Multiscale Analysis of River Networks using the R Package linbin


To link to this article: http://dx.doi.org/10.1080/02755947.2015.1044764

Published online: 17 Jul 2015.

Submit your article to this journal

Article views: 206

View related articles

View Crossmark data

Citing articles: 1 View citing articles
Management Brief

Multiscale Analysis of River Networks using the R Package linbin

Ethan Z. Welty*1 and Christian E. Torgersen
U.S. Geological Survey, Forest and Rangeland Ecosystem Science Center, Cascadia Field Station, School of Environmental and Forest Sciences, University of Washington, Box 352100, Seattle, Washington 98195, USA

Samuel J. Brenkman
U.S. National Park Service, Olympic National Park, 600 East Park Avenue, Port Angeles, Washington 98362, USA

Jeffrey J. Duda
U.S. Geological Survey, Western Fisheries Research Center, 6505 Northeast 65th Street, Seattle, Washington 98115, USA

Jonathan B. Armstrong
Wyoming Cooperative Fish and Wildlife Unit, University of Wyoming, 1000 East University Avenue, Laramie, Wyoming 82071, USA

Abstract
Analytical tools are needed in riverine science and management to bridge the gap between GIS and statistical packages that were not designed for the directional and dendritic structure of streams. We introduce linbin, an R package developed for the analysis of riverscapes at multiple scales. With this software, riverine data on aquatic habitat and species distribution can be scaled and plotted automatically with respect to their position in the stream network or—in the case of temporal data—their position in time. The linbin package aggregates data into bins of different sizes as specified by the user. We provide case studies illustrating the use of the software for (1) exploring patterns at different scales by aggregating variables at a range of bin sizes, (2) comparing repeat observations by aggregating surveys into bins of common coverage, and (3) tailoring analysis to data with custom bin designs. Furthermore, we demonstrate the utility of linbin for summarizing patterns throughout an entire stream network, and we analyze the diel and seasonal movements of tagged fish past a stationary receiver to illustrate how linbin can be used with temporal data. In short, linbin enables more rapid analysis of complex data sets by fisheries managers and stream ecologists and can reveal underlying spatial and temporal patterns of fish distribution and habitat throughout a riverscape.

Ecological patterns and processes occur at multiple spatial and temporal scales within a river network (Levin 1992; Fausch et al. 2002), and this complexity is increasingly being examined in fisheries and water resources management (Arthington et al. 2010; Wheaton et al. 2010; Nakagawa 2014). The riverscape approach to investigating and managing stream fishes emphasizes the importance of considering the many spatial scales that are relevant to the diverse life histories of fish and the objectives of fisheries managers (Fausch et al. 2002). However, despite considerable advancements during the last decade, existing tools do not allow analysts to nimbly toggle between scales during stream analysis and modeling (Burnett et al. 2007; Brenkman et al. 2012; Carbonneau et al. 2012; Lawrence et al. 2012; Klett et al. 2013; McMillan et al. 2013). Although better tools have been developed to bridge the gap between spatial data in a GIS and statistical packages (Benda et al. 2007; Isaak et al. 2014; Peterson and Ver Hoef 2014; Ver Hoef et al. 2014), none of these tools allows users to quickly and easily evaluate the influence of scale on patterns of species distribution and aquatic habitat.

*Corresponding author: ethan.welty@colorado.edu
1Present address: Institute of Arctic and Alpine Research, University of Colorado, Campus Box 450, Boulder, Colorado 80309, USA.
Received November 13, 2014; accepted April 14, 2015
Analysis of riverscape data first requires that geographically referenced variables be extracted from a GIS and plotted as a function of river distance. We define these plots of biotic and abiotic variables (y-axis) versus distance along the river channel (x-axis) as longitudinal profiles. Creating such profiles from complex riverine data sets (e.g., with braided channels, gaps in sampling effort, and irregular sampling intervals) is difficult and time consuming. Hence, these analyses may be beyond the reach of many stream ecologists and fisheries managers who use fish distribution, habitat, and water quality data collected in elaborate spatial and temporal arrangements. The need to analyze the associations between fish species and their habitat at multiple scales led to the development of flexible and automated routines for longitudinal data aggregation and plotting. Used by Brenkman et al. (2012), Lamperth (2012), Lawrence et al. (2012, 2013), and Fullerton et al. (in press), these functions have now been formalized in the R package linbin (“linear binning”; Welty 2015) and published as open-source code on the Comprehensive R Archive Network (CRAN; R Development Core Team 2014). In this paper, we describe how linbin works, and we provide examples of its applications.

**LINBIN: A TOOL FOR MULTISCALE ANALYSIS OF RIVERSCAPES**

The R package linbin was originally conceived for the analysis of detailed, spatially continuous riverscape data (e.g., Torgersen et al. 2006) and has been used by the U.S. Geological Survey, the U.S. National Park Service, and other investigators studying spatial patterns in riverine landscapes (e.g., Brenkman et al. 2012; Lamperth 2012; Lawrence et al. 2012). It is now actively maintained and available as free, open-source code on GitHub (github.com) and in the CRAN package repository (Welty 2015).

**Linbin Workflow and Core Functions**

The linbin package contains a suite of functions to perform multiscale analysis. Steps include (1) creating, converting, and reading input data from a file (events, as_events, and read_events); (2) designing the “bins”—that is, the intervals over which the data are summarized (e.g., event_range, event_coverage, and seq_events); (3) assigning data to the bins (sample_events); and (4) creating sets of bar plots from the binned data to visualize longitudinal profiles of riverscapes at multiple scales (plot_events; see Figure 1 for an example).

**Linearly Referenced Input Data**

The linbin package is broadly conceived for any data that are arranged along a single dimension, whether the dimension is spatial (e.g., distance upriver), temporal (e.g., time elapsed), or neither (e.g., a sum or percentage variable). As an example of how riverscape data are reduced to one dimension for linbin, Figure 2a shows riverine habitat mapped in a GIS as a series of adjacent units (Radko 1997; ESRI 2003, 2010). In such spatial data, position can be expressed as distance upstream measured from a downstream reference point (e.g., river mouth or confluence), equivalent to the “river mile” or “river kilometer” used in cartography. In GIS science, this technique of expressing geographic positions as measurements along a line is known as linear referencing. The line measures are used to locate “events” along the line: either (1) point events (e.g., salmon redd, logjams, and road crossings) with one measure or (2) line events with two endpoint measures (e.g., the habitat units of Figure 2a). The linbin package stores linearly referenced data in an “event table,” wherein each row includes a point event or a line event and the values of any variables (e.g., water depth or number of fish) associated with that event.

**Designing the Bins**

An important concept for bin design in linbin is event “coverage,” or the intervals over which the data contain no gaps (see Figure 2b). Groups of sequential bins can be generated automatically from the event coverage with the function seq_events by using one of three strategies (illustrated in Figure 3): (1) a fixed number of bins (if the data contain gaps, then the bin endpoints are adjusted so that each bin contains an equal share of the total event coverage); (2) a fixed bin length (the bin endpoints are adjusted so that each bin contains the specified length of coverage; when bin length does not evenly divide the total coverage, a bin with the remainder is added to the end of the sequence); and (3) an adaptive bin length (bin lengths are varied about the specified length such that a whole number of bins fits within each interval of coverage, thereby preserving gaps and minimizing edge effects). By choosing to preserve breaks between adjacent or overlapping units (Figure 2b) or inserting custom breaks in the coverage (e.g., with the cut_events function), the third strategy can also ensure that the bin sequence corresponds to hydrologic or geomorphic features (e.g., stream reaches or tributary confluences).

**Assigning Data to the Bins**

The binning function sample_events can process event variables of all types (numeric, logical, character, etc.) since it allows the use of all functions that compute a single value from one or more vectors of values. Functions that are commonly used on single numeric variables include sum (e.g., channel length or fish counts), mean (e.g., depth, wetted width, or percent substrate), and min and max (e.g., minimum or maximum depth). A function that is commonly used on multiple variables is weighted.mean (e.g., mean depth weighted by channel unit length). Categorical variables, such as channel unit type (e.g., pool or riffle), can be applied as filters to any computation (e.g., the mean depth of pools or the sum of fish in riffles).
Binning begins by cutting events at bin endpoints via the cut_events function (Figure 2c). When events are cut, any user-specified variables (typically sums) can be scaled to the relative lengths of the resulting events (i.e., an assumption of uniform distribution); all other variables remain unchanged. Finally, the variables are computed according to the specified sampling functions from the (cut) events that fall within each bin (example 1 in Figure 2c).

In cases where data are collected in braided stream channels, overlapping events can be merged together in a preliminary binning step (example 2 in Figure 2c). In this way, for example, contributions of parallel channels to a bin mean can be weighted by their width, while contributions from adjacent units can be weighted by their length.

**Plotting the Binned Data**

The plotting function plot_events produces a grid of bar plots for all variables and groups of bins. Batch processing and plotting in linbin make it possible to explore patterns in riverine data at multiple scales without needing to laboriously compute and plot individual longitudinal profiles. To incorporate information on sampling effort, bins with no data are not shown, whereas bins with a value of zero are drawn as a thin black line.

**APPLICATIONS IN RIVERINE MANAGEMENT AND RESEARCH**

To illustrate the applications of linbin to riverscape science, we provide examples from rivers and streams in Washington and Alaska. We demonstrate the utility of linbin for examining spatial patterns in long river sections or throughout an entire stream network and for analyzing temporal patterns of fish movement. The data for these case studies are included in the linbin package; the code is provided in the package documentation (Welty 2015).

A 53-km, spatially continuous snorkel survey of the Quinault River, Washington (August 2009), illustrates how linbin can be used to quantify multiscale patterns. Figure 1
depicts trout abundance (from visual counts) at a range of bin sizes from 100 to 25,600 m. The scales at which fish counts are binned can be specified by the user to reveal different patterns attributable to a hierarchy of ecological processes (Frissell et al. 1986; Turner et al. 1989; Levin 1992).
Similar riverscape surveys were conducted throughout 65 km of the Elwha River, Washington, during summer low flow in August 2007 and August–September 2008 (Brenkman et al. 2012). These surveys had differing spatial gaps where no data were collected due to high water velocities in canyon sections that were unsafe for snorkeling. Furthermore, long reaches were sampled more coarsely in 2007 than in 2008. The linbin package facilitated the spatially explicit comparison of fish abundance between years by aggregating the data from both surveys into bins corresponding to their largest common intervals of coverage. Despite differences in hydrology between the two study years, the patterns in adult fish abundance were similar (Brenkman et al. 2012). During the 2008 survey, physical habitat variables were collected concurrently with fish counts; linbin was used to resample variables (e.g.,

![Longitudinal profiles of mean wetted width throughout the Elwha River, Washington, in 2008, illustrating the different strategies for automatic bin generation in the linbin package. Resampling of (a) the original survey data is illustrated as follows: (b) equal-length bins (ignoring gaps), (c) equal-coverage bins (straddling gaps), and (d) variable-length bins locally adapted to fit the coverage of the data (preserving gaps). The latter strategy was used by Brenkman et al. (2012). The conventional solid line on the x-axis is not displayed in order to reveal gaps in the data. The formatting of the bars, axes, ticks, and tick labels illustrates the default plotting output of linbin.](https://example.com/figure3)

![Mainstem and Network longitudinal profiles for NetMap variables (in distance upstream from the river mouth) for (b) the main stem (5.57-km bins containing 5.57 km of stream) and (c) the entire river network (5.62-km bins containing 14–404 km of stream). The variables were binned as means weighted by stream length and include intrinsic potential (IP; a modeled estimate of the likelihood of fish occurrence, as defined by Burnett et al. 2007) for Chinook Salmon (IP_CHINOOK), Coho Salmon (IP_COHO), and steelhead (IP_STEELHD); the fraction of favorable habitat for North American beaver (BeavHab); and mean channel depth (DEPTH_M).](https://example.com/figure4)
mean wetted width) by using bins that were adapted to the survey coverage so as to preserve gaps and minimize edge effects (Figure 3).

The linbin package can also be used to produce longitudinal profiles for entire stream networks. NetMap (Benda et al. 2007; www.terrainworks.com) employs digital elevation models to generate detailed river networks and to compute biophysical variables for spatially continuous hydrologic units (Figure 2a) throughout the networks. Figure 4 depicts longitudinal profiles that were computed by linbin from NetMap output for both the main-stem and the entire network of the Dungeness River, Washington. For Coho Salmon *Oncorhynchus kisutch*, Chinook Salmon *O. tshawytscha*, and steelhead *O. mykiss*, the mean intrinsic potential (a modeled predictor of species occurrence developed by Burnett et al. 2007) declined rapidly with distance upstream in the tributaries—from a near-power-law decline for Chinook Salmon to a near-linear decline for steelhead. In contrast, for distance upstream in the main stem, intrinsic potential declined more slowly (for Coho Salmon and Chinook Salmon) or even increased (for steelhead). Similarly, the majority of habitat for the North American beaver *Castor canadensis* occurred in tributaries just upstream from the river mouth (Figure 4), a pattern that a main-stem-only analysis failed to reveal.

The linbin package can process temporal data in a manner similar to the processing of spatial data. For example, rather than locating fish counts on a river based on distance upstream, the time elapsed can be used to locate the fish in time. This approach can be helpful for identifying population-level trends from individual-level movement data. Here, we use linbin to visualize cyclic habitat use by juvenile Coho Salmon in a thermally heterogeneous stream. In Bear Creek, southwest Alaska, Coho Salmon prey on the eggs of Sockeye Salmon *O. nerka*, which spawn only in cold, groundwater-dominated habitats in the lower 1 km of the stream. Antenna arrays revealed that many PIT-tagged Coho Salmon gorged on eggs during the night and then moved upstream 500 m or more, where warmer temperatures accelerated digestion (Armstrong et al. 2013). Computation of fish abundance from individual residence time intervals with linbin revealed superimposed diel and seasonal patterns of Coho Salmon abundance in the stream region where Sockeye Salmon spawned (Figure 5). The high-frequency pattern reflected cyclic feeding movements at night, while the positive trend indicated an accumulation of nonmoving fish as the Sockeye Salmon run declined through mid-August. This may reflect a reduction in postfeeding movements as (1) the abundance of Sockeye Salmon eggs declines and (2) Coho Salmon, which have to spend more time foraging, are less likely to be digestively constrained.

In conclusion, we have demonstrated that linbin can be used to analyze complex data sets and reveal underlying patterns by generating multiscale summaries of variables collected throughout a riverscape. A key advantage to using linbin for multiscale analysis is that it includes a flexible and automated bin generation routine; although the provided examples are either spatial or temporal, linbin can process any data that are arranged along one dimension. It accounts for overlaps and gaps in sampling and is thus especially well-suited for the dendritic structure of streams. Furthermore, by

![Figure 5](https://example.com/f5.png)

**Figure 5.** Temporal abundance patterns of tagged Coho Salmon in the downstream 1 km of Bear Creek, Alaska, from July 29 to August 19, 2008 (Armstrong et al. 2013), normalized by the studywide abundance of Coho Salmon that were first tagged in this river reach. The vertical gray lines mark the start of each day (0000 hours [midnight] in local time).
reducing such complex data to longitudinal profiles, linbin facilitates direct interpretation and analysis by conventional plotting, smoothing, and modeling routines (e.g., locally weighted scatterplot smoothing). Once linbin has been used to summarize fish occurrence and potential explanatory variables (e.g., habitat structure) in a stream network, statistical tools can be used to model the probability of fish occurrence within each bin. The same analysis can be repeated for a range of bin lengths; in this manner, hypotheses can be tested regarding the relative influences of explanatory variables on fish distribution at different spatial scales (see Bult et al. 1998).

The linbin package can also be used in conjunction with other geospatial hydrologic tools, such as the National Hydrography Dataset (U.S. Geological Survey; nhd.usgs.gov/tools.html) and Arc Hydro (ESRI 2011), to explore longitudinal patterns in modeled watershed attributes. Furthermore, the use of linbin with other R packages that implement spatial statistical analysis (e.g., Ver Hoef et al. 2014) will make it possible, for example, to assess spatial autocorrelation in stream networks, to develop appropriately scaled covariates for use in predictive spatial statistical models (e.g., Peterson et al. 2013; Isaak et al. 2014), and to plot and analyze model output.

ACKNOWLEDGMENTS

We thank Jeremy M. Cram, Katherine J. C. Klett, David J. Lawrence, and other early adopters of the code for their testing and useful feature suggestions, and we thank Katherine Beirne for preparing the Quinault River data for analysis. We are grateful to the Working Group on Dam Removal at the U.S. Geological Survey’s John Wesley Powell Center for Analysis and Synthesis for contributing to the ideas behind this work; and the U.S. National Park Service, the U.S. Geological Survey’s Ecosystems Mission Area, and the U.S. Fish and Wildlife Service for contributing financially. We appreciate Nathaniel P. Hitt, Daniel J. Daugherty, and two anonymous reviewers for their assistance and comments on earlier drafts of the manuscript. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

REFERENCES


