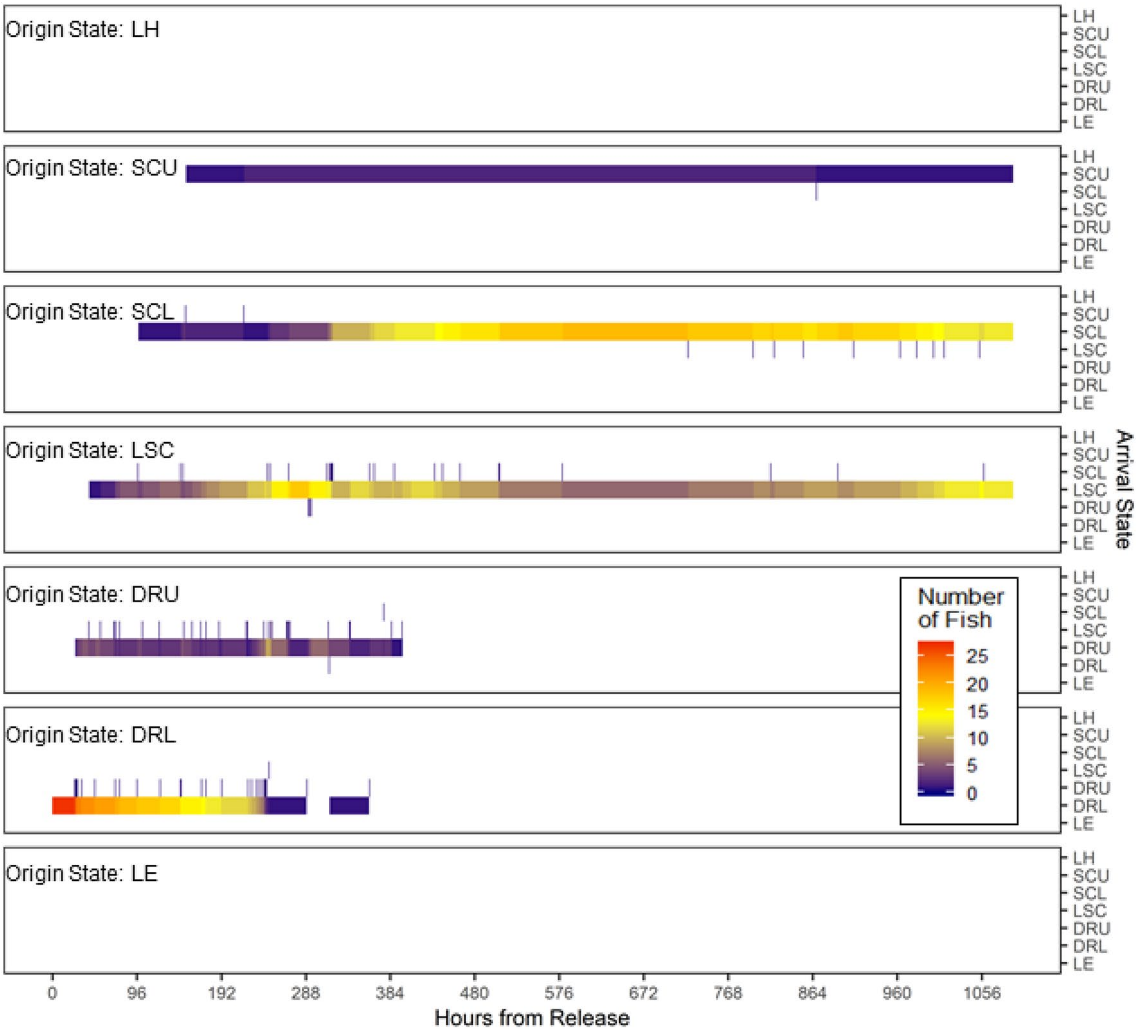

Fig. 4 Time-varying transition frequencies. Transition frequencies for sea lamprey (n = 27) in the St. Clair River Detroit River system. The color scale grades from blue (1 individual) to red (27 individuals) and each cell represents the number of fish, at 1 h intervals, remaining in the same state, moving upstream (rug plot above the same state transition lines), or moving downstream (rug plot below the same state transition lines). States are ordered from downstream to upstream; *LE* Lake Erie, *DRL* Lower Detroit R., *DRU* Upper Detroit R., *LSC* L. St. Clair, *SCL* Lower St. Clair R., *SCU* Upper St. Clair R., and *LH* Lake Huron. All fish were released on May 16, 2014 in the Lower Detroit River (DRL) and each tick represents 4 days

imputation (e.g., individuals '011', '012', and '018'; Fig. 3a), to long, detailed sequences that needed comparatively less imputation (e.g., individuals '004', '007', and '021'; Fig. 3a).

Seventy-six upstream transitions were observed (Fig. 4). Only 2.6% of upstream transitions (n = 2) were second-order movements, including a transition from the lower Detroit River to Lake St. Clair (i.e., missed in the upper Detroit River; ~ 220 h from release) and a transition from the upper Detroit River to the lower St. Clair River (i.e., missed in Lake St. Clair; ~ 380 h from release; Fig. 4). Fourteen downstream transitions were observed (Figs. 3b, 4), representing 12 distinct fallback events from 11 individuals. All downstream transitions were first-order movements. Three fallback events (3 fish) were initiated in Lake St. Clair and, in all three cases, the fish continued upstream through Lake St. Clair after the initial fallback (individuals '003', '009', and '013' in Fig. 3b). Nine fallback events (9 fish) were initiated in the St. Clair River (8 lower St. Clair R.; 1 upper St. Clair R.) and eight of those fallback events represented the final movement of the fish, terminating in Lake St. Clair. One fallback event initiated in the St. Clair River was followed by continued upstream migration after detection in Lake St. Clair.

Twenty-two of the 27 (81%) acoustic-tagged sea lamprey entered the St. Clair River, including 20 individuals that ceased upstream migration in the lower St. Clair River and two individuals that ceased migration in the upper St. Clair River (Figs. 3b, 4). Further, all 22 individuals that reached the St. Clair River arrived there by 10 June (576 h post release; Fig. 4) and remained there until 15 June when fish began exhibiting fallback behavior. Five sea lamprey ceased upstream migration in Lake St. Clair (Figs. 3b, 4) and no sea lamprey were detected in the Detroit river after 2 June 2014 (392 h after release). Some fish moved upstream quickly with advancement to Lake St. Clair, the lower St. Clair River, and the upper St. Clair River occurring in as little as 44 (Fig. 3b; individual = '002'), 99 (Fig. 3b; individual = '019'), and 153 (Fig. 3b; individual = '019') hours, respectively.

Dissimilarity values ranged from 13.0 to 1859.2 for the two movement history sequences that were most (individuals '005' and '016') and least (individuals '002' and '019') similar, respectively (Fig. 5a). Five significant clusters were identified among the 27 movement sequences (Fig. 5b). Cluster 1 was defined by three fish that moved quickly through the Detroit River and Lake St. Clair, ceased upstream migration in the lower St. Clair River, and then "fell back" to Lake St. Clair (Fig. 6). Fish in Cluster 2 (n = 3) also ceased upstream migration in the lower St. Clair River but moved more slowly (2–3 weeks) through the Detroit River and Lake St. Clair. Fallback behavior was only apparent in one fish from this group (Fig. 6). The largest group, Cluster 3, contained 13 fish that moved through the Detroit River and Lake St. Clair in 2 weeks followed by extended periods in the lower St. Clair River. Despite the diversity of representative movement sequences and relatively small sample size, within cluster variability was small relative to among-cluster variability, indicating that the sequences were more similar within clusters than among clusters. Fish '009' was the lone exception due to multiple fallbacks during its migration (Fig. 3b). Cluster 4 was comprised of two fish that ceased upstream migration in the upper St. Clair River ('018' and '019; Figs. 3b; 5b). The final cluster, Cluster 5, contained three representative sequences (Fig. 6) for the six individuals that ceased upstream migration

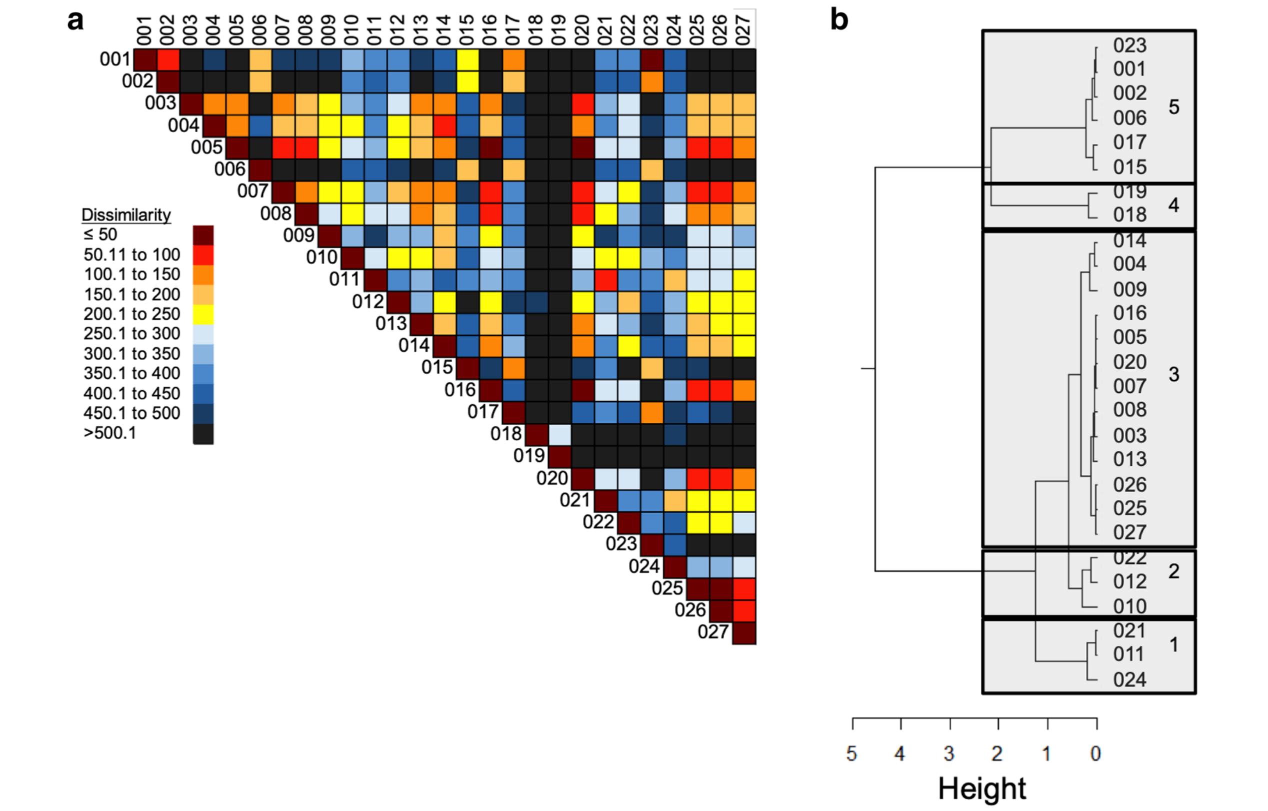

Fig. 5 Dissimilarity and Cluster Dendrogram. Dissimilarity matrix **a** calculated using optimal matching based on the custom cost regime. Colors grade from warm to cold for the most similar and least similar movement histories, respectively. Agglomerative clustering dendrogram **b** comparing movement sequences among the 27 individual fish. Gray boxes indicate significant clusters at α = 0.05 as determined by bootstrap resampling

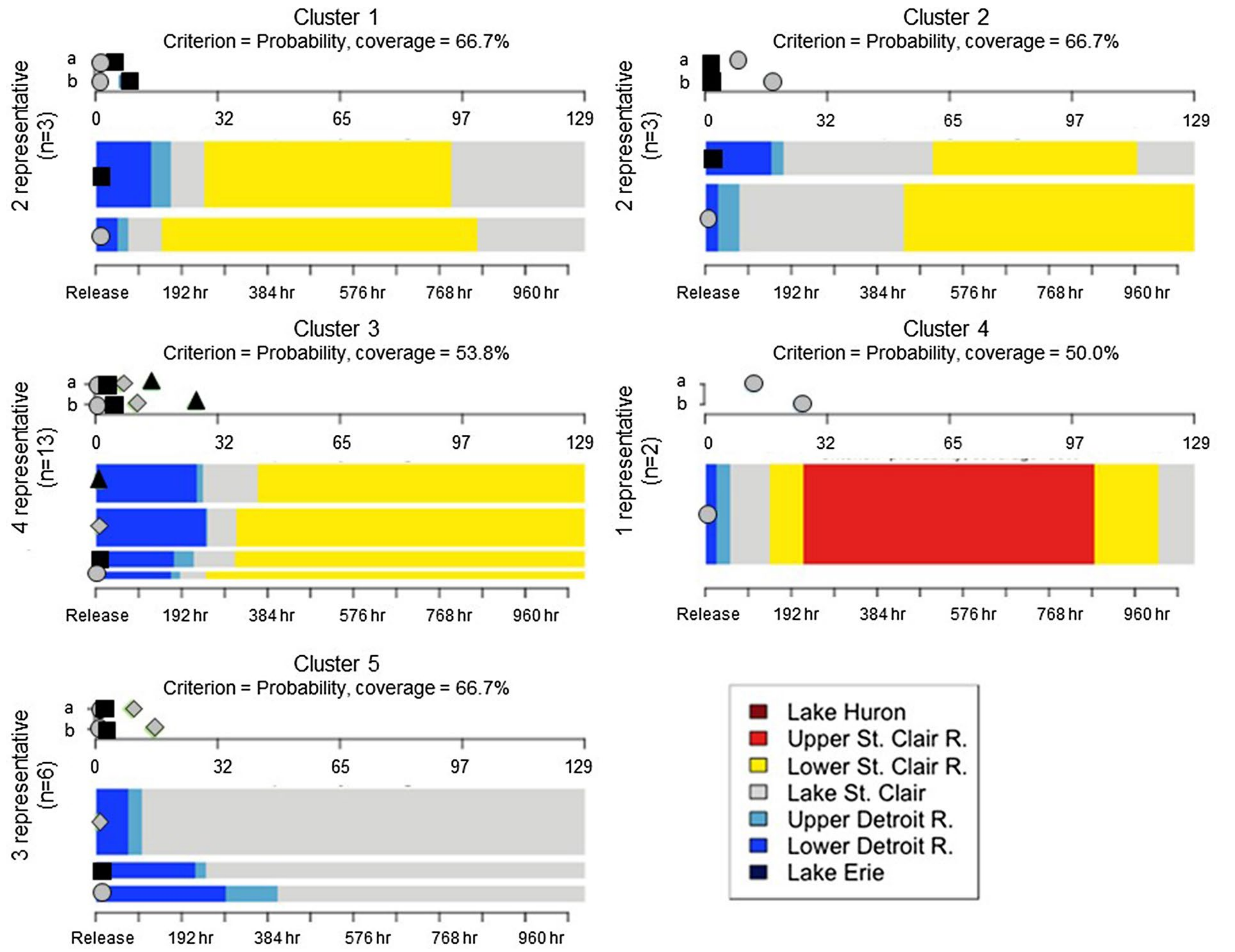

Fig. 6 Representative movement histories. Movement histories for sea lamprey in the St. Clair River Detroit River system grouped by agglomerative cluster analysis in Fig. 5b. Representative sequences were selected using the sequence probability and 50% minimum coverage criteria. Movement histories (horizontal bars) are plotted from the bottom-up according to their representativeness with the bottom bar being the centroid for that cluster (i.e., most representative). Bar thickness is proportional to the number of individuals assigned to that sequence. The symbols on the top, correspond to the symbols on the left of each representative movement history, indicate the **a** within representative sequence spread and **b** the mean distance to the cluster centroid

in Lake St. Clair. Like the fish in the first three clusters, these individuals had highly variable transit times through the Detroit River system.

Discussion

Sequence analysis methods allowed us to construct individual-level movement histories using more objective, repeatable methods than commonly-used ad-hoc alternatives, and to further use individual movement history sequences in a flexible, statistical framework to identify distinct movement patterns representing groups of individuals with common movement characteristics. Importantly, results from our analysis of movement history sequences (this paper) are consistent with a previous analysis of the same dataset [33] and both approaches lead to the conclusion that the lower St. Clair River was the most likely spawning area for sea lamprey in the SCDRS. Rather than reiterate the ecological interpretations of these movement patterns detailed in Holbrook et al. [33], we focus here on critical decisions in the sequence analysis workflow (e.g., spatial state definition, sequence time resolution, and dissimilarity cost regime) that are specifically relevant to fish movement applications.

Movement history sequences are a convenient data storage format for consistent and direct summaries of space use by individual fish, but require spatial state definitions and time resolutions that are ecologically-relevant and allow accurate imputation of missing data points. Missing data are frequently encountered in acoustic telemetry data due to numerous reasons outlined previously (e.g., imperfect receiver coverage, missed detections, etc.). Last observation carried forward (LOCF) is a popular state imputation method in medical and clinical research due its simplicity, but its use has also been criticized for apparent subjectivity [47, 48]. Such criticisms are alleviated in well-designed telemetry studies by ensuring that receivers adequately delineate states and reliably detect fish moving among states. Further, the assumption that a fish remained in a 'state' between detection events was likely accurate for our study given the closed dynamics of the SCDRS. Imputation methods should be carefully considered in open systems like lakes, estuaries, and oceans and may not be appropriate unless the system is covered by an extensive receiver array or data support broad state classification schemes. In such cases, missing data points may be filled using statistical models to interpolate an individual's position based on detection events, environmental variables, and fish swimming speeds [49–52]. It is also worth noting that some sequence alignment algorithms are capable of handling missing data points and sequences with unequal lengths [42] and can be further tuned using creative cost structures to inform missing values [53].

Identifying the appropriate temporal resolution for analyzing fish movements is a key consideration when using sequence analysis and should be explored a priori. Movement sequences constructed at multiple temporal scales resulted in highly variable data granularity that influenced our interpretation of fish movement patterns considerably. For example, while the 6 and 12 h intervals isolate peak movement times (i.e., night time) for sea lamprey from times with reduced activity [33] and reduce the number of imputed data points, the intervals were too broad and important movement patterns were not observed. Conversely, the 1 h interval captured those important features but resulted in 6 to 12 times as many imputations, even though the proportions of imputed data were not different. Ultimately, there is a trade-off between limiting the number of imputed points and the amount of information lost at coarse time scales.

Selecting the appropriate algorithm for deriving dissimilarity measures is perhaps the most important step in identifying movement typologies that represent group-level movement and space use characteristics. Though a review of the methods is beyond the scope of this paper (but see [42]), it is important to note that approaches other than those used here (including Euclidean and Chi square distances) are available for calculating pairwise similarities (based on common features) and dissimilarities between movement sequences. For example, Kessel et al. [2] used the proportion of time intervals in which the state differed between two individuals (essentially 'Hamming' distance; [54]) to identify migratory contingents among lake sturgeon. Hamming distance is an intuitive choice because it captures both spatial and temporal dynamics [55]. However, by weighting all time intervals equally, Hamming distance can fail to recognize ecologically-important differences that occur over short time scales (e.g., spawning migrations) and favor more protracted residency events.

Understanding which dissimilarity measures are best suited for specific ecological questions is also critical when using sequence analysis to study fish movements. While the number of possible algorithms and parameterizations can be overwhelming [42], sequence analysis is a flexible statistical framework that can be used to ask a number of questions. While we chose OM parameters that focused on the order of state transitions, we could have adopted many other approaches. For example, we could have used Levenshtein II, Euclidean distance with the number of periods (K) set to 2, or OMspell with a high expansion cost to group movement histories based on the duration of state occupancy. Likewise, clusters could have been based on the timing of state transitions by using dynamic Hamming or Euclidean distance with K equal to sequence length [56]. The distinction among the various algorithms is not arbitrary and selecting appropriate method depends largely on the movement aspect or question of interest [42]. Within the context of fish movement ecology, we interpret those aspects as (1) *experienced states*—total count of states occupied, (2) *sequencing*—the order of distinct successive states occupied by an individual, (3) *distribution*—total time spent in each state during the movement sequence, (4) *timing*—age, date, or time of day when an individual transitions into a state of interest, (5) *duration*—length of time individuals spend in the same state, and (6) *spacing*—elapsed time that occurs while transitioning between two states of interest.

Conclusions

Sequence analysis offers a flexible statistical framework for studying individual- and group-level fish movement histories, behavioral shifts, and habitat use that can be implemented in a reproducible manner using widely accessible software. Beyond our focus on finding commonalities in fish movement histories, additional statistical approaches have been developed specifically for analyzing sequential data such as fish movement histories [57–59]. Likewise, the dissimilarity measures derived from alignment algorithms such as OM are analogous to those found in community ecology and could be used to ask increasingly complex questions regarding fish movement patterns. Sequence analysis is not intended as a panacea or as an alternative to spatially explicit movement models that allow for more rigorous prediction of habitat use [55]. Rather, it may be viewed as either a complement to those models (i.e., exploratory tool for informing hypothesis development) or a stand-alone semi-quantitative method for generating a simplified, temporally and spatially structured view of complex acoustic telemetry data and hypothesis testing when observed patterns warrant further investigation.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s40317-020-00195-y.

Additional file 1. Selecting the appropriate temporal resolution. **Figure 1.** Cluster analysis of individual movement histories at different time intervals.

Acknowledgements

We thank Scott Colborne, Todd Hayden, and the anonymous reviewers for their insightful comments during the development of this manuscript. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US Government. This work was funded by the Great Lakes Fishery Commission (Grant ID #2013_BIN_44024) by way of Great Lakes Restoration Initiative appropriations (Grant ID #GL-00E23010). This paper is contribution 76 of the Great Lakes Acoustic Telemetry Observation System.

Authors' contributions

CMH and DWH contributed to acquisition of data. MRL, CMH, and DWH contributed to conception and study design. MRL and CMH contributed to analysis of data. MRL, CMH, and DWH contributed to interpretation of results and writing of the manuscript. All authors read and approved the final manuscript.

Funding

The data used in this project were funded by the Great Lakes Fisheries Commission through Great Lakes Restoration Initiative appropriations (GL-00E23010-3).

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

All applicable international, national, and/or institutional guidelines for the care and use of animals were followed.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1] Hammond Bay Biological Station, Great Lakes Science Center, United States Geological Survey, 11188 Ray Rd., Millersburg, MI 49759, USA. [2] Great Lakes Science Center, United States Geological Survey, 1451 Green Rd., Ann Arbor, MI 48105, USA.

Received: 28 March 2019 Accepted: 21 February 2020
Published online: 19 March 2020

References

1. Crossin GT, Heupel MR, Holbrook CM, Hussey NE, Lowerre-Barbieri SK, Nguyen VM, et al. Acoustic telemetry and fisheries management. Ecol Appl. 2017;27:1031–49.
2. Kessel ST, Hondorp DW, Holbrook CM, Boase JC, Chiotti JA, Thomas MV, et al. Divergent migration within lake sturgeon (*Acipenser fulvescens*) populations: multiple distinct patterns exist across an unrestricted migration corridor. J Anim Ecol. 2018;87:259–73.
3. Dionne PE, Zydlewski GB, Kinnison MT, Zydlewski J, Wippelhauser GS. Reconsidering residency: characterization and conservation implications of complex migratory patterns of shortnose sturgeon. Can J Fish Aquat Sci. 2013;127:119–27.
4. Melnychuk MC, Dunton KJ, Jordaan A, Mckown KA, Frisk MG. Informing conservation strategies for the endangered Atlantic sturgeon using acoustic telemetry and multi-state mark-recapture models. J Appl Ecol. 2017;54:914–25.
5. Lin HY, Roberts DT, Brown CJ, Fuller RA, Dwyer RG, Harding DJ, et al. Impacts of fishing, river flow and connectivity loss on the conservation of a migratory fish population. Aquat Conserv Mar Freshw Ecosyst. 2018;28:45–54.
6. Holbrook CM, Bergstedt RA, Barber JM, Bravener GA, Jones ML, Krueger CC. Evaluating harvest-based control of invasive fish with telemetry: performance of sea lamprey traps in the Great Lakes. Ecol Appl. 2016;26:1595–609.
7. Coulter AA, Brey MK, Lubejko M, Kallis JL, Coulter DP, Glover DC, et al. Multistate models of bigheaded carps in the Illinois River reveal spatial dynamics of invasive species. Biol Invasions. 2018;20:3255–70.
8. Donaldson MR, Hinch SG, Suski CD, Fisk AT, Heupel MR, Cooke SJ. Making connections in aquatic ecosystems with acoustic telemetry monitoring. Front Ecol Environ. 2014;12:565–73.
9. Hussey NE, Kessel ST, Aarestrup K, Cooke SJ, Cowley PD, Fisk AT, et al. Aquatic animal telemetry: a panoramic window into the underwater world. Science. 2015;348:1255642.
10. Hayden TA, Holbrook CM, Fielder DG, Vandergoot CS, Bergstedt RA, Dettmers JM, et al. Acoustic telemetry reveals large-scale migration patterns of walleye in Lake Huron. PLoS ONE. 2014;9:e114833.
11. Kristensen ML, Birnie-gauvin K, Aarestrup K. Routes and survival of anadromous brown trout *Salmo trutta* L. post-smolts during early marine migration through a Danish fjord system. Estuar Coast Shelf Sci. 2018;209:102–9.
12. Perry RW, Skalski JR, Brandes PL, Sandstrom PT, Klimley AP, Ammann A, et al. Estimating survival and migration route probabilities of juvenile chinook salmon in the Sacramento-San Joaquin River Delta. North Am J Fish Manag. 2010;30:142–56.
13. Dean MJ, Hoffman WS, Zemeckis DR, Armstrong MP. Fine-scale diel and gender-based patterns in behaviour of Atlantic cod (*Gadus morhua*) on a spawning ground in the Western Gulf of Maine. ICES J Mar Sci. 2014;71:1474–89.
14. Lidgard DC, Bowen WD, Jonsen ID, Iverson SJ. Animal-borne acoustic transceivers reveal patterns of at-sea associations in an upper-trophic level predator. PLoS ONE. 2012;7:e48962.
15. Espinoza M, Farrugia TJ, Webber DM, Smith F, Lowe CG. Testing a new acoustic telemetry technique to quantify long-term, fine-scale movements of aquatic animals. Fish Res. 2011;108:364–71.
16. Binder TR, Farha SA, Thompson HT, Holbrook CM, Bergstedt RA, Riley SC, et al. Fine-scale acoustic telemetry reveals unexpected lake trout, *Salvelinus namaycush*, spawning habitats in northern Lake Huron, North America. Ecol Freshw Fish. 2018;27:594–605.
17. Pincock DG, Johnston SV. Acoustic telemetry overview. In: Adams NS, Beeman JW, Eiler JH, eds. Telem Tech a user Guid Fish Res. 2012. p. 305–37.
18. Havrylkoff BJ, Peterson MS, Slack WT. Assessment of the seasonal usage of the lower Pascagoula River estuary by Gulf sturgeon (*Acipenser oxyrinchus desotoi*). J Appl Ichthyol. 2012;28:681–6.
19. Binder TR, Marsden JE, Riley SC, Johnson JE, Johnson NS, He J, et al. Movement patterns and spatial segregation of two populations of lake trout *Salvelinus namaycush* in Lake Huron. J Great Lakes Res. 2017;43:108–18.
20. Gurarie E, Bracis C, Delgado M, Meckley TD, Kojola I, Wagner CM. What is the animal doing? Tools for exploring behavioural structure in animal movements. J Anim Ecol. 2016;85:69–84.

21. Martins EG, Gutowsky LFG, Harrison PM, Flemming JEM, Jonsen ID, Zhu DZ, et al. Behavioral attributes of turbine entrainment risk for adult resident fish revealed by acoustic telemetry and state-space modeling. Anim Biotelemetry. 2014;2:1–13.
22. Gahagan BI, Fox DA, Secor DH. Partial migration of striped bass: revisiting the contingent hypothesis. Mar Ecol Prog Ser. 2015;525:185–97.
23. Abbott A, Forrest J. Optimal matching methods for historical sequences. J Interdiscip Hist. 1986;16:471–94.
24. Mount DW. Bioinformatics: sequence and genome analysis. In: Mount DW, editor. Bioinforma Seq Genome Anal. 2nd ed. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 2004.
25. Bargeman B, Joh C-H, Timmermans H. Vacation behavior using a sequence alignment method. Ann Tour Res. 2002;29:320–37.
26. Smith B, Tibbles J. Sea lamprey (*Petromyzon marinus*) in lakes Huron, Michigan, and Superior: history of invasion and control, 1936–78. Can J Fish Aquat Sci. 1980;37:1780–801.
27. Christie GC, Goddard CI. Sea Lamprey International Symposium (SLIS II): advances in the Integrated Management of Sea Lamprey in the Great Lakes. J Great Lakes Res. 2003;29:1–14.
28. Manion PJ, Hanson LH. Spawning behavior and fecundity of lampreys from the upper three great lakes. Can J Fish Aquat Sci. 1980;37:1635–40.
29. McLain AL, Smith BR, Moore HH. Experimental control of sea lampreys with electricity on the south shore of Lake Superior, 1953–60, Gt. Lakes Fish. Comm Tech Rep. 1965;10:1–48.
30. Morman RH. Distribution and ecology of lampreys in the lower peninsula of Michican, 1957–75. Comm: Gt. Lakes Fish; 1979.
31. Beamish FWH. Biology of the North American anadromous sea lamprey, *Petromyzon marinus*. Can J Fish Aquat Sci. 1980;37:1924–43.
32. Flescher D, Martini FH. Order Petromyzontiformes. In: Collette BB, Klein-MacPhee G, editors. Bigelow Schroeder's fishes Gulf Maine. 3rd ed. Washington and London: Smithsonian Institution Press; 2002. p. 16–9.
33. Holbrook CM, Jubar AK, Barber JM, Tallon K, Hondorp DW. Telemetry narrows the search for sea lamprey spawning locations in the St Clair-Detroit River System. J Great Lakes Res. 2016;42:1084–91.
34. Applegate VC. Natural history of the sea lamprey, Petromyzon marinus, in Michigan. Special Scientific Report-Fisheries, No. 55. U.S. Department of the Interior; 1950.
35. Holtschlag DJ, Koschik JA. A Two-Dimensional Hydrodynamic Model of the St. Clair—Detroit River Waterway in the Great Lakes Basin. Lansing, Michigan; 2002.
36. Anderson EJ, Schwab DJ, Lang GA. Real-time hydraulic and hydrodynamic model of the St Clair River, Lake St. Clair Detroit River System. J Hydraul Eng. 2010;136:507–18.
37. Edwards CJ, Hudson PL, Duffy WG, Nepszy SJ, McNabb CD, Haas RC, et al. Hydrological, morphometrical, and biological characteristics of the connecting rivers of the International Great Lakes: a review. In: Dodge DP, editor. Proc Int Large River Symp. Kansas: Canadian Special Publications in Fisheries and Aquatic Sciences; 1989. p. 240–64.
38. Hondorp DW, Roseman EF, Manny BA. An ecological basis for future fish habitat restoration efforts in the Huron-Erie Corridor. J Great Lakes Res. 2014;40:23–30.
39. Beeman JW, Perry RW. Bias from false-positive detections and strategies for their removal in studies using telemetry. In: Beeman JW, Eiler JH, editors. Telem. Tech. a user Guid. Fish. Res. Bethesda: American Fisheries Society; 2012. p. 505–18.
40. Simpfendorfer CA, Huveneers C, Steckenreuter A, Tattersall K, Hoenner X, Harcourt R, et al. Ghosts in the data: false detections in VEMCO pulse position modulation acoustic telemetry monitoring equipment. Anim Biotelemetry. 2015;3:1–10.
41. Pincock DG. False detections: What they are and how to remove them from detection data. VEMCO Appl. Note. 2012.
42. Studer M, Ritschard G. What matters in differences between life trajectories : a comparative review of sequence dissimilarity measures. J R Stat Soc Ser B Statistical Methodol. 2016;179:481–511.
43. Suzuki R, Shimodaira H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics. 2006;22:1540–2.
44. Gabadinho A, Ritschard G, Studer M, Müller NS. Mining sequence data in R with the TraMineR package: A user's guide. Univ. Geneva, 2010. 2011.
45. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2018. https://www.r-project.org/.
46. Gabadinho A, Ritschard G, Mülller NS, Studer M. Analyzing and visualizing state sequences in R with TraMineR. J Stat Softw. 2011;40:1–37.
47. Lachin JM. Fallacies of last observation carried forward analyses. Clin. Trials. 2016;13:161–8.
48. Kenward MG, Molenberghs G. Last observation carried forward: a crystal ball? J Biopharm Stat. 2009;19:872–88.
49. Hedger RD, Martin F, Dodson JJ, Hatin D, Caron F, Whoriskey FG. The optimized interpolation of fish positions and speeds in an array of fixed acoustic receivers. ICES J Mar Sci. 2008;65:1248–59.
50. Bergé J, Capra H, Pella H, Steig T, Ovidio M, Bultel E, et al. Probability of detection and positioning error of a hydro acoustic telemetry system in a fast-flowing river: intrinsic and environmental determinants. Fish Res. 2012;125–126:1–13.
51. Grothues TM, Davis WC. Sound pressure level weighting of the center of activity method to approximate sequential fish positions from acoustic telemetry. Can J Fish Aquat Sci. 2013;70:1359–71.
52. Binder TR, Holbrook CM, Hayden TA, Krueger CC. Spatial and temporal variation in positioning probability of acoustic telemetry arrays: fine-scale variability and complex interactions. Anim Biotelemetry. 2016;4:4.
53. Lazar A, Jin L, Spurlock CA, Wu K, Sim A. Data quality challenges with missing values and mixed types in joint sequence analysis. 2017 IEEE Int. Conf. Big Data (Big Data). 2017. p. 2620–7.
54. Hamming RW. Error detecting and error correcting codes. Bell Syst Tech J. 1950;29:147–60.
55. De Groeve J, de Weghe N, Van Ranc N, Neutens T, Ometto L, Rota-stabelli O, et al. Extracting spatio-temporal patterns in animal trajectories: an ecological application of sequence analysis methods. Methods Ecol Evol. 2016;7:369–79.
56. Biemann T. A transition-oriented approach to optimal matching. Sociol Methdol. 2011;41:195–221.
57. Gauthier J-A, Widmer ED, Bucher P, Notredame C. 1. Multichannel sequence analysis applied to social science data. Sociol Methodol. 2010;40:1–38.
58. Studer M, Ritschard G, Gabadinho A, Mu NS. Discrepancy analysis of state sequences. Sociol Methods Res. 2011;40:471–510.
59. Helske S, Helske J, Eerola M. Combining sequence analysis and hidden Markov models in the analysis of complex life sequence data. In: Helske S, editor. Seq Anal Relat Approaches. New York: Springer; 2018. p. 185–200.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.