COLUMBIA BASIN RESEARCH

Standard Operating Procedure for Diagnosing and Addressing Predator Detections in Salmon Telemetry Data

31 May 2025

REBECCA A. BUCHANAN[1], STEVEN L. WHITLOCK[1], LAUREN A. YAMANE[2], AND BRYAN G. MATTHIAS[2]

[1]Columbia Basin Research, School of Aquatic and Fishery Sciences, University of Washington
1325 Fourth Avenue, Suite 1515, Seattle, Washington 98101-2540

[2]U.S. Fish and Wildlife Service
850 South Guild Ave, Suite 105, Lodi, CA 95240

SCHOOL OF AQUATIC AND FISHERY SCIENCES | UNIVERSITY OF WASHINGTON

Table of Contents

Background

Problem of tag predation

- Tag predation is the transferal of an active tag from the original study subject to a predator during a predation event
- Tag predation may pose a problem if the tag is detected while still within the predator's body
- Tag predation may bias study results
    - Examples:
        - Cumulative survival through the study area: positive bias
        - Reach-specific survival: variable bias
        - Survival rate through time: positive bias
        - Behavior (habitat use, residence time, migration rate, etc.): variable bias
    - The magnitude of the bias depends on study objectives and predator behavior
- Researchers have implemented various methods of identifying post-predation detections ("predated detections" or "predated tags") in studies of juvenile salmon. We refer to these methods as "predator filters"
    - Filtering approaches fall into two main categories based on the sources of information that they use:
        - **Behavior-based:** filters that rely on inferred tag movements and assumed differences in behavior between smolts and predators
        - **Signal-based:** filters that rely on changes in tag signal, signal strength or frequency, or data collected from built-in sensors (e.g., temperature, pressure, etc.)
    - Identifying predated detections is challenging, time-consuming, subjective, and difficult to verify. Filtering methods are often poorly documented
    - The lack of standardized methods is problematic because it:
        - Hinders the timely analysis of acoustic-telemetry data
        - Limits the possibilities of effective peer review
        - Obstructs data sharing and collaboration
        - Impedes the effective use of active-tag data in adaptive management
- There is a need for standardized approaches for diagnosing tag predation in juvenile salmon survival studies based on acoustic telemetry data

Purpose of this document

- Provide a Standard Operating Procedure (SOP) for diagnosing tag predation in smolt survival studies based on acoustic telemetry data
- Provide guidance on
    - Selection of a filtering approach
    - Use of detection **metrics**
    - Use of predator data
- Identify important considerations regarding

  - Spatial structure of acoustic receiver network
  - Preparation of data for predator filter
  - Assessment of predator filter output
- Frequently asked questions

Context

- Acoustic telemetry tagging studies of juvenile salmonid survival
- Central Valley and Sacramento–San Joaquin River Delta of California
- The primary focus of the SOP is behavior-based filters rather than signal-based filters.
  - Some information is provided for use of predation detection tags or data storage tags, but the majority of the recommendations here pertain to behavior-based filters.

Additional Notes

- Recommendations provided here are intended to be widely applicable for salmon survival studies in the Central Valley and Delta
  - General recommendations provided here are also suitable for salmon survival studies in other settings
- The recommendations are not prescriptive—there may be better approaches in some cases
- Researchers should review the recommendations in the context of their study objectives and available data
- This document and additional resources are available at https://www.cbr.washington.edu/analysis/apps/tagpredation:
  - A resource guide and literature review on predator filters in the Central Valley are provided in Kelley et al. (2022)
  - An R package, "DPF" (for "Delta Predator Filter"), that implements many of the analytical steps identified in the SOP

Predator Filter Framework

Definition

- A predator filter is a data analysis procedure used to identify invalid portions of detection histories due to tag predation

Characteristics

- Formal process
  - The predator filter constitutes a discrete step in data analysis, acting as a precursor to and isolated from the main analytical objective(s) of the study, such as model fitting, parameter estimation, prediction, etc.
- Systematic

- The steps used in the predator filter are identified before the filtering process begins and/or evolve in an intentional way throughout the filtering process
- The decisions used in creating and/or implementing the predator filter are consistent across all tags

- Repeatable
  - Implementation of the predator filter by another researcher leads to the same outcome in predator diagnoses
  - All decisions and steps taken in defining and applying the predator filter are documented

Desired properties

- Outcome is robust to small changes in filtering decisions
- Makes efficient use of existing information, including:
  - Detections from all receivers available, including those not used in primary analysis (e.g., survival modeling)
  - Concurrent or historical predator tagging data from the same region
  - Pre-existing understanding of salmon and predator behavior
- Inference is on a scale that is suitable to the study objectives

Study design

- Tagged juvenile salmonids are released upstream and are monitored as they migrate downstream past a network of acoustic receivers
  - Space-for-time mark-recapture model
  - Multistate mark-recapture model
- Acoustic receivers are deployed in such a way that salmon progress during movement through the study area can be monitored and data structure is suitable for the study design (e.g., survival to a particular location can be estimated)
- Acoustic receivers whose expected detection ranges overlap spatially are grouped in **general locations**
- Detections from unique general locations do not overlap temporally
- Auxiliary data may be available, such as environmental conditions or operable barrier schedules

Telemetry data structure

- Detection data are aggregated into **detection events** at general locations
  - Each detection event is delineated by the time of its start and end
  - Unique detection events at the same general location are separated by:
    - Detection at a different general location, or
    - Time gap longer than pre-specified minimum (e.g., 1 hour; **max-delay**) in detections at the general location in question.
- **Visit** = detection event at a given general location
- **Transition** = movement from one general location to another

Level of inference

- Tag predation diagnosis may be made on various scales
  - **Tag level**: the predator filter classifies tags as being either predated or not predated by the end of their detection history
    - Suitable for estimating total survival through the study area but insufficient for estimating reach-specific survival or survival rate through time
  - **Event level**: the predator filter classifies tags as being either predated or not predated for each detection event
    - Suitable for estimating reach-specific survival, survival rate through time, and total survival through study area

Role of detection metrics

- Detection metrics ("metrics") are used to characterize the temporal status of the tag based on the detection data
  - Behavior of the tagged individual (e.g., migration rate)
  - Environment of the tag within the tagged individual or habitat (e.g., temperature)
  - Status of predation tag (e.g., signal change at trigger of predation state)
- The predator filters use the detection metrics as the basis of predation diagnosis

Spatial Framework

- Some detection metrics depend on interpretation of the tag's movement through space. Examples:
  - Migration rate
  - Movement against flow
  - Residence in high-risk zone (see below)
- Transition length (distance traveled), transition type (direction), and use of specific regions of the study area can be computed by relating receivers or general locations to a **flowline map**
- Flowline map
  - Consists of nodes and segments connecting the nodes
    - Nodes are at channel bends, flow splits, junctions, etc.
  - Distance between two locations is the sum of the lengths of the intervening segments
  - Directionality can be assigned to sequence of locations
    - Downstream = closer to the Pacific Ocean
    - Upstream = further from the Pacific Ocean
  - Specific regions may be identified by the user on the flowline map for comparison to tag movement or residence. Examples:
    - **High-risk zone** = region where predation risk is thought to be higher
    - **Unidirectional flow zone** = region where water moves in only one direction (e.g., downstream in riverine segments) rather than in multiple directions (e.g., reverse flows in tidal segments)
    - **Trans-fluvial transition zone (TFTZ)** = region where water direction changes from unidirectional to multi-directional, depending on discharge and possibly barrier status

- Barriers may affect routes available or spatial extent of the TFTZ
- Examples of flowline map, high-risk zones, TFTZs, and segments affected by barriers: see Appendix D

Detection Metrics

- Metrics are variables that represent the observed movement or experience of the tag. These variables are calculated for each tag's detection data and are used as the basis of quantitative analysis to distinguish between detections that represent the live study subject and those that occur after predation
    - Metrics are categorized in several ways:
        - Scale: event-level vs tag-level
        - Basis of data: spatiotemporal vs signal
        - Some signal-based metrics may be available only from specialized tags
    - Metrics may incorporate auxiliary data
    - Some metrics are listed below and a fuller set is defined in the Appendix
        - Researchers may define their own metrics. Review the material in this section for considerations in metric definitions

Metric scale

- Event-level
    - Calculated for each detection event
    - The length of the metric(s) time series varies among tags
- Tag-level
    - Calculated for each tag
    - Each tag has a single observed value for each metric

Types of metrics

- Spatiotemporal metrics (with examples)
    - These metrics reflect the time and location of detection
    - All spatiotemporal metrics are defined in terms of fundamental metrics:
        - Time between detection events
        - Distance between detection events (transition length)
        - Duration of detection events (residence time)
        - Transition type (downstream, upstream, etc.)
    - Metrics may be defined to characterize multiple features of a detection history
        - Multiple metrics may be defined under each category
        - Use of highly correlated metrics may be inappropriate in some methods (e.g., rule-based filter)
        - Types of metrics and examples under each category are shown below

        - This list is not exhaustive
    - Time of movement
        - Time between detection events (fundamental metric)
    - Time of residence
        - Duration of current detection event (fundamental metric)
        - Length of time spent in high-risk zone
    - Distance of movement
        - Length of transition from previous event to current event (fundamental metric)
    - Movement rate
        - Migration rate = distance scaled by time
        - Velocity (signed migration rate: positive = downstream, negative = upstream)
    - Movement pattern
        - Transition type: downstream, upstream, repeated, lateral (fundamental metric)
        - Movement against direction of river flow
- Signal-based metrics (examples)
    - Signal strength
    - Predation or trigger signal from predation tags
    - Temperature
    - Depth

Auxiliary data

- Types of auxiliary data used in metric calculations are identified below (with examples)
    - Environmental conditions
        - River discharge at Freeport or Vernalis
    - Flow barrier status
        - Delta Cross Channel Gates
        - Head of Old River temporary barrier
    - Fish condition
        - Fork length at tagging
    - User-defined classification of region
        - High-risk zones (regions where predation risk or predator density is considered high)

Considerations in metric selection

- Geospatial considerations
    - Some metrics are sensitive to the density of the receiver network. Examples:
        - Time between detection events
        - Transition length

  - Some metrics are sensitive to the size of the detection range of individual locations. Examples:
    - Residence time
  - Recommendations:
    - Define metrics in such a way that limits dependence on spatial construction of detection network. Examples:
      - Time between detection should be measured as time lag between first detections of consecutive detection events, rather than last detection of one event to first detection of next event
    - Omit general locations whose detection field varies widely over the study period
- Data structure considerations
  - Avoid NAs: some filtering methods do not allow for missing data
    - Either the metrics with NAs or the tags with missing data will be omitted from analysis
    - Recommendation: Metrics should be defined for every detection event or detection history, depending on the scale
  - Counts of rare behavior will include many zeros
    - Some filtering methods are sensitive to small changes in observations from the norm and may over-fit to zero-dominant metrics that reflect rare behavior
    - Recommendation: Convert counts or absolute measures of rare behavior to the proportion scale in order to reduce the sensitivity to zero-dominant metrics
    - Examples:
      - Use the proportion of transitions that switch from downstream to upstream, rather than the count
      - Use the proportion of time since release that the tag has spent in a high-risk zone, rather than the absolute time
  - Numeric vs. categorical
    - The recommended filtering methods are quantitative and require numeric metrics
    - Categorical metrics should be converted to numeric metrics for analysis
    - Example:
      - Convert transition type (categorical metric) to proportion of transitions that are directed upstream (numeric metric)
  - Center and scale metric data
    - Multivariate methods generally use metrics that are on the same scale
    - Not recommended for expert-opinion rule-based filter

Overview of Filtering Methods

Approaches

Behavior-based filters

- Assumptions
    - Assumption 1: There are differences in the movement behaviors of target study fish (juvenile salmonids) and their predators
    - Assumption 2: Detection metrics reflect the behavior of the animal that bears the tag
- Types of behavior-based filters
    - **Pattern Recognition**
        - Use observed multivariate or univariate patterns in full suite of detection event data to classify tags as predated
        - Tags that do not follow the average smolt-tag pattern are classified as predated
        - Example: Cluster analysis
            - Hierarchical cluster analysis
            - Gaussian mixture model
            - Resulting clusters must be interpreted as smolt or predated
        - Example: Principal component analysis
        - Tag-level classification
            - Event-level classification requires additional interpretation
                - Ordination
                - Expert opinion (see "Hybrid")
    - **Rule-based**
        - Predation is assigned when metrics exceed critical thresholds in a series of univariate tests
        - Thresholds may be defined via expert opinion or statistical basis
            - Example: outlier from observed distribution
        - Weight of evidence approach: assign predation status if minimum number of metrics exceed critical thresholds
        - Event-level or tag-level classification
    - **Hybrid**
        - Combine pattern recognition with rule-based approaches. Examples:
            - Use expert opinion to assign clusters to fate (smolt vs predated)
            - Use expert opinion to assign predation to detection event

Signal-based filters

- Predation is assigned when tag signal switches to "predation" pattern

- Decision rule is required to assign predation to detection event

Filter scale

- Tag-level filter
  - Predation assignment is made to a tag: the tag is predated or not by the end of the tag's detection history
  - Additional interpretation is required if predation assignment is required on the event level
  - Input data are at the tag level
- Event-level filter
  - Predation assignment is made to the detection event
  - Input data are at the event level
- Summarization functions: used to convert event-level metrics to tag-level metrics
  - Cumulative measures: maximum
  - Non-cumulative measures: mean
  - Other summarization functions may be used

Data from known tagged predators

- Data from known tagged predators can be used directly or indirectly to improve diagnosis of tag predation
  - Directly
    - Data from known tagged predators are used as a training set in machine learning
    - Data from known tagged predators are used in defining clusters in pattern recognition methods
      - Predated smolt-tag clusters may differ from known predator clusters if predation occurred late in the smolt-tag detection history
  - Indirectly
    - Data from known tagged predators are used to interpret clusters in pattern recognition methods but are not used in the clustering process
    - Data from known tagged predators are used to test ability of rule-based filter or ordination methods to identify predation (sensitivity)
    - Data may be detection event data or expected distributions of metric observations from predators
- Concurrent vs non-concurrent predator data
  - Concurrent
    - Definition: data from known tagged predators from the same time period, region, and tag technology (i.e., frequency, manufacturer, etc.) as the salmon study

        - Concurrent data are preferred in order to avoid invalid comparisons between smolt-tag data and predator behavior that is not representative of the smolt-tag region, season, water year, and tag technology
        - Small sample sizes of concurrent known tagged predator data may limit the benefits of having concurrent data
    - Non-concurrent
        - Definition: data from known tagged predators from outside the region, time period, and/or tag technology of the salmon study
        - Non-concurrent data may be inappropriate to use if predator behavior depends on region, season, water year type, or tag technology
        - Non-concurrent data may be suitable if the distinguishing factor is not relevant for predator behavior ("comparable" predator data)
            - May be used to augment concurrent predator data in case of small concurrent data set
        - Expected distributions of metrics for predators may be used in place of detection event data
- Level of predator representation
    - Effective use of predator data depends on how well the likely salmon predators are represented in the data set or metric distribution
    - Small data sets may provide biased or imprecise information on predator behavior, even if predator data are concurrent with smolt-tag data
        - Too few individuals tagged
        - Too few species
        - Detection histories are too short
        - Spatial network of general locations used in predator data is too sparse
    - Researchers should review filter output for "reasonable" results even if known tagged predator data are used

Implementation

Below are an outline and step-by-step guidelines for implementing a predator filter. Some steps may be iterative, such as step 1 (Data preparation) and step 3 (Compute and review metrics).

Outline of the process

1. Data preparation
2. Choose a filter
3. Compute and review metrics
4. Implement filter
5. Assess and interpret results
6. Reporting

Step by step

1. **<u>Data preparation</u>**
    a. Clean the data
        i. Release data: tag ID, release date, release latitude and longitude
        ii. Receiver deployment data:
            1. Receiver name, deployment dates, latitude and longitude
            2. Ensure that all receivers are assigned to a general location
        iii. Detection data: remove false positives (incorrect record of tag presence)
    b. Aggregate ping or interval data to event-level data
        i. Detection event is delineated by first and last detection time at a general location
        ii. Successive detection events at a given general location are separated by detections elsewhere or by a time gap in detections at the general location that exceeds a user-defined minimum (**max-delay**)
        iii. Select a value for max-delay: the longest time gap between detections at a single general location that is allowed in a single detection event
            1. Shorter max-delay: number of observed detection events will increase
            2. Longer max-delay: residence time will increase
            3. Document the value used for max-delay
            4. Note: data sets that were compiled using different max-delay thresholds are not comparable.
    c. Validate spatial and temporal structure of the event-level data
        i. Time between detection events should be positive for all successive pairs of events (i.e., no temporal overlap)
        ii. Cumulative sum of (a) time between detection events and (b) detection event durations should equal time since release to end of detection event
        iii. Violation of either of these checks indicates problem in data
            1. Error in deployment data or detection data
            2. Error in spatial structure (assignment of general locations; see next item in list)
    d. Review general location definitions
        i. Consolidate neighboring receivers when necessary
            1. Temporal overlap in detection events
                a. Reflects spatial overlap of detection fields
                b. Time between detection events is negative
                c. Assignment of tag to unique general location is not possible
                d. Small movements of tag may be interpreted as movement directed upstream
            2. Skewed metrics from close proximity
                a. Detection fields do not overlap but boundaries are close in space

b. Measurement error in distance between events is high relative to actual distance
    i. Measured distance between events is larger than actual distance
    ii. Migration rate and related metrics are artificially increased (positive bias)
c. More likely to be problematic in high velocity or high flow conditions

ii. Alternatives to consolidation
1. Remove problematic receivers from data set
    a. If consolidated general location is too large to be informative
    b. Data loss
2. Omit observations of metrics biased by close proximity
    a. In Step 3: Compute and review metrics
    b. Example: migration rate
    c. Introduces NAs into data

2. Choose a filter

a. Pattern recognition vs. rule-based filter: Pros and cons
    i. Pattern recognition
        1. Pros
            a. Removes the subjectivity of an expert-opinion filter
            b. Allows the patterns in the data drive the diagnosis process
            c. Reduces the knowledge of the system and species required by the user
            d. Uses the joint distribution of metrics as the basis of diagnosis
        2. Cons
            a. Small changes in behavior may be amplified
            b. May overestimate predation if very few smolt tags are actually predated, or underestimate predation if most smolt tags are predated
            c. Does not accept missing data (NAs)
            d. Some methods produce results only on the tag level
        3. Neutral
            a. Although less subjective than expert-opinion filter, still depends on implementation decisions and interpretation of results
            b. Concurrent predator data may be used to interpret clusters and reduce sensitivity to small changes in smolt behavior
    ii. Rule-based filters
        1. Pros
            a. Makes use of user's knowledge of system and species

b. Easily understandable
c. Allows for either very high or very low rate of tag predation
d. Less sensitive to small changes in behavior
e. Does not require concurrent data from known tagged predators
f. Assigns predator classification at event level

2. Cons
   a. Subjective
   b. High effort to define critical thresholds
   c. Increased user effort for each additional metric
   d. Uses marginal distribution of each metric as basis of diagnosis (does not consider correlation among metrics)
3. Neutral
   a. Requires sensitivity analysis

b. Considerations
   i. Do you need event-level or tag-level classifications?
   ii. What type of predator data do you have?
      1. Detection data from known predator tags concurrent with smolt-tag data
      2. Detection data from known predator tags non-concurrent with smolt-tag data
      3. Distributions of metrics from predators observed in similar region and water year type ("comparable" predator data)
      4. Distributions of metrics from predators observed in different region or different water year type ("non-comparable" predator data)
      5. None
   iii. Does your data set have many NAs among the metrics?
   iv. How confident are you in your understanding of smolt vs predator behavior in your system?
   v. Do you have predation tag data?
   vi. Do you have temperature or depth data from a data storage tag?

3. Compute and review metrics

a. Compute metrics for each detection event for each tag
b. Review metric distributions
   i. Basic summarization tools
      1. Mean, median, range
      2. Plots
         a. Histograms
         b. Box plots
   ii. Potential issues
      1. Negative values in migration rate or time between events

2. Very high values in migration rate
3. Errors in transition type
4. Zero-inflated metrics
5. Scale dependence

4. Implement filter

a. Summarize metrics to appropriate scale (e.g., tag-scale for pattern-recognition filters)
b. Transform metrics if necessary. Examples:
    i. Log transformation for skewed distributions for pattern-recognition filters
    ii. Scale metrics to ensure common scale among all metrics for pattern-recognition filters
c. Document all researcher decisions required for the filter (see Step 6: Reporting)
d. Implement filter via replicable code set rather than "by hand" or via researcher examination of the data
e. Document code and make available
f. Implement filter on data from salmon tagging study (i.e., the primary data set)
g. Implement filter on known tagged predator data (if applicable)
h. Implement filter on combined data set of smolt-tag and predator data (if applicable)
i. Follow tag-level diagnosis with event-level diagnosis (if applicable)

5. Assess and interpret results

a. Interpret and visualize results
    i. Assign predated status to clusters (tag-level)
    ii. Compile number of tags classified as predated: total, by general location, by region
    iii. Review results: do they make sense?
        1. It is unlikely both that there were no predated tags and that all tags were predated
        2. Examine the metric values for the suite of predated and non-predated tags
        3. Compare the metric values for the same transitions (i.e., between the same pair of general locations) from both the smolt tags flagged as predated and the known predator tags.
    iv. Examine metrics that were flagged most often (rule-based filters)
        1. Were outcomes associated among metrics (i.e., duplicative metrics)?
        2. Example: Examine confusion matrix or other measure of association
    v. Examine the characteristics of clusters or ordination axes (pattern recognition filters)
        1. Are metric distributions within clusters sensible?
        2. Examine amount of variability explained by first few principal component axes (if applicable); higher degree of variability accounted for indicates higher confidence in filter output
    vi. Sensitivity analysis

1. Sensitivity to filter approach used
2. Sensitivity to suite of metrics used
3. Sensitivity to transformations used
4. Sensitivity to settings of pattern recognition analysis
5. Sensitivity to critical thresholds used in rule-based filter
6. Sensitivity to weight-of-evidence threshold in rule-based filter

vii. Map locations of predator assignments (if applicable)

viii. Examine performance on known tagged predators (if applicable)

1. Ideal: All known tagged predators are flagged as predated
2. More likely: Some known tagged predators may not be flagged as predated (e.g., if their detection histories are short or they exhibit downstream migratory behavior)

b. Compare among filters (if applicable)

i. It is recommended but not necessary to apply multiple predator filters. If multiple filters are implemented, it is recommended to compare output among filters

ii. Examine consensus in tag-level or event-level fate assignments among filters

iii. Consider study objectives in deciding whether to use the union or intersection of tags or detection events flagged as predated for the final predation assignments across multiple filters

6. Reporting

1. Key information for replicating the filter
   i. Acoustic data cleaning details (e.g., false positive filtering method)
   ii. Event definitions (e.g., max-delay value used)
   iii. General location definitions
   iv. Use of known tagged predator data
2. Metrics
   i. Metrics used
   ii. Summary functions (if applicable)
   iii. Transformations (if applicable)
3. Predator filter settings
   i. Details for replicating pattern recognition analysis (e.g., clustering algorithm, distance matrix, number of groups, etc.)
   ii. Rule set (if using rule-based filter)
   iii. Weight-of-evidence threshold (if applicable)

Recommendations

- Recommendations provided here are intended to be widely applicable for salmon survival studies in the Central Valley and Delta

  - General recommendations provided here are also suitable for salmon survival studies in other settings
- The recommendations are not prescriptive – there may be better approaches in some cases
- Researchers should review the recommendations in the context of their study objectives and available data

General

1. Researchers should make tag predation diagnosis and filtering a formal, systematic, repeatable, and documented component of analysis of telemetry data
2. Use detections from all receivers available for diagnosis of tag predation, including receivers not otherwise used in the primary analysis (e.g., survival modeling)
3. Document all stages, decisions, and code used in the predator filter process
4. Perform a sensitivity analysis

Prepare data

1. Clean your tagging, deployment, and detection data before starting the predator filter
2. Aggregate detection data to the detection event level for tag predation diagnosis
3. Use the same value of max-delay for the smolt-tag data set and the known predator tags
4. Avoid using general locations whose detection field varies widely over the study period
5. Construct a flowline map for implementing a spatial framework
6. Review the flowline map and metrics that depend on it; errors can be introduced by automated interpretation of spatial relationships or by manual construction of the map
7. Validate the spatial and temporal structure of your data before implementing a predator filter (look for negative values or inconsistent metric values, review general location definitions)

Metrics

1. Review the “Metrics” section before defining your own detection metrics
2. Define metrics in such a way that limits dependence on spatial construction of receiver network (e.g., network density, size of general location detection field)
3. Metrics should be defined for every detection event and tag, depending on the scale; avoid NAs
4. Convert counts or absolute measures of rare behavior to the proportion scale in order to reduce the sensitivity to zero-dominant metrics and to variations in detection history or travel length
5. Categorical metrics should be converted to numeric metrics for analysis
6. Center and scale metric data for multivariate (pattern recognition) methods; not recommended for expert-opinion rule-based filters
7. Summarize cumulative metrics using the maximum and non-cumulative metrics using the mean for tag-level analysis
8. Avoid use of highly correlated metrics in rule-based filter

9. Review distribution of metrics before implementing filter

Select filter

1. Review “Considerations” under “Choose a filter” when selecting among possible filters
2. Avoid filters that assign predation based on a single metric

Predator data

1. Use concurrent known predator data instead of non-concurrent known predator data when available
2. Consider using non-concurrent known predator data if distinguishing factors (season, region, time period, tag technology) are not likely to affect predator behavior (comparable predator data)
   i. Identify comparable predator data by evaluating predator data sets for behavior differences represented in multiple metrics rather than a single behavior metric
   ii. The metrics used in the predator filter do not need to be the same as those used in identifying comparable predator data as long as those used in the predator data analysis represent a range of behavior characteristics
   iii. Recommended metrics to consider are in identifying comparable predator data are: migration rate, residence time, and upstream movement
   iv. Evaluation of available predator data from the Central Valley found differences in predator behavior associated with season and water year type

b. The “DPF” R package assists the user in identifying concurrent and comparable predator data stored in the package for a given smolt tag data set: see functions “lookup_pred_metrics” and “get_pred_dedf”

3. Beware of placing too much dependence on small data sets of concurrent known predator data
4. Do not compare data sets that use different max-delay settings in construction of the event-level data

Implement filter and review output

1. Use appropriate summarization and transformation of metrics for your chosen filter
2. Implement filter via replicable code set rather than manually or via ad hoc decision process
3. Review the results of the predator filter: do they make sense?
4. Researchers should review filter output for reasonable results even if known tagged predator data are used
5. Perform a sensitivity analysis of the predator filter output to decisions made during the filtering process

6. Examine the performance of the predator filter on data from known tagged predators (if available)
7. Report summary statistics of predator filter output: how many tags were flagged as predated, where (if applicable)

Acknowledgements

This document was created as part of the project “Standard Operating Procedure for Diagnosing and Addressing Predator Detections in Salmon Telemetry Data.” Funding came from the U. S. Bureau of Reclamation through the Delta Science Program (Delta Stewardship Council) under the 2021 Delta Science Awards (agreement DSC-21032). Development and testing of tag predation methods was performed in collaboration with Russell Perry, Adam Pope, Dalton Hance, Michael Dodrill, and Jake Kelley at the U. S. Geological Survey. Predator tagging data were compiled by Cyril Michel (NOAA, University of California Santa Cruz). We are grateful to the many fisheries biologists and telemetry technicians in the Central Valley and Delta who have provided their telemetry data and expressed interest in this project.

Appendices

Appendix A: Details of Predator-filtering Approaches

Pattern recognition methods

Hierarchical cluster analysis

Ward (1963) described a procedure called hierarchical cluster analysis to iteratively group n individuals, initially each in their own group, by first combining the two 'most alike' members resulting in $n-1$ groups, and then repeating the procedure until the desired number of mutually exclusive groups remained. Determining which groups were 'most alike' during any given iteration of this procedure is achieved by minimizing loss of an objective function and is referred to as Ward's minimum variance method.

Gibson et al. (2015) used hierarchical cluster analysis to determine which tags implanted into Atlantic salmon smolts had been consumed by striped bass. The analysis described in Gibson et al. (2015) involved three primary steps: 1) select metrics, 2) select an appropriate dissimilarity matrix, and 3) perform the clustering of metrics into the desired number of groups. For their analysis, the 11 metrics selected were computed for each tag (for both tagged smolts and predators) to form a multivariate matrix with $n =$ number of tags rows and $i = 11$ metrics columns. Each column was standardized to negate the varying effect differing metric scales would have on the cluster analysis. The cluster analysis operated on clustering around the minimum Euclidean distance as defined by the standardized multivariate vectors between groups and was performed using the function 'hclust' in the 'stats' R package (R Core Team 2024).

Because the hierarchical cluster analysis method as described in Gibson et al. (2015) determines only which tags were likely predated and not at what point in its capture history a tag was predated, it is necessary to first summarize each dataset from event-level to tag-level. The tag-level data thus contain a single row for each tagged fish, with each performance metric column summarized over the capture history for each tag. In general, metrics that are cumulative in nature are summarized by taking the maximum over each tag's capture history, while non-cumulative, event-level metrics are summarized by taking the mean value over each tag's capture history. It is reasonable to analyze the smolt-tag data set first alone and then combined with known tagged predators in order to assess the impact of known tagged predator data on the fate assignments for the smolts tags.

Decisions required by the analyst include:

- The suite of metrics to include
- The number of clustered groups to use (≥2)
- Transformations of metrics
- Distance measure used in the distance matrix (e.g., Euclidean distance)
- How to interpret the fate of the tags in each cluster (predated vs. not predated):
    - For smolt-only datasets in the absence of known predator reference data, the largest cluster (group) may be interpreted as non-predated and all other groups as predated under the assumption that the majority of smolt tags are not detected as predators.

- For datasets containing both tagged smolts and tagged predators, the group with the largest ratio of smolt tags to known predators may be assigned the "non-predated" status and all other groups the "predated" status.
- A possible expansion of the clustering performed using the combined smolt and predator data is to perform a second-stage clustering of the smolt-only data in which the outcome of the combined smolt and predator analysis (i.e., first-stage cluster) is used to inform the clustering. In the second-stage clustering outcome, the group with the highest ratio of non-predated to predated tags (using labels from the first-stage clustering) is classified as non-predated in the second stage, while all other second-stage groups are classified as predated.

The USGS has used this technique to identify potentially predated salmon smolts in multiple studies (Perry et al. 2018, Pope et al. 2021) and has found several particular metrics useful in differentiating salmon smolts from predators in the North Delta, including number of upstream transitions, total time in the study area, and average rate of downstream movement.

Gaussian mixture models

Gaussian mixture models (GMMs) are a model-based clustering approach that assumes data arise from groups that follow multimodal Gaussian distributions. GMMs allow for heterogeneous variance-covariance matrix shapes, volumes, and orientations, providing flexibility in modeling clusters. This method has been used to distinguish predators (bass species) from Chinook Salmon smolts using track-level, spatially-explicit movement metrics (turning angle and movement rate; Romine et al. 2014).

Maximum likelihood estimation of the distributional models is conducted by an expectation-maximization (EM) algorithm. If the number of groups (*G*) is supplied, the GMM will partition the data into *G* distinct clusters and estimate the means, variance-covariance matrices, and probability of each data point belonging to each group. An advantage to GMMs is that *G* may alternatively be determined through a model selection process (e.g., based on BIC) if the number of groups is not known.

As with the Hierarchical Cluster Analysis method, event-level data are summarized per tag prior to applying this clustering method. Summary functions calculate either the maximum value or mean of each metric per tag; users may consider using the geometric mean instead. To minimize the effects of rare events on metric distributions, proportions (of the total number of events) are used rather than counts of rare events. Tags with missing data must be excluded from the analysis, so metrics with many missing values may need to be dropped from analysis; GMMs are not robust to missing data (although approaches could be taken to address missing points; see R package "MGMM").

The estimation routine for GMMs can be highly sensitive to the cluster initialization. Often, hierarchical model-based agglomerative clustering is used to provide initial clustering partitions but these can be problematic if data are not continuous or are coarse (Scrucca et al. 2016). A scaled singular value decomposition (SVD) transformation on the input data can be used to address this potential issue, following the advice of Scrucca and Rafferty (2015). This is the approach used in the "gmm_filter" function in the "DPF" R package (Whitlock et al. 2025).

Issues/Considerations

- Overlap in the distributions of metrics of known tagged predators and tags deemed to be in smolts reduces the applicability of GMMs for diagnosing tag predation. GMMs need to be able to identify peaks in distributions for each cluster.
- GMMs assume that the data's true distributions arise from a mixture of Gaussian distributions. Thus, the further away the data are from meeting this assumption, the less reliable the method will be for classification.
- GMMs can help determine the number of clusters that best fit the data. If it is suspected that clusters may segregate on species (e.g., multiple predator species plus smolts), then the user may choose to not specify the number of groups *G*. However, GMMs may identify too many groups to be useful, especially if applied to a combined smolt-tag and known predator data set. It may be more appropriate to set *G* to either 2 (smolts, predators) or 3 (smolts, and two types of predators) instead.
- Users of these approaches will need to assign each cluster to be either predator or smolt, using an understanding or expectation of how metric values generally align with these group identities. Visualizing the distributions of individual predator filter metrics and comparing these distributions to known predator and/or smolt metric values for the purpose of assigning clusters to these groups is key. Both the choice of *G* and the assignment of cluster identity will greatly influence how many and which tags are labeled predator vs. smolt.
    - A more straightforward option that is available when there are some known predators in the dataset is to assign the cluster with the highest ratio of smolt tags to known predators as the "non-predated" cluster and all other clusters as the predator clusters.

Robust Principal Component Analysis

The Robust Principal Component Analysis (RPCA) pattern recognition method applies an alternative form of principal component analysis to data sets of tag-level metrics and identifies outliers, which are considered to be predated tags. Traditional principal component analysis is a well-known dimensionality reduction and exploratory data analysis technique that is frequently applied to data sets with variables that are known to be correlated in an attempt to look beyond any collinearity and reveal key insights about the underlying structure of the data set. In a nutshell, PCA is the eigendecompostion of the covariance matrix of a data set into orthogonal vectors known as principal components (PCs), which are then ordered decreasingly based on the proportion of total variance that they describe (e.g., PC1, PC2, etc.). A major weakness of traditional PCA is that it can perform poorly when applied to datasets with outliers that are the result of contamination. The RPCA technique is an alternate version of PCA that uses resampling methods and robust measures of scale and spread to compute estimates and can account for a high degree of contamination (>50% of observations). Our RPCA implementation also makes use of two robust measures of outlyingness described in Hubert et al. (2009) including: (1) score distance, which measures how far an observation is from the center of the data's distribution along the most important (i.e., lowest ranking) PC axes; and (2) orthogonal distance, which measures the degree to which a given observation is projected in a direction orthogonal to the major axes defined by the full data set (PCs 1 and 2). Cutoff values for classifying observations as outliers based on either measure are determined by specifying quantiles of theoretical distributions (i.e., tolerance ellipse) or based on the empirical data. A modified version of the outlyingness criteria can be computed if some of the input distributions are skewed.

The rpca_filter() in the "DPF" R package (Whitlock et al. 2025) function carries out the RPCA and creates an object containing tabular output and standard RPCA plots. Tags with measures that exceed either the score or the orthogonal distance cutoff value are automatically classified as predators. After taking the time to understand the data structure and the reason that tags were flagged, the analyst can either decide to accept the results of the filter at face value or use their new understanding of the data structure to identify what they consider to be predated tags using an alternative strategy. The RPCA is only carried out once, but the outputs from the procedure have several further uses. It is possible to compute RPCA coordinates and outlyingness measures for new data sets with the same set of metrics based on the estimated eigenvalues and eigenvectors. The RPCA output is used to identify where in a smolt tag's detection history a predation event is estimated to have occurred. This is accomplished by subsetting and summarizing increasing portions of the tag's event-level metric data chronologically, transforming the resultant matrix of tag-level summaries into scores using outputs from RPCA analysis, and then assigning the tag predation event to that point in the tag's history where a cutoff was first exceeded. A further use of the RPCA output is that it enables analysts to examine the extent to which tags from a predator reference data set would be characterized as an outlier by the RPCA.

Classification and regression tree

Classification regression tree (CART) models and random forest are a form of machine learning that partitions data into categories. These models construct a classification tree using categorical and/or continuous predictor variables by repeatedly splitting data into homogeneous groups (see Breiman et al. 1984; De'ath and Fabricius 2000; Krzywinski and Altman 2017). At each step, data are split into two mutually exclusive groups using simple rules to make each group as homogeneous as possible (Breiman et al. 1984; De'ath and Fabricius 2000; Krzywinski and Altman 2017). A tree is constructed by applying this splitting procedure to each group (and subsequent groups) separately to partition the response variable into a set of homogeneous groups (Breiman et al. 1984; De'ath and Fabricius 2000; Krzywinski and Altman 20174). Random forests are constructed by developing many CART models (e.g., trees) using a random subset of the data and averaging the results (Breiman 2001; Cutler et al. 2007). Regression trees are powerful models that can handle many predictor variables, are also robust to outliers, do not require transformation of variables, and are able to capture non-linear relationships and complex interactions (Moisen 2008). Further, they allow users to use many predictor variables within the model without a-priori choosing which predictor variables to include/exclude in the model.

There are some challenges in using CART models to identify tag predation within smolt telemetry data. Regression tree models require known smolt and predator tracks to initially develop the classification models. While users may have known predator tracks (i.e., known tagged predator data), use of conventional acoustic tags in a smolt survival study does not provide known smolt data because the smolt datasets contain a mixture of smolts and predators. This is further complicated because we do not know when smolt tracks transition to predation tracks. Pursuing these methods in the future would be useful for developing a predator filter. One way to do this would be to use predator detection tags (PDAT) developed by InnovaSea (see https://www.innovasea.com/fish-tracking/applications/predation/) to ensure known smolts are in the training datasets. Once regression tree models are developed using both known smolt and known predator tracks, we can apply these classification models to historical smolt telemetry data and compare results to other predator filtering methods.

Rule-based filters

A rule-based event-level data filter can be constructed to flag detection events considered to represent predated tags. For each detection event, the observed value of the metric is compared to a pre-specified threshold value(s) and the number of metrics whose observation exceeded the threshold is recorded ("predator score"). Detection events whose predator score exceeds a pre-specified critical (i.e., minimum) predator score are classified as predated on arrival of the general location observed in the detection event. All subsequent detection events for the tag are also classified as predated. Critical thresholds can be defined by expert opinion, literature review, simple mathematical models, sensitivity analysis, or a combination thereof. Alternatively, critical thresholds may be defined based on features of the observed distribution of metrics, such as the most extreme value that is not considered an outlier.

The critical predator score required for classification of a detection event as predated is a weight-of-evidence threshold for tag predation assignment. Buchanan and Whitlock (2022) set this value at 2 for the South Delta, meaning that at least two metric thresholds must be violated for a single detection event in order to classify that event as predated. Requiring at least two observed metrics beyond their thresholds allows for some variability in smolt behavior on an individual level and as they move through diverse habitats. The sensitivity of the filter outcome to the minimum predator score threshold should be investigated. Sets of metrics that include high levels of pairwise correlation will require higher minimum predator score thresholds to avoid double-counting a single tag behavior in computing the predator score.

Decisions required by the analyst include:

- Suite of metrics to use
- Critical thresholds for each metric, including whether the threshold is a maximum or a minimum or both
- Whether to include a spatial component to the critical threshold rule set
- The minimum predator score (weight-of-evidence critical threshold) for classification as predated
- Precision of critical thresholds
- Whether to explore alternative values of critical thresholds, and if so, what are they

Appendix B: Definitions of Metrics

Definitions for a selection of spatiotemporal detection history metrics are provided here. Additional metrics may be defined—see "Metrics" section in body of document for considerations in defining metrics.

Table 1. Spatiotemporal metrics defined in this document. Metrics are computed using general location-detection event data unless otherwise noted. Other metrics may be used.

Metric	Category	Definition/Additional information
Time Since Release	Time of movement	Defined as time since first detection within study period if tag was released before study period (e.g., known predators)
Time Between*	Time of movement	Time lag between start of prior detection event and start of the current detection event
Time Since Last at General Location	Time of movement	Time delay between start of the most recent detection event at the current general location to the start of the current detection event
Near-field Residence Time	Residence time	Duration of detection event
Mid-field Residence Time	Residence time	Time lag from first to last detections of string of consecutive detection events at general location
Visit Event*	Residence time	Numeric index of strings of unbroken detection events
Cumulative Proportion of Time in High-risk Zones	Residence time	Starts at estimated entrance of high-risk zone, assuming constant rate of movement
Cumulative Residence Time (Near-field)	Residence time	Total near-field residence time to date
Transition Length*	Distance	Estimated distance between prior detection event and current detection event, measured along the shortest route
Migration Rate	Migration rate	Measured between consecutive detection events
Body Lengths per Second (BLPS)	Migration rate	Body Lengths per second; use mean smolt length in computation of metric for predator data set
Transition Type*	Movement pattern	Categorical variable with values "downstream", "upstream", "repeated", and "lateral" (or NA)
Cumulative (Running) Count of Detection Events to General Location	Movement pattern	Count continues if tag returns to general location after detection elsewhere
Count (Running) of Consecutive Detection Events at General Location	Movement pattern	Count restarts if tag returns to general location after detection elsewhere
Cumulative Proportion of Transitions that Switch from Downstream to Upstream Movement	Movement pattern	Increments for each switch from downstream to upstream; ignores lateral transitions and repeat consecutive visits to same general location
Detection in High-risk Zone*	Movement pattern	Logical variable: was the detection event in a high-risk zone?
Transition under Unidirectional Flow*	Movement/ environment	Logical variable: was the transition in a region with unidirectional water flow?
Against Flow*	Movement/ environment	Logical variable: was the transition known to be directed against the water flow?
Cumulative Proportion of Transitions Against Flow	Movement/ environment	Increments for each additional transition known to be directed against the direction of river flow

* Metrics marked with asterisk (*) are used in computing other metrics and may be unsuitable for direct inclusion in some tag predation diagnosis methods.

Time Since Release

Time Since Release is used to distinguish between migrating smolts and resident predators, with the expectation that tags detected especially long after release were more likely to come from predators than from smolts. Similar metrics have been used in previous studies (e.g., Gibson et al. 2015, Perry et al. 2018, Buchanan et al. 2021, Buchanan and Whitlock 2022). Because the acoustic tags used in some predator tagging studies lasted multiple years and the objective of the predator data analysis is to characterize predator behavior for use in tag predation diagnosis for single-season smolt tagging studies, Time Since Release is defined for predator data using a virtual release date equal to either their first detection during the smolt tagging study (if the predator was released before the smolt study) or their actual release date (if the predator was released during the smolt study).

Time Between

Time Between measures the time lag between the start of the previous detection event and the start of the current event (fundamental metric).

Time Since Last at Station

The metric Time Since Last at Station is used to distinguish smolts from predators that are either largely residential with home ranges near but not including a telemetry station (i.e., general location) or are highly pelagic with a cycling behavior throughout some part of the study area, both of which patterns may result in lengthy time gaps between successive detections at a single station.

Near-field Residence Time

Near-field Residence Time is defined as the duration of a single detection event at a general location (fundamental metric). It includes detection gaps of up to a user specified maximum ("max_delay") between the start and end of the detection event, conditional on the tag not being detected elsewhere.

Mid-field Residence Time

The Mid-field Residence Time metric measures the total elapsed time from the start of the first detection event at a general location to the end of the current detection event at that general location within a sequence of consecutive detection events at the same general location unbroken by detection elsewhere. It includes both the Near-field Residence Time during the included detection events and the time between the detection events (in this case, equal to Time Since Last At Station).

Visit Event

Visit Event is a numeric index of the sequences of unbroken detection events at a given general location without intervening detections elsewhere. It is used for identifying events reflected by Mid-field Residence Time.

Cumulative Residence Time

The Cumulative Residence Time metric measures the running total of the Near-field Residence Time metric for successive detection events at the same general location (Perry et al. 2018). It differs from the Mid-field

Residence Time by omitting the time between successive visits and by including return detection events at the general location even after detection elsewhere.

Cumulative Proportion of Time in High-risk Zones

Daniels et al. (2018) used the time spent on defined spawning grounds of Striped Bass Morone saxatilis as one metric for assessment of predation of Atlantic Salmon smolts Salmo salar. A similar metric can be used to track how long smolt tags are estimated to be present in zones of high predation risk compared to an assumed baseline predation risk for the study area ("high-risk zones"). High-risk zones must be specified (e.g., based on expert opinion; Appendix D).

Residence time within the high-risk zones can be estimated using the mapped extent of the high-risk zones. The entry and departure times into and out of the high-risk zones are estimated under the assumption of a constant movement rate through the surrounding reaches delineated by the telemetry receivers in the vicinity. The metric is computed by summing the near-field residence times at stations located within high-risk zones and the estimated travel time within high-risk zones while transitioning to or from a site outside of the zone. Travel time is estimated using linear interpolation using the known segment lengths. The cumulative residence time increases for every transition that includes parts of one or more high-risk zones.

The cumulative residence time is converted to the proportion of time since release spent in high-risk zones in order to limit the incidence of zeros in the metric data.

Cumulative Count of Detection Events at General Location

The Cumulative Count of Detection Events at General Location metric is defined as a running count of visits to a given general location, regardless of whether the tag is detected elsewhere between successive detection events at the general location in question.

Count of Consecutive Detection Events at General Location

The metric Count of Consecutive Detection Events at General Location (Perry et al. 2018) differs from Cumulative Count of Detection Events at General Location by counting only those detection events that are unbroken by detection elsewhere. Thus, the Consecutive Detection Events at General Location is lower than the Cumulative count of Detection Events at General Location if the tag is detected at a general location, then at a different general location, and then returns to the first general location.

Transition Length

Transition Length estimates the spatial extent of the transition of the tag from the previous detection event to the current detection event (fundamental metric). Distances are estimated using a flowline map and a distance metric on the flowline map (e.g., using the 'riverdist' package in R; R Core Team 2024, Tyers 2016). Briefly, the receivers that compose a given general location are associated with the centroid of the receiver locations for that general location, the centroids are "snapped" to the closest node along the river flow line, and the distance between the nodes associated with the previous detection event and the current detection event is measured along the shortest available route. Available routes are identified by daily comparison of date and river flow conditions with barrier or gate status and flow thresholds associated with accessibility of temporarily available routes (e.g., overtopping of a weir).

Migration Rate

Migration Rate is defined as the ratio of Transition Length to Time Between (units = km/day). In cases in which the tag is observed at the same general location on consecutive detection events (i.e., no apparent travel) and in which the release site is included within the detection range of the first detection site, the Migration Rate value is assigned NA.

This definition of Migration Rate is scaled to the transition level. Migration Rate can also be defined as the average rate since release.

Body Lengths per Second (BLPS)

Body Lengths Per Second (BLPS) is an alternative Migration Rate measure that is scaled to the size of the smolt at the time of tagging. It is reasonable to include under the expectation that movement rates may be related to fish size for some stocks, with larger fish moving faster (Giorgi et al. 1997). Even if size differences in movement rates are not obvious among individual smolts, it is possible that some species and life stages of predators, such as adult Striped Bass Morone saxatilis, have a noticeably higher movement rate when scaled by the average smolt size. Thus, for smolt tags, BLPS is computed using the observed fork length measured at the time of tagging. For predators, BLPS is computed using the mean fork length at tagging observed for the smolt data set under consideration.

Transition Type

A number of researchers have used observations of upstream movement as evidence of predation of the tagged smolts (e.g., Gibson et al. 2015, Daniels et al. 2018, Henderson et al. 2018, Johnston et al. 2018, Hause 2020, Buchanan and Whitlock 2022). In the Delta, the direction of each transition between detection events can be categorized as either “downstream,” “upstream,” or NA, or alternatively as “downstream,” “upstream,” “repeated,” and “lateral”). Repeated transitions are those in which the tag did not appear to move (i.e., consecutive detection events at the same general location). Lateral transitions are those in which a tag moves from one branch of a river junction to the other. Transition Type (fundamental metric) is used to compute the number of switches from downstream to upstream movement and identify movements against the flow.

Cumulative Proportion of Switches from Downstream to Upstream Movement

The Cumulative Proportion of Switches from Downstream to Upstream Movement metric is defined as a numerical representation of Transition Type and the extent of upstream movement or changes in movement direction exhibited by the tag. The expectation is that predators are more likely to exhibit upstream movement or bi-directional movement than are migrating smolts. A proportion is used to limit the overabundance of zeros in the metric observations.

Transition under Unidirectional Flow

Transitions that are known to be entirely in a region with unidirectional flow are identified and used in computation of the Against Flow metric. In the context of the Central Valley and Delta, unidirectional flow is interpreted as wholly downstream-directed flow, in contrast to bidirectional flow in tidally dominated regions of the Delta.

The extent of unidirectional flow is modeled simply as a function of Delta inflow and, for the South Delta, the presence of the rock barrier at the head of Old River. Regions are assumed to have unidirectional flow if they were upstream of the Tidal-Fluvial Transition Zone, where the river flow transitions to tidally influenced, bidirectional flow (Appendix D).

The value of the Transition Under Unidirectional Flow metric is either TRUE, if the transition occurred wholly within the modeled unidirectional flow region for the given day, FALSE if the transition occurred within the bidirectional flow region, and NA otherwise. All transitions within the tidally dominated regions are assigned NA values for this metric. A more completely defined metric that represents subdaily tidal flow in the Core Delta would require a hydrodynamic model of the Delta.

Against Flow

Several studies have used observations of transitions directed against the prevailing flow direction as evidence of predation (Perry et al. 2018, Hause 2020, Buchanan et al. 2021, Buchanan and Whitlock 2022). The Against Flow metric is defined as a logical variable indicating whether the transition is known to be directed against the direction of river flow. It is defined only within regions of unidirectional flow; in regions of bidirectional flow, Against Flow is assigned NA. In unidirectional regions, Against Flow is defined as TRUE if the transition is directed upstream (Transition Type = Upstream), FALSE if the transition is directed downstream (Transition Type = Downstream), and NA otherwise.

Cumulative Proportion of Transitions Against Flow

The Cumulative Proportion of Transitions Against Flow is a numeric version of the Against Flow metric. It is defined as a running count of transitions known to be directed against the flow, scaled by the number of detection events so far. Values of NA in the Against Flow metric are treated as FALSE in computing the Cumulative Proportion of Transitions Against Flow. As a result, all values of this metric will be 0 if the entire release and detection history of the tag is limited to the tidally dominated regions of the Delta.

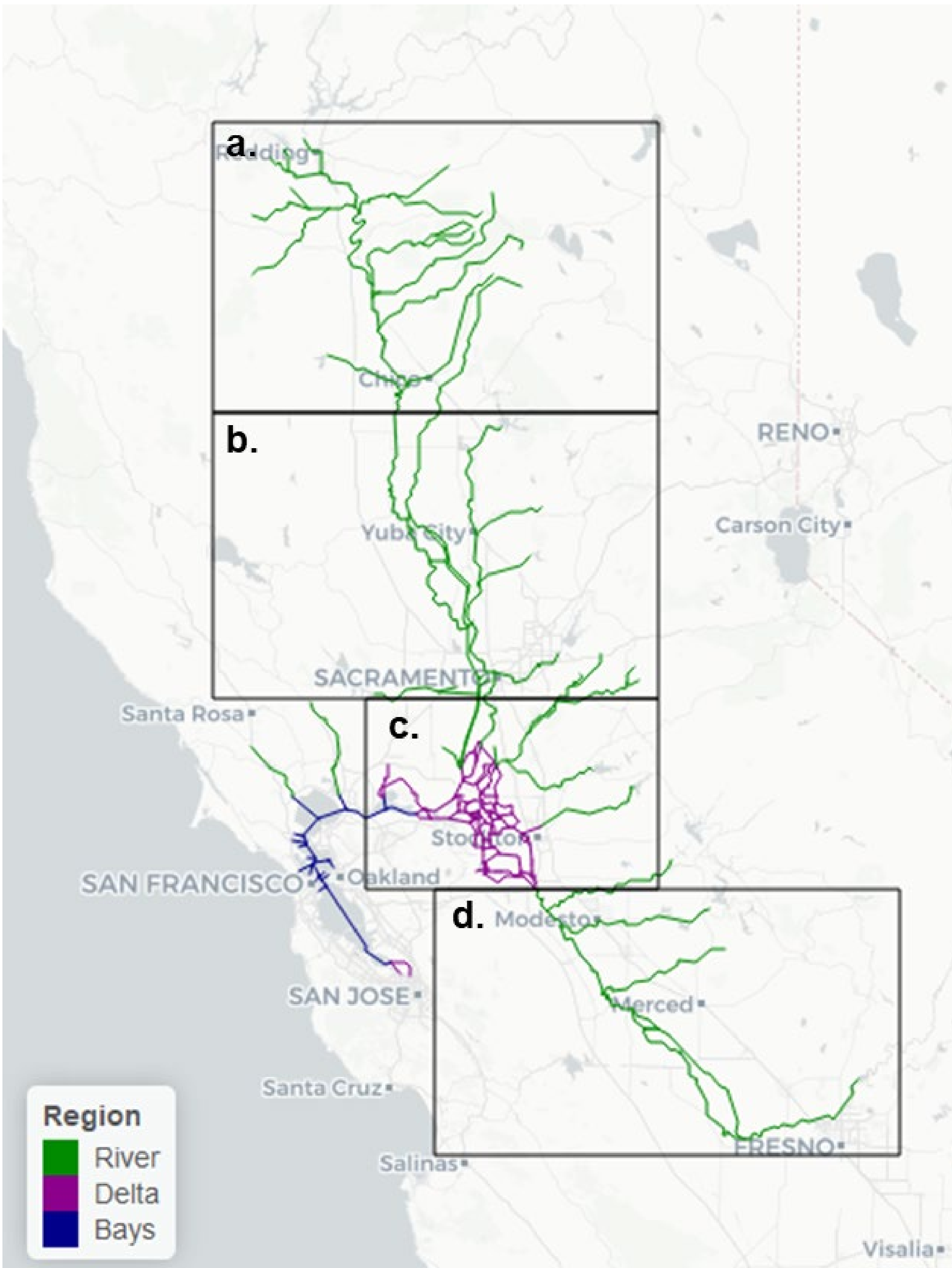
Appendix C: Flowline Network

- Various methods are available for constructing the flowline map
    - Manual digitization of flow lines
        - Use existing software (e.g., ArcMap or Google Maps) to measure the transition distances
        - Tedious but can check as you go
        - May use software to identify routes for measuring transition length
    - Flowline network approach
        - Develop a flowline layer
            - Consists of nodes and segments connecting the nodes
        - Snap general locations to the node network
        - Fast and automated but may introduce errors, requires careful review of outcome
        - Use software to identify routes for measuring transition length
            - Warning: errors in the flowline map cause problems
- Challenges
    - Ambiguity in routes
        - Options:
            - Use shortest available route
            - Average over all available routes
        - Document which approach was used
- Document all decisions made in constructing flowline network and assigning routes

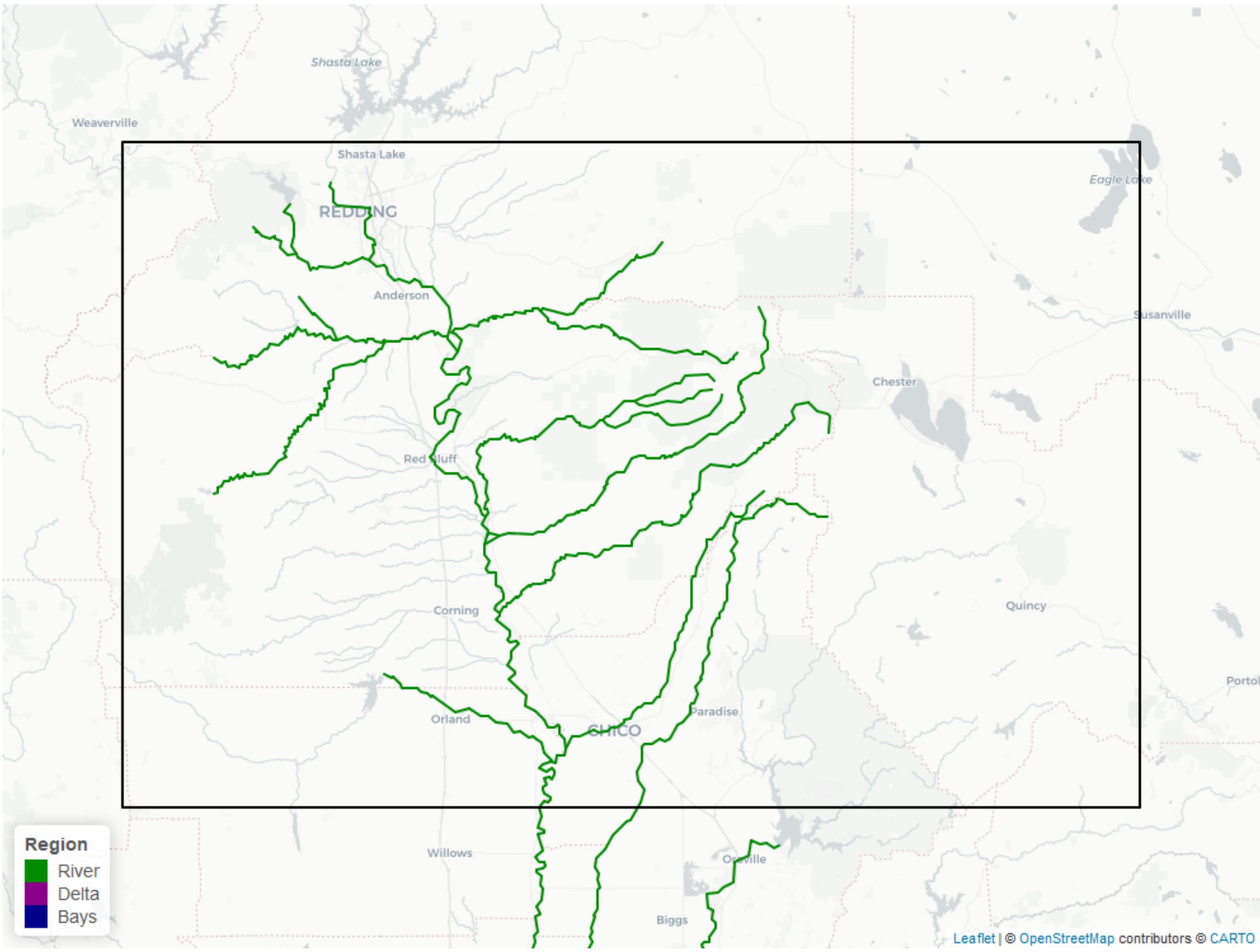
Appendix D: Maps

Flowline map

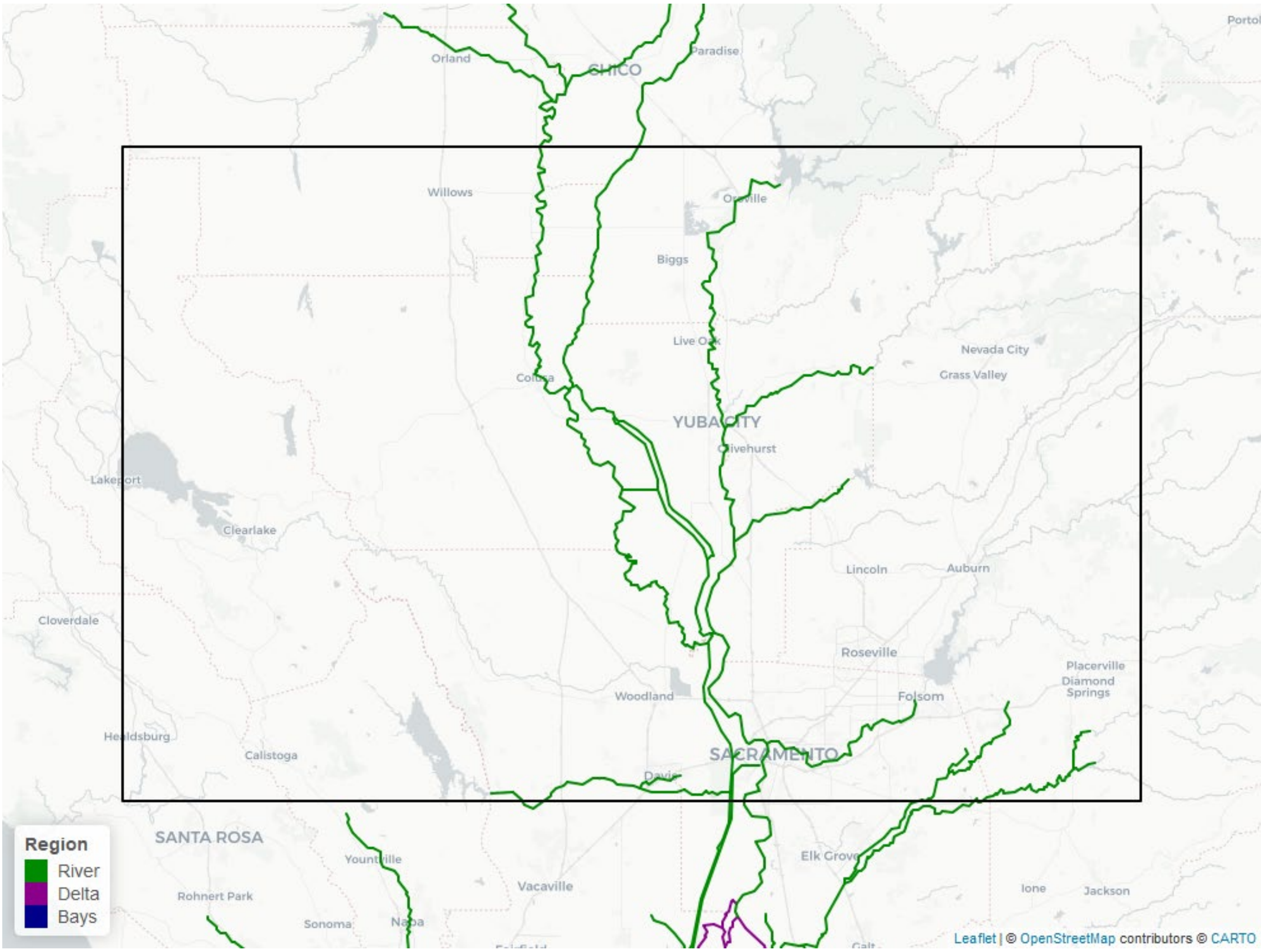
Full extent of the flowline map used in estimating detection metrics related to routes, distances, and auxiliary variables for telemetry studies in the Central Valley. Areas outlined by boxes include the Upper (a) and Lower (b) Sacramento River basins, the Delta (c), and the San Joaquin basin (d). These areas are shown in greater detail on subsequent pages.



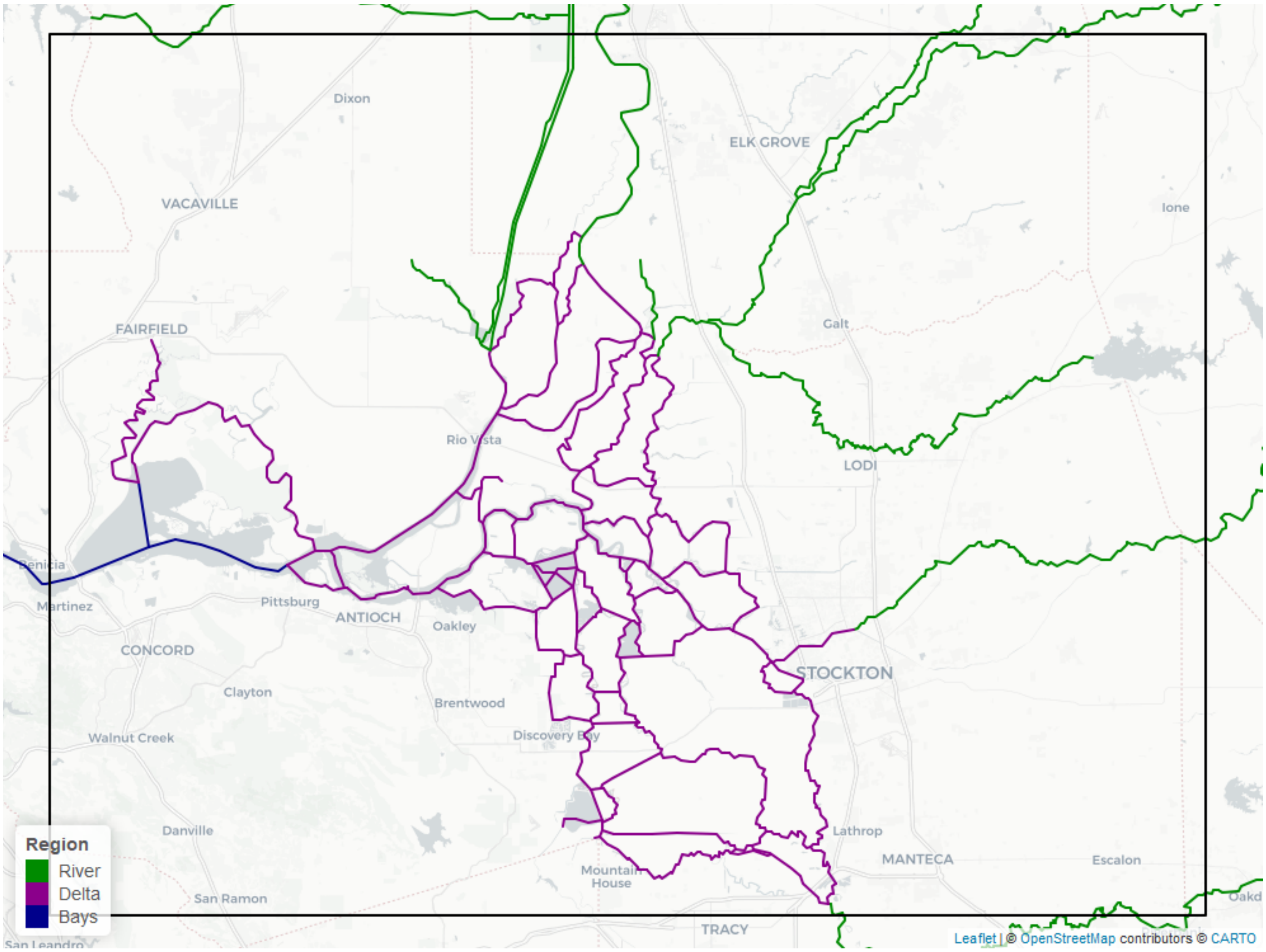
Flowline map—Upper Sacramento Basin (a)



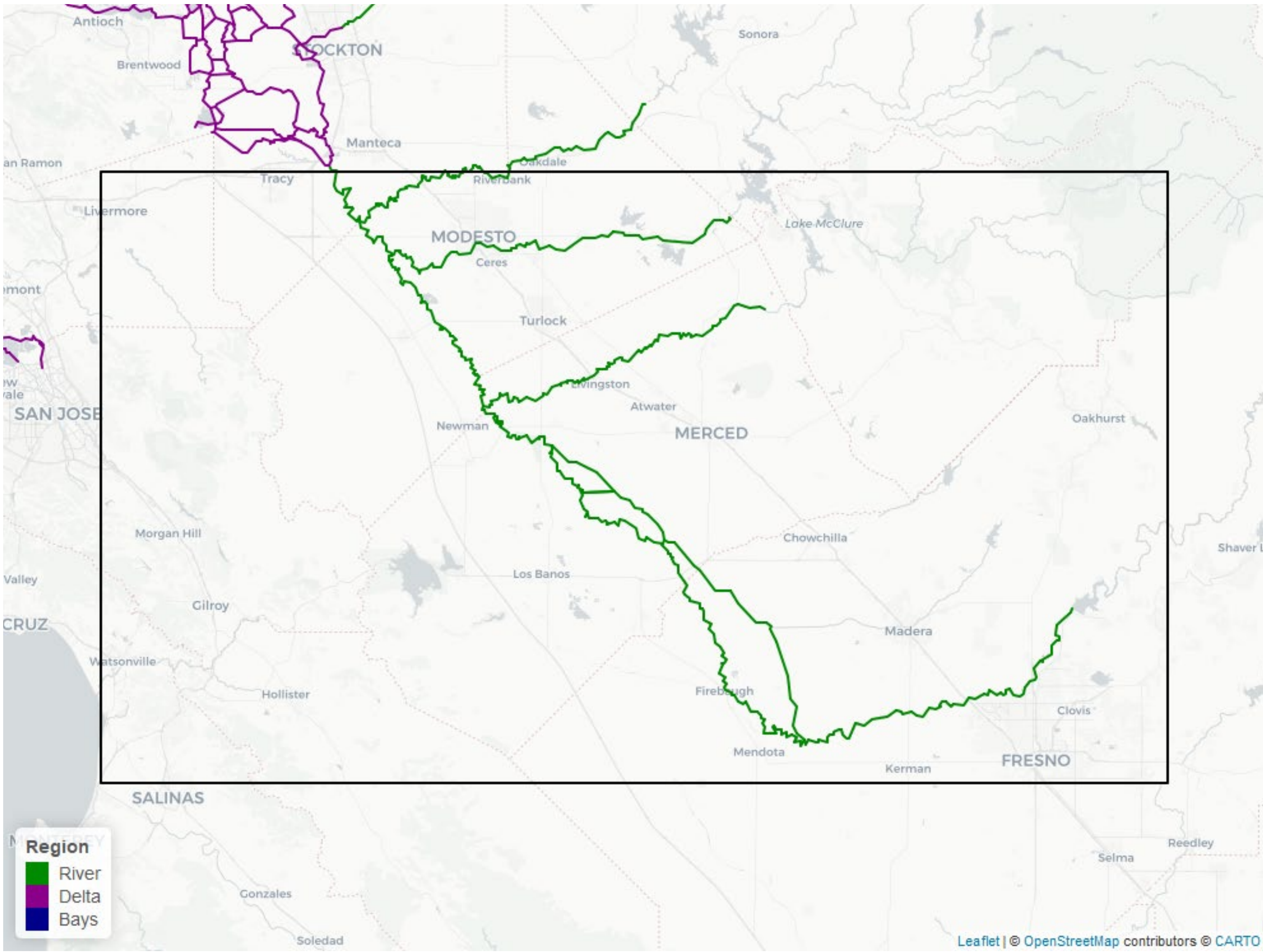
Flowline map—Lower Sacramento River Basin (b)



Flowline map—Sacramento–San Joaquin Delta (c)

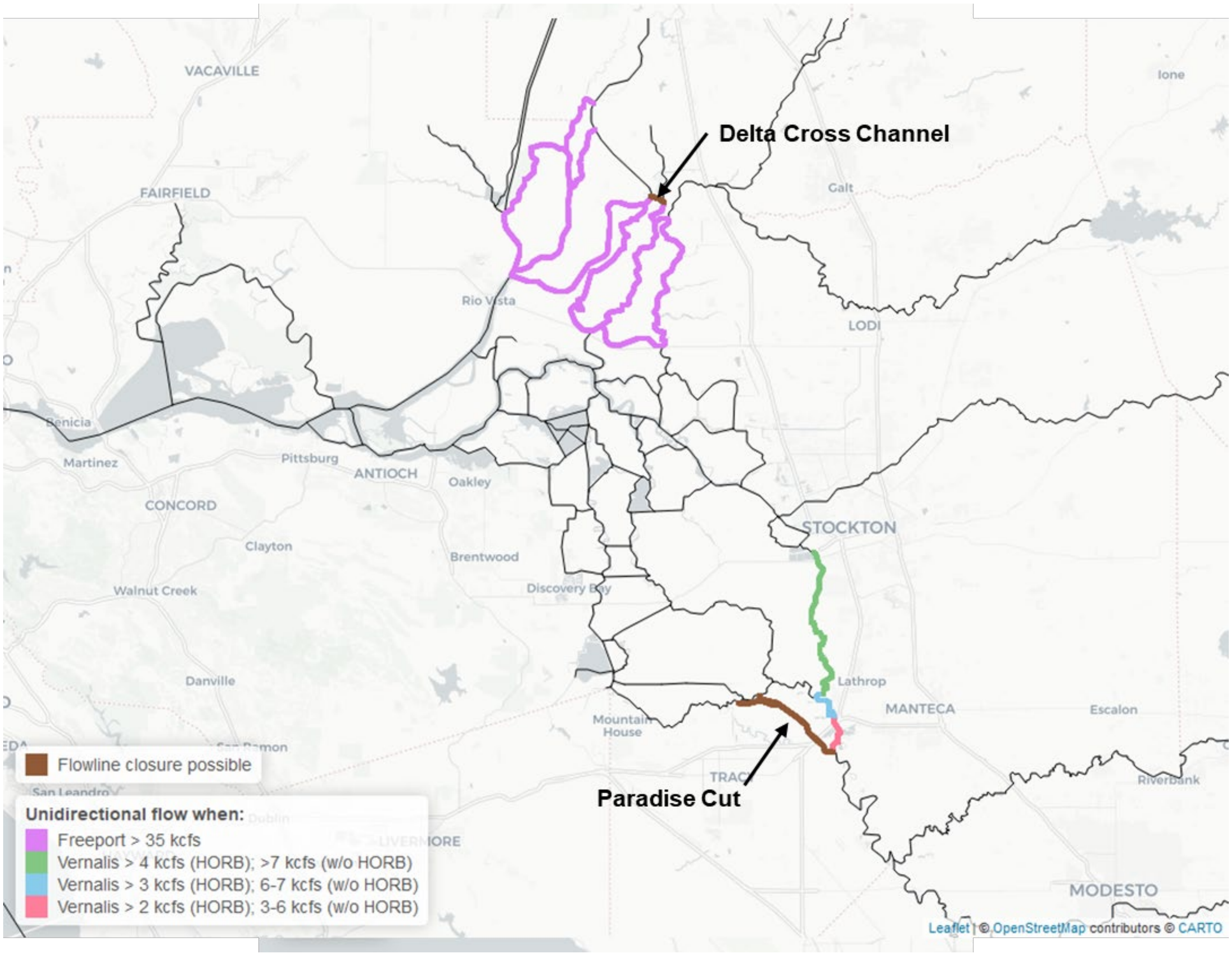


Flowline map—San Joaquin River Basin (d)



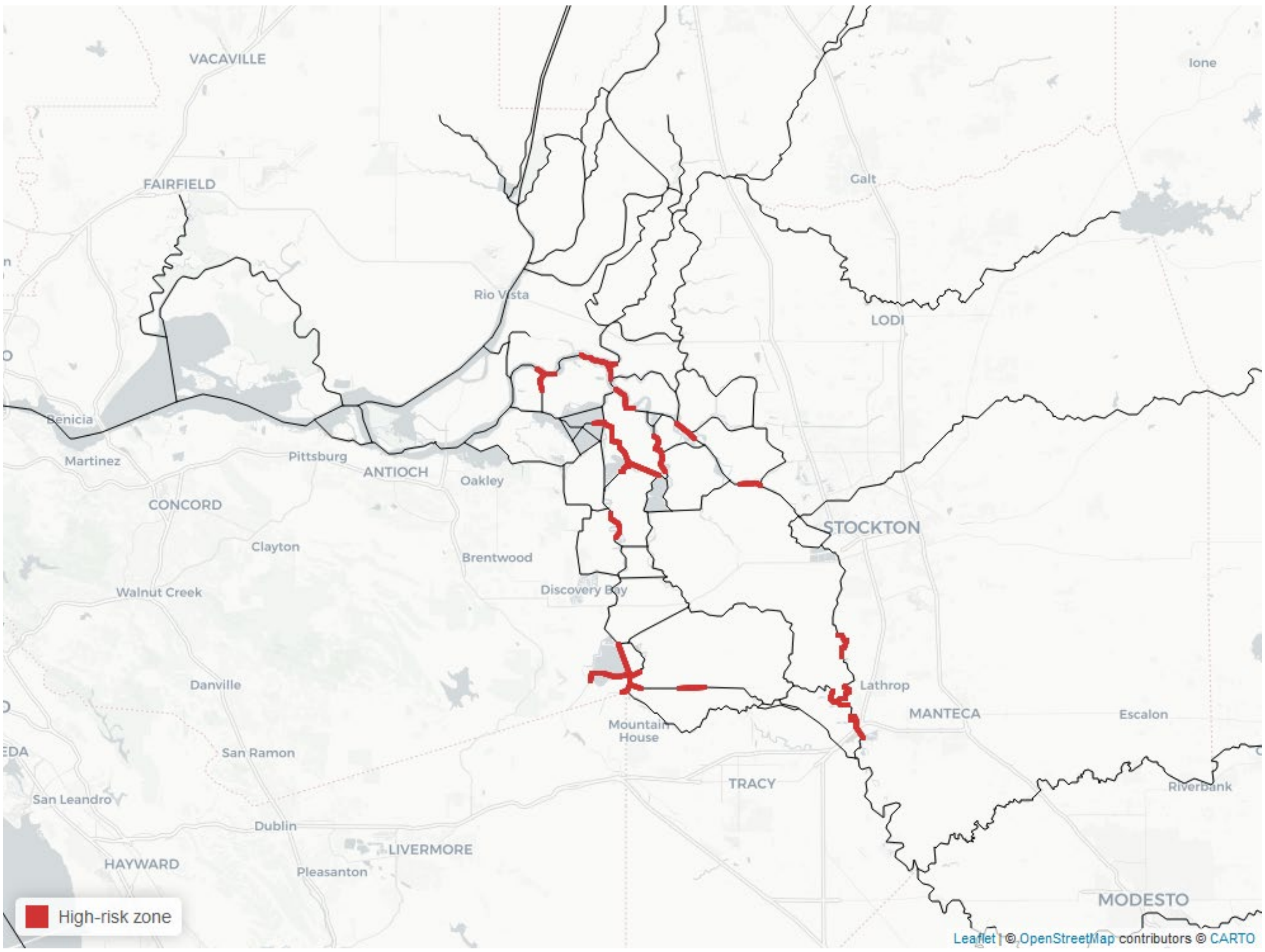
Flowline map—Barriers and Tidal-Fluvial Transition Zones (TFTZ)

Maroon colored segments denote segments that are accessible under specific flow or management conditions (“Flowline closure possible”) and segments that switch between bidirectional to unidirectional flow states based on discharge thresholds and barrier placement (Tidal-fluvial Transition Zones). The Paradise Cut segment is not affected by a barrier but is dewatered at low flows.



Flowline map—High-risk Zones

Red segments denote reaches where salmon smolts are assumed to be exposed to high predation risk. High-risk zones were delineated in the Core Delta and at the head of Old River based on a landscape-scale predation risk assessment reported by Michel et al. (2020). The entrances to the Central Valley Project and State Water Project, and the river channels in the nearby vicinity, were identified as high-risk zones (Schultz et al. 2015, Grossman 2016, Moyle et al. 2017). Also classified as a high-risk zone was the reach in Grant Line Canal that included the temporary agricultural barriers. Finally, the San Joaquin River reaches near Mossdale Bridge and Brandt Bridge were classified as high-risk zones (Vogel 2010). Additional high-risk zones can be added.



Appendix E: Considerations in Implementing a Predator Filter

In addition to the issues raised elsewhere in this document, the following questions should be considered by the analyst when implementing a predator filter. Brief answers are provided.

- Spatial considerations of general locations
  - What is the maximum spatial extent recommended for general locations (how wide)?
    - The wider the general location, the more imprecise the detection location is and the larger the error in the transition length and migration rate metrics.
  - How close together can general locations be?
    - The distance between the general locations should be considerably larger than the detection range of each general location.
  - How close to the release site can general locations be?
    - Investigation of the distribution of release lengths in South Delta data sets suggests that 2 km is a minimum distance to avoid biasing transition length and migration rate high.
  - What to do if the receiver network includes general locations that are closer together than recommended?
    - Either combine general locations or thin them to avoid spatial overlap.
  - Is it better to retain metrics that might be biased because sites are too close together, or to omit metrics that actually are not biased?
    - It is important to have an unbiased metric set.
- What to do if preconceived ideas of predator and smolt behavior contradict majority clusters?
  - Review both the preconceived ideas and the methods used in the predator filter.
- Should the filter method be spatially explicit (assign different criteria for different regions or define separate clusters for different regions)?
  - This may be more important in larger study areas in which study subjects experience widely different habitats with different predator communities.
- Should the filter be defined differently for different seasons?
  - This depends on the nature of the study subject behavior in different seasons and of the predator community in different seasons.
- Rule-based filter
  - Should the critical thresholds be defined based on environmental conditions, such as water velocity or temperature?
    - This should be assessed via a sensitivity analysis.
  - What to do if my “expert opinion” contradicts the observed data (e.g., all tags are flagged as predated)?
    - Return to the literature or additional experts to update your expert opinion.
- How many metrics should be used for predator filters? Is more better?
  - Using fewer metrics represents fewer types of possible predator behavior.
  - Only metrics that have a biological justification should be used.

- Using more metrics does not always result in either more or fewer flagged tags.

- Should pairs of highly correlated metrics be excluded from the metric set (e.g., migration rate and BLPS)?
    - This is more important for the rule-based filter than for pattern recognition methods.
- Is it recommended to use multiple diagnostic approaches? If so, how to combine them?
    - Yes, use multiple diagnostic approaches.
    - Compare results on the tags and/or detection events that are flagged.
    - It is reasonable to assign the predated status to those tags that are flagged by all methods (intersection in Euler diagram of flagged tags).
    - It may be reasonable to assign the predated status to those tags that are flagged by any methods (union in Euler diagram of flagged tags), but this depends on whether all methods are considerable equally suitable.
    - Methods that are of dubious worth should be omitted (e.g., assumptions are unlikely to be met or expertise is lacking).

Appendix F: List of References

Breiman L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. Classification and regression trees. Wadsworth and Brooks/Cole, Monterey, California, United States.

Breiman, L. 2001. Random forests. Machine Learning 41:5–32, https://doi.org/10.1023/A:1010933404324.

Buchanan, R. A., E. Buttermore, and J. Israel. 2021. Outmigration survival of a threatened steelhead population through a tidal estuary. Canadian Journal of Fisheries and Aquatic Sciences. 78(12):1869–1886, https://doi.org/10.1139/cjfas-2020-0467.

Buchanan, R. A., and S. L. Whitlock. 2022. Diagnosing predated tags in telemetry survival studies of migratory fishes in river systems. Animal Biotelemetry 10:13, https://doi.org/10.1186/s40317-022-00283-1.

Cutler, D. R., T. C. Edwards Jr, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. Random forests for classification in ecology. Ecology 88:2783–2792, https://doi.org/10.1890/07-0539.1.

Daniels, J., G. Chaput, and J. Carr. 2018. Estimating consumption rate of Atlantic salmon smolts (*Salmo salar*) by striped bass (*Morone saxatilis*) in the Miramichi River estuary using acoustic telemetry. Canadian Journal of Fisheries and Aquatic Sciences 75(11):1811–22, https://doi.org/10.1139/cjfas-2017-0373.

De'ath, G., and K. E. Fabricius. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81:3178– 3192, https://doi.org/10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2.

Gibson, A. J. F., E. A. Halfyard, R. G. Bradford, M. J. Stokesbury, and A. M. Redden. 2015. Effects of predation on telemetry-based survival estimates: insights from a study on endangered Atlantic salmon smolts. Canadian Journal of Fisheries and Aquatic Sciences 72(5):728–741, https://doi.org/10.1139/cjfas-2014-0245.

Giorgi, A. E., T. Hillman, J. S. Stevenson, S. G. Hays, and C. M. Peven. 1997. Factors that influence the downstream migration rates of juvenile salmon and steelhead through the hydroelectric system in the mid-Columbia River basin. North American Journal of Fisheries Management 17(2):268–282, https://doi.org/10.1577/1548-8675(1997)017<0268:FTITDM>2.3.CO;2.

Grossman, G. D. 2016. Predation on fishes in the Sacramento–San Joaquin Delta: current knowledge and future directions. San Francisco Estuary and Watershed Science 14(2), https://doi.org/10.15447/sfews.2016v14iss2art8.

Hause, C. 2020. Outmigration survival of juvenile spring-run Chinook salmon in relation to physicochemical conditions in the San Joaquin River. Master’s Thesis: University of California Davis.

Henderson, M. J., I. S. Iglesias, C. J. Michel, A. J. Ammann, and D. D. Huff. 2018. Estimating spatial-temporal differences in Chinook salmon outmigration survival with habitat- and predation-related covariates. Canadian Journal of Fisheries and Aquatic Sciences 76(9):1549–1561, https://doi.org/10.1139/cjfas-2018-0212.

Johnston, M. E., A. E. Steel, M. Espe, T. Sommer, A. P. Klimley, P. Sandstrom, D. Smith. 2018. Survival of juvenile Chinook salmon in the Yolo Bypass and the lower Sacramento River. San Francisco Estuary and Watershed Science 16(2), https://doi.org/10.15447/sfews.2018v16iss2art4.

Kelley, J. R., S. L. Whitlock, R. A. Buchanan, and R. W. Perry. 2022. Resource guide and literature review for addressing the problem of tag predation in salmonid studies in the Central Valley of California. Technical report to Delta Science Stewardship Council. 41 pages. Available: https://www.cbr.washington.edu/analysis/apps/tagpredation.

Krzywinski, M., and N. Altman. 2017. Classification and regression trees. Nature Methods 14:757–758, https://doi.org/10.1038/nmeth.4370.

Michel, C. J., M. J. Henderson, C. M. Loomis, J.M. Smith, N. J. Demetras, I. S. Iglesias, B. M. Lehman, and D. D. Huff. 2020. Fish predation on a landscape scale. Ecosphere 11(6), https://doi.org/10.1002/ecs2.3168.

Moisen, G. G. 2008. Classification and regression trees. In Encyclopedia of Ecology, volume 1, ed. S. E. Jørgensen, and B. D. Fath. Elsevier, Oxford, U.K., 582–588.

Moyle, P. B., R. A. Lusardi, P. J. Samuel, and J. V. E. Katz. 2017. State of the salmonids: status of California's emblematic fishes 2017. Center for Watershed Sciences, University of California–Davis, Davis, and California Trout, San Francisco.

Perry, R. W., A. C. Pope, J. G. Romine, P. L. Brandes, J. R. Burau, A. R. Blake, A. J. Ammann, and C. J. Michel. 2018. Flow-mediated effects on travel time, routing, and survival of juvenile Chinook salmon in a spatially complex, tidally forced river delta. Canadian Journal of Fisheries and Aquatic Sciences 75(11):1886–1901, https://doi.org/10.1139/cjfas-2017-0310.

Pope, A. C., R. W. Perry, B. N. Harvey, D. J. Hance, and H. C. Hansel. 2021. Juvenile Chinook salmon survival, travel time, and floodplain use relative to riverine channels in the Sacramento–San Joaquin River Delta. Transactions of the American Fisheries Society 150:38–55, https://doi.org/10.1002/tafs.10271.

R Core Team. 2024. R: a language and environment for statistical computing. R Foundation for Statistical Computing, http://www.R-project.org.

Romine J. G., Perry R. W., Johnston S. V., Fitzer C. W., Pagliughi S. W., Blake A. R. 2014. Identifying when tagged fishes have been consumed by piscivorous predators: application of multivariate mixture models to movement parameters of telemetered fishes. Anim Biotelem. 2(1):3, https://doi.org/10.1186/2050-3385-2-3.

Schultz, A. A., K. K. Kumagai, and B. B. Bridges. 2015. Methods to evaluate gut evacuation rates and predation using acoustic telemetry in the Tracy Fish Collection Facility primary channel. Animal Biotelemetry 3:13, https://doi.org/10.1186/s40317-015-0034-y.

Scrucca L., Fop M., Murphy T. B. and Raftery A. E. 2016. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal,* 8(1):205-233.

Scrucca L. and Raftery A. E. 2015. Improved initialisation of model-based clustering using Gaussian hierarchical partitions. Advances in Data Analysis and Classification, 4(9):447–460.

Tyers, M. 2016. riverdist: River Network Distance Computation and Applications. R package. Version 0.15.5 (December 31, 2021).

Ward, J. H. Jr. 1963. Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association 58(301):236¬–244, https://doi.org/10.1080/01621459.1963.10500845.

Whitlock, S., R. Buchanan, L. Yamane, A. Pope, and M. Dodrill. 2025. Delta Predator Filter (DPF). R package version 0.0.1.0, download DPF from: https://www.cbr.washington.edu/analysis/apps/tagpredation.

Vogel, D. A. 2010. Evaluation of acoustic-tagged juvenile Chinook salmon movements in the Sacramento–San Joaquin delta during the 2009 Vernalis Adaptive Management Program. Technical Report for San Joaquin River Group Authority. Available from https://www.waterboards.ca.gov/waterrights/water_issues/programs/bay_delta/bay_delta_plan/water_quality_control_planning/docs/sjrf_spprtinfo/vogel_2010.pdf. Accessed 5 June 2021.

Appendix G: Glossary

Behavior-based predator filter: A methodological approach for identifying tag predation that relies on contrast in movement capability, habitat use, or other tendencies between study subjects and predators.

Detection event: Aggregation of acoustic signal detections from an individual tag on one or more receivers uninterrupted by detections of the tag elsewhere or (optionally) by time gaps beyond a specified maximum duration (event time threshold; max-delay); individual events are identified by the tag, spatial scale, and timing, where spatial scale may range from a single receiver to full telemetry station, and timing is indicated by the first and last detection times.

Detection metric ("metric"): Statistics calculated from the tag's detection data and (optionally) auxiliary data and used to diagnose tag predation events within a predator filter. These may be continuous, ordinal, or categorical and are based directly on measured variables, including summarizations, transformations, and/or combinations of measured variables.

Event-level: Classification of a study subject's status as predated or not predated based on a summarization of attributes or apparent movements of a tag between discrete detection events; indicates timing of predation event in relation to tag's sequence of detection events; may also refer to a variable that is computed on the scale of detection events.

Flowline map: Collection of nodes and segments that connect the nodes, where the nodes represent deviations from linear water channels, including channel bends, flow splits, channel junctions, etc.

General location: A collection of one or more receivers grouped based on evidence or expectation of overlapping detection ranges.

High-risk zone: Reaches within a flowline map where study subjects are thought to be exposed to especially high predation risk.

Hybrid filter: Approach for diagnosing tag predation that combines pattern recognition approach with rule-based approach.

Known predator data: Detection data from telemetry tags that were implanted in fish considered to be predators of juvenile salmonids.

Max-delay: the longest time gap between detections at a single receiver or general location that is allowed for observations to be grouped within a single detection event.

Metrics: Time-varying variables representing the observed movement or experience of the tag, calculated for each tag's detection data and used as the basis for distinguishing between detections representing live study subjects and those occurring after predation.

Pattern recognition filter: Approach for diagnosing tag predation that involves application of one or more statistical or machine learning procedures. Although automated, these approaches still require subjectivity in the selection of metrics, transformations, and tuning parameters.

Predator filter: a formal, systematic, and repeatable data analysis procedure used to identify invalid portions of detection histories due to tag predation.

Predation tag: An electronic tag that transmits a uniquely identifiable acoustic signal that switches to an alternative signal upon predation of the tag, such as when a coating is dissolved within the gut of a predator or when loss of equilibrium is detected by an accelerometer.

Rule-based filter: Approach for diagnosing tag predation that involves the application of a predefined set of rules based on past research, expert judgement, or statistical features of the observed metric data.

Signal-based predator filter: An approach for identifying tag predation that relies entirely on data sent from the tag, such as temperature, depth, or predation signal triggered by the predation or digestive process (see "predation tag"), or else relies on interpreting temporal patterns in signal strength.

Smolt-tag data: Data from telemetry tags that were implanted in juvenile salmonids ("smolts"), including tags that were eventually classified as predated and data from the predated state.

Tag-level: Classification of a tag as representing the original study subject or a predator based on a summary of the full detection history; indicates whether the tag is or has been in a predator by the end of its detection history but not when the predation event occurred. Also used in reference to summaries of event-level variables.

Tag predation: the transferal of an active tag from the original study subject to a predator during a predation event.

Tidal-fluvial transition zone (TFTZ): Segment within a flowline map that can switch between bidirectional to unidirectional flow states depending on environmental conditions or management decisions (e.g., discharge thresholds and barrier placement).

Transition: Movement from one general location to another.

Unidirectional flow zone: Region within a flowline map where water moves in only one direction (e.g., downstream in riverine segments) rather than in multiple directions (e.g., reverse flows in tidal segments).

Visit: Detection event at a given general location.

.