

Advancing provenance assignment using machine learning and time series analysis of chemical chronologies in archival tissues

Kohma Arai^{a,*}, Malte Willmes^b, Rachel C. Johnson^{a,c}, Anna M. Sturrock^{a,d}

^a Center for Watershed Sciences, University of California Davis, One Shields Ave., Davis, CA 95616, USA

^b Norwegian Institute for Nature Research, PO Box 5685, Torgarden, 7485 Trondheim, Norway

^c National Marine Fisheries Service, Fisheries Ecology Division, Southwest Fisheries Science Center, 110 McAllister Way, Santa Cruz, CA 95060, USA

^d School of Life Sciences, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK

ARTICLE INFO

Keywords:

Otolith chemistry
Strontium isotope ratios (⁸⁷Sr/⁸⁶Sr)
Chinook salmon
Random forest
Dynamic time warping (DTW)
Ensemble methods

ABSTRACT

Accurate provenance assignment is critical for understanding ecological connectivity and guiding conservation and management. Natural chemical chronologies stored in metabolically inert, incrementally growing tissues (e. g., otoliths) provide a powerful tool for this purpose. However, traditional approaches face biological challenges, collapse chronological data into single values, and require subjective interpretation, limiting accuracy and scalability. We present a novel framework that integrates machine learning, time series analysis, and ensemble modeling to improve provenance assignment from archival tissue chemistry. Using otolith ⁸⁷Sr/⁸⁶Sr profiles from 17 distinct natal sources ($n = 255$) of California Central Valley Chinook salmon, we developed automated feature extraction, explicit time series classification (including dynamic time warping [DTW] with k -nearest neighbors [KNN]), and ensembles combining multiple classifiers. Furthermore, to address gaps in under-sampled life histories within natal sources, we added simulated chemical profiles to the reference baseline and tested whether they improved model performance. Time series approaches (mean accuracy: 0.56–0.62) consistently outperformed traditional methods (0.48), particularly for sources influenced by maternal signatures or early dispersal. Feature extraction approaches performed best when profiles followed predictable life-stage patterns, while explicit time series classification (DTW + KNN) excelled for distinct profile “shapes”. Ensembles leveraged complementary strengths and outperformed any single method. Our results highlight the advantages of treating archival chemical data as time series and applying machine learning and ensemble strategies to enhance the accuracy and scalability of provenance assignment. This framework is broadly applicable across taxa, tissue types, and chemical markers, offering a roadmap for advancing ecological inference and conservation.

1. Introduction

Tracing the origins and movements of organisms is key to understanding ecological processes and informing conservation and management. Provenance tools are broadly used in contexts ranging from migration and connectivity studies to wildlife forensics, food traceability, and species management (Bataille et al., 2020; Carter and Chesson, 2017; Hobson, 1999; Reich et al., 2024). Chemical records in metabolically inert, incrementally growing tissues such as bird feathers, mammal teeth, and fish otoliths have been widely used for this purpose (Newsome et al., 2010; Reis-Santos et al., 2023; van Dijk et al., 2014). In aquatic organisms, the chemistry of these tissues reflects the physicochemical properties of ambient water conditions (Hüsey et al., 2020),

enabling reconstruction of provenance and movement histories across taxa and ecosystems (Doubleday et al., 2022). However, robust and flexible analytical frameworks are still needed to translate these chemical signals into reliable provenance estimates.

Two major biological challenges limit the effectiveness of natural chemical tags for provenance assignment, both tied to the rate at which environmental chemistry is incorporated into tissues (i.e., turnover rate; Lewis et al., 2025; Yang et al., 2023). First, mobile species may disperse before the relevant structure has recorded the natal chemical signal (Fig. 1). This is especially problematic in marine organisms, where eggs and larvae are passively transported long distances (Cowen and Spoungle, 2009). Many geolocators in marine waters (e.g., carbonate $\delta^{18}\text{O}$) require considerable material, often integrating signals over long

* Corresponding author.

E-mail address: kharai@ucdavis.edu (K. Arai).

<https://doi.org/10.1016/j.ecoinf.2026.103604>

Received 8 September 2025; Received in revised form 4 December 2025; Accepted 6 January 2026

Available online 7 January 2026

1574-9541/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

timescales (Hanson et al., 2010). As a result, many studies focus on broader spatial units such as post-settlement juvenile nursery areas (e.g., Arai et al., 2021). Within-population variation in migration behavior (i.e., partial migration; Chapman et al., 2011) presents an additional challenge. Even when individuals originate from the same natal region, differences in migration strategies (e.g., early vs. late migrant; Fig. 1) can produce distinct chemical chronologies, complicating provenance assignment.

Second, maternal effects can confound natal chemical signals. For egg-laying species, the yolk can nourish offspring for extended periods, resulting in “maternal overprinting” of early chemical signatures (Hegg et al., 2019; Fig. 1). This is particularly problematic for species with larger eggs and where maternal environments differ from those of the offspring at the start of life. For example, anadromous salmonids undergo vitellogenesis in the ocean but spawn in freshwater rivers, creating a salinity-based mismatch between the chemical signature of the yolk versus the offspring's natal environment (Chittaro et al., 2023).

This mismatch, combined with early dispersal, can result in chemical chronologies that never reflect the offspring's natal signature.

In addition to these biological constraints, methodological challenges limit the scalability of chemical records for provenance assignment. For instance, visual stress checks in otoliths can provide valuable markers of key life history transitions, such as emergence or settlement, and are commonly used to guide targeted chemical analyses (Barnett-Johnson et al., 2007). However, these checks do not always align with natal signature formation. For example, a larva displaced by high flows may not deplete its yolk and form a stress mark until well after it has left its natal area, potentially decoupling chemical signals from geographic origin. Moreover, identifying regions of interest within chemical time series typically relies on manual inspection (Brennan et al., 2015), introducing subjectivity, limiting throughput, and constraining use in large-scale applications.

A further limitation in the analysis of chemical time series is the tendency to oversimplify their inherently chronological structure.

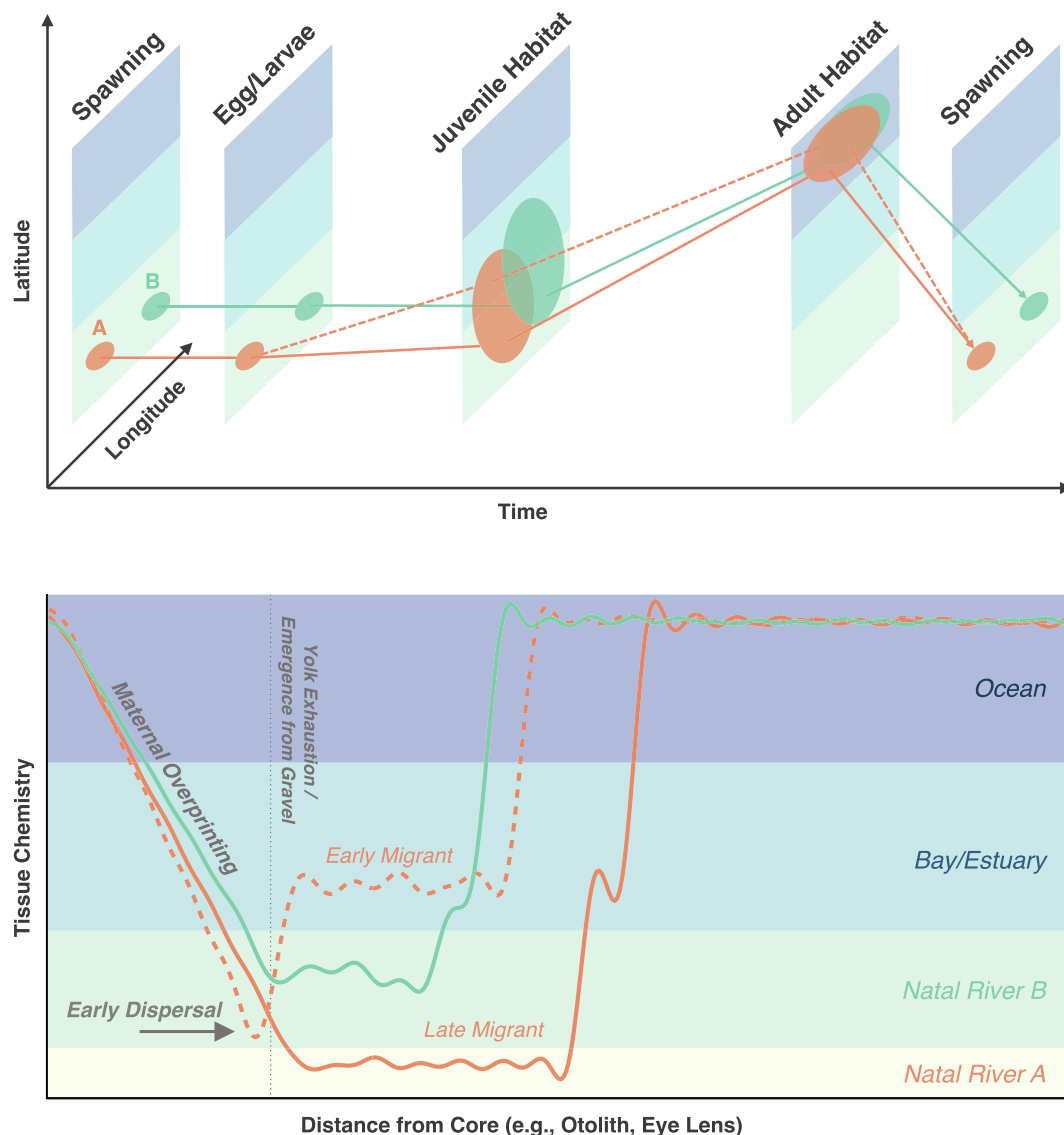


Fig. 1. Schematic showing geographic distributions of life history stages and corresponding chemical chronologies in archival tissues (e.g., otolith, eye lens) of anadromous salmonids. Life histories for two distinct natal sources (A and B) are depicted. For source A, fish exhibit early (dashed) or late (solid line) downstream migration. (i) Early migrants may leave the natal river before the tissue records its chemical signature. (ii) The initial portion of the time series reflects maternal chemical signature (“maternal overprinting”) until yolk reserves are exhausted (dotted line), after which ambient signals emerge. Ellipses represent potential geographic distributions; background colors indicate different habitats. Note that this example was inspired by otolith $^{87}\text{Sr}/^{86}\text{Sr}$ (uniform at sea), but for markers exhibiting spatial variation among adult feeding grounds (e.g., eye lens $\delta^{13}\text{C}$; Quaeck, 2017), their offspring's archival structures could reflect such differences in the core. Figure inspired by Avigliano (2022) and Secor (1999).

Rather than leveraging the full sequence of measurements, many studies reduce these data to summary statistics (e.g., mean values) within a fixed window, leading to a loss of temporal resolution and biological nuance (Barnett-Johnson et al., 2008). Yet time series contain information not only in the individual values, but also in their sequence and overall structure, all of which are obscured when condensed into static summaries (Chatfield and Xing, 2019).

Time series methods are increasingly used to identify migratory phenotypes (e.g., Cordoleani et al., 2021; Vignon, 2015), and to characterize larval trajectories and partial migration using unsupervised clustering such as Dynamic Time Warping (DTW; Arai et al., 2024; Hegg and Kennedy, 2021; Teichert et al., 2024). Machine learning classifiers such as random forests (RF) and *k*-nearest neighbors (KNN) outperform traditional methods (e.g., linear discriminant analysis) in capturing non-linear relationships and temporal patterns (Arai et al., 2023; Mercier et al., 2011). Ensemble approaches, which integrate multiple classifiers, are also gaining popularity for their ability to improve predictive accuracy by leveraging the strength of diverse models (Dietterich, 2000).

Provenance assignment generally follows one of two frameworks: continuous or nominal assignment (Wunder, 2012). Continuous methods generate spatial probability surfaces to estimate an individual's most likely origin across a continuous landscape (e.g., Brennan et al., 2019b; Reich et al., 2024; Vander Zanden et al., 2018). Nominal approaches classify individuals into a finite set of predefined sources based on chemical similarity to reference data (e.g., Arai et al., 2021; Barnett-Johnson et al., 2008; Rooker et al., 2014). In this study, we focus on nominal assignment and explore methods that incorporate more of the underlying temporal structure in chemical time series to improve provenance accuracy.

A key assumption in supervised nominal assignment is that all contributing sources to the mixed group are well represented in the reference data (Elsdon et al., 2008). However, this assumption can break down when rapid dispersal during early life stages or geographically widespread populations makes it logistically challenging to collect sufficient known-origin samples. For these poorly characterized life history traits or natal sources, the lack of real, direct observations can introduce substantial bias into provenance models. In such cases, simulation techniques based on established biological principles, such as chemical uptake dynamics, known water chemistries, and fish movement patterns, can provide a realistic proxy for actual measurements, effectively serving as a reference dataset for model training (Artetxe-Arrate et al., 2025). Together, these approaches offer a promising framework for overcoming the biological and technical challenges inherent in provenance assignment from chemical time series in archival tissues.

Otolith strontium isotope ratios ($^{87}\text{Sr}/^{86}\text{Sr}$) are particularly effective for tracking fish movements among rivers with distinct geologies, as these isotopic ratios vary predictably with underlying bedrock and incorporated reliably into otoliths (Barnett-Johnson et al., 2008; Bataille and Bowen, 2012). California's Central Valley, with its geologically diverse watersheds, is ideal for such provenance applications (Barnett-Johnson et al., 2008; Johnson et al., 2016). California Central Valley fall-run Chinook salmon (*Oncorhynchus tshawytscha*) display diverse life history traits, both spatially (across distinct spawning populations) and temporally (through varied juvenile emigration timings within populations; Williams, 2006). Adults migrate upstream into freshwater during late summer to early fall, spawning primarily from October to December. Juveniles typically emerge and begin their downstream migration between winter and spring, rearing in the Delta before entering the ocean (Williams, 2006; Yoshiyama et al., 1998). Otolith $^{87}\text{Sr}/^{86}\text{Sr}$ has been widely used to assign natal origin and estimate juvenile outmigration timing (Sturrock et al., 2020; Willmes et al., 2024). However, juveniles that leave natal rivers shortly after emergence (early migrants) may retain maternal signals and lack sufficient time to incorporate distinct natal river signatures into their otoliths, making them particularly susceptible to misclassification using traditional methods (Sturrock et al., 2020; Willmes et al., 2024; see also Fig. 1).

These early migrants are typically too small to tag at the time of dispersal, and if later sampled downstream, their natal origin cannot be confirmed, leading to gaps in known-origin reference data for this rapidly dispersing phenotype within a given population. Additionally, conventional assignment approaches rely on manual inspection of individual $^{87}\text{Sr}/^{86}\text{Sr}$ profiles and data summarization, which can reduce both throughput and assignment accuracy.

Here, we use otolith $^{87}\text{Sr}/^{86}\text{Sr}$ time series data from Central Valley fall-run Chinook salmon to develop and compare nominal assignment approaches for classifying natal origin across 17 distinct sources within the Sacramento-San Joaquin River system (Fig. 2). We evaluate (1) manual versus automated natal region selection, (2) summary statistics versus full time series representations, and (3) a range of classification techniques. To advance beyond conventional approaches, we introduce novel methodologies that integrate time series feature extraction, machine learning classification, and ensemble modeling. We also incorporate simulated training data to represent under-sampled early migrant phenotypes and test their contribution to model performance. Finally, we validate our framework using both known-origin and unknown-origin field samples to assess its accuracy and relevance for management decisions. Our goal is to evaluate trade-offs in accuracy, scalability, and efficiency, and to provide a robust, flexible workflow for provenance assignment using archival tissue chemical chronologies.

2. Materials and methods

2.1. Datasets

2.1.1. Otolith $^{87}\text{Sr}/^{86}\text{Sr}$ reference data

The classification model included 17 natal sources (12 natural, 5 hatchery) that represent the majority of the Central Valley Chinook salmon stock complex. Model training primarily used otolith $^{87}\text{Sr}/^{86}\text{Sr}$ profiles measured from the core through the early juvenile growth region of fall-run juveniles sampled in natal rivers or hatcheries prior to outmigration (Table S1, Fig. 2). For hatcheries, we also included older fish with confirmed origins via coded wire tags (CWT). Hatchery fish were included in the reference baseline because their otolith $^{87}\text{Sr}/^{86}\text{Sr}$ can differ from those of natural fish from the same river, due to marine-derived feed in hatcheries (Barnett-Johnson et al., 2008).

Due to limited known-origin fall-run juvenile samples for the upper Sacramento River (SAC), Butte (BUT), Mill (MIL), and Deer (DEE) Creeks, we included otolith $^{87}\text{Sr}/^{86}\text{Sr}$ data from post-spawned adult carcasses recovered during the spawning season—winter-run in the SAC and spring-run in BUT, MIL, and DEE. While not fall-run, these fish likely represent the natal sources of their recovery sites based on high homing rates (Sturrock et al., 2019) and close agreement between their natal otolith signatures and local water chemistry.

Otolith samples from each source were randomly selected across years, resulting in a broad temporal range (Table S1). Although some high-latitude or glaciated watersheds exhibit notable seasonal variation in $^{87}\text{Sr}/^{86}\text{Sr}$ due to shifts in water-source contributions (e.g., Douglas et al., 2013; Voss et al., 2014) previous work in our temperate study system has documented minimal interannual variation in river $^{87}\text{Sr}/^{86}\text{Sr}$ (Barnett-Johnson et al., 2008). Accordingly, our multi-year sampling design is unlikely to introduce temporal bias into the $^{87}\text{Sr}/^{86}\text{Sr}$ baselines used for assignment.

2.1.2. Synthetic early migrant data

To address key gaps in the reference dataset, we simulated otolith $^{87}\text{Sr}/^{86}\text{Sr}$ profiles for “early migrants”—juveniles that emigrate shortly after emergence as fry, and rear in non-natal habitats. These individuals are underrepresented in empirical data due to their small size and tagging limitations but often dominate outmigration cohorts, making their exclusion a potential source of bias (Sturrock et al., 2020). Simulated profiles were informed by water $^{87}\text{Sr}/^{86}\text{Sr}$ data, otolith growth rates, and migration timing, and incorporated maternal, natal, and

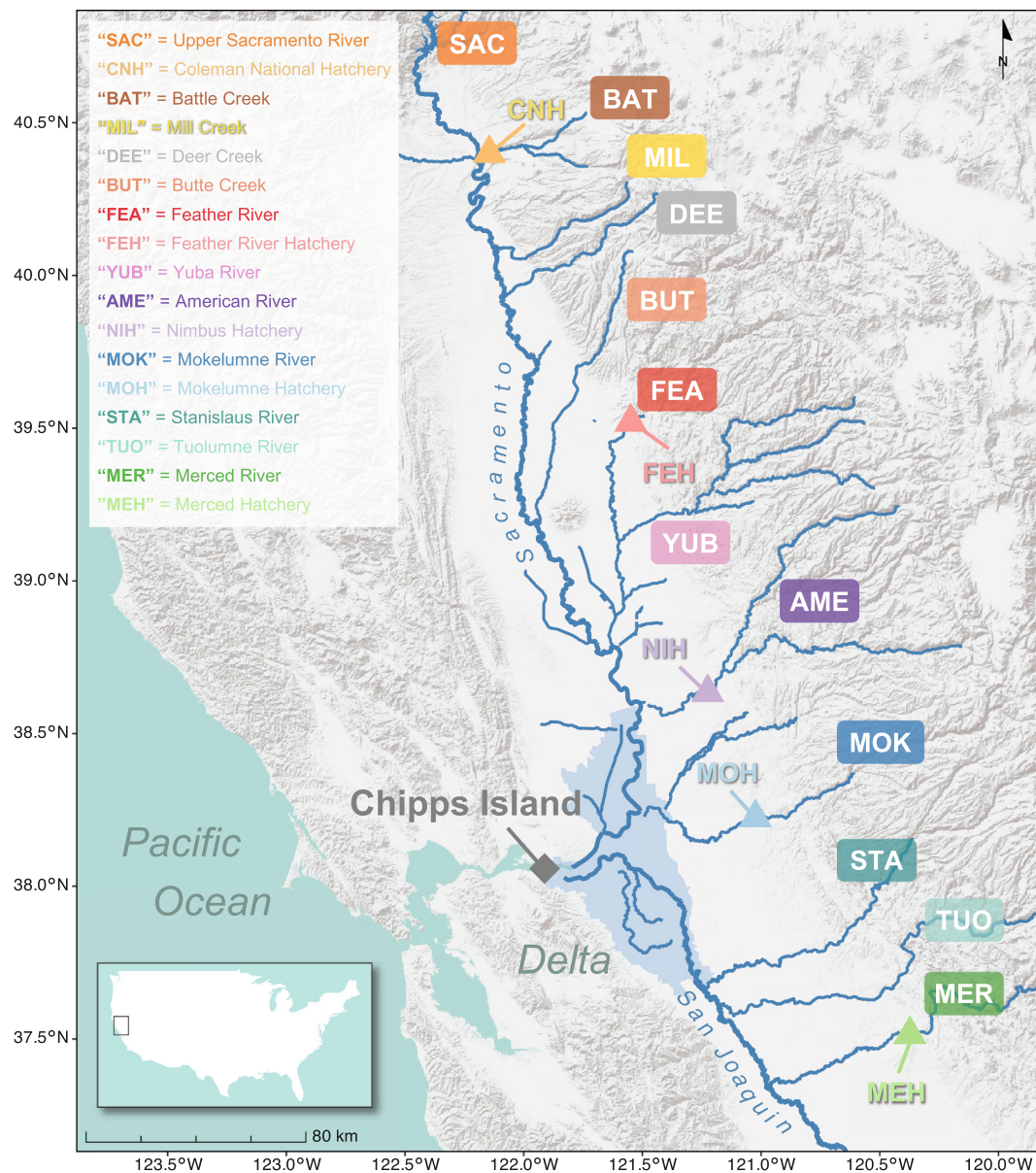


Fig. 2. Map of California's Central Valley showing the locations of Chinook salmon-producing rivers and hatcheries (triangles), representing a total of 17 natal sources, along with their corresponding three-letter source codes used throughout the paper. Juvenile Chinook salmon of unknown origin were collected at Chipps Island (diamond), located at the downstream terminus of the Delta.

downstream phases. Parameter values were grounded in empirical data; for instance, otolith size at emergence was parameterized as $225 \pm 30 \mu\text{m}$ (mean $\pm$ SD), representing the midpoint between two independent otolith datasets (mean $\pm$ SD = $217 \pm 25 \mu\text{m}$, $n = 3113$, and $236 \pm 28 \mu\text{m}$, $n = 960$; Sturrock et al., unpublished). Early migrants were modeled to emigrate soon after emergence, typically reaching otolith sizes of $250 \pm 50 \mu\text{m}$ before moving into downstream freshwater habitats (Sturrock et al., 2020; Williams, 2006). Downstream rearing otolith growth of $250 \pm 40 \mu\text{m}$ was informed by the mean otolith radii at natal and freshwater exit (Cordoleani et al., 2021; Willmes et al., 2024). We generated five synthetic otolith $^{87}\text{Sr}/^{86}\text{Sr}$ profiles per natural-origin source (excluding hatcheries, which do not release fish at fry size; Huber and Carlson, 2015), resulting in a total of 255 samples across 17 natal sources (15 per source; Table S1) used to develop the assignment models. Full simulation procedures, parameter ranges, and biological rationales are provided in Supplementary Text S1 and Table S2.

2.1.3. Otolith microstructure features

In addition to otolith $^{87}\text{Sr}/^{86}\text{Sr}$ profiles, we incorporated otolith microstructure features (mean exogenous feeding check score) into the natal assignment model to improve classification accuracy (see Supplementary Text S2 for details). The mean exogenous feeding check score was included as an additional predictor alongside otolith $^{87}\text{Sr}/^{86}\text{Sr}$ to inform natal assignment.

2.2. Approaches for provenance assignment

We developed and compared seven models for assigning salmon natal origin based on otolith $^{87}\text{Sr}/^{86}\text{Sr}$ profile (Fig. 3). These span a continuum from expert-driven, discrete-feature approaches to fully automated, time series-based classification methods. This continuum reflects a progression in how much temporal structure is retained and how much user interpretation is required. The most complex model integrates predictions from multiple approaches into an ensemble, leveraging the strengths of each method.

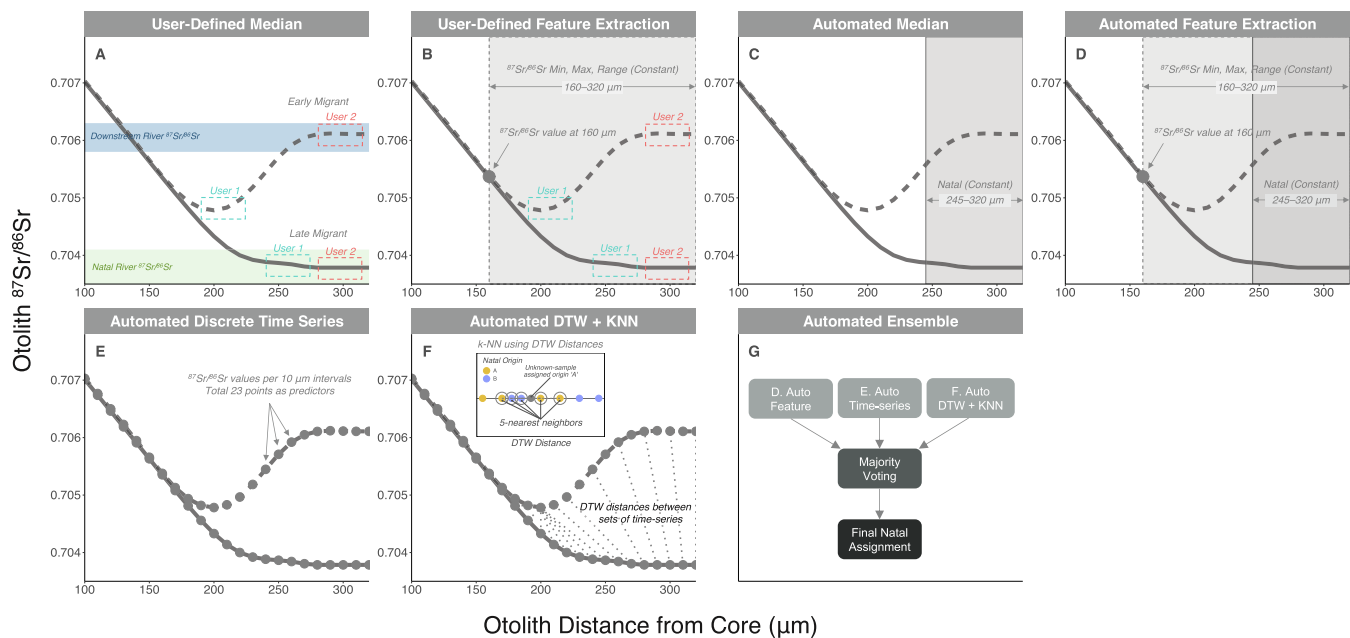


Fig. 3. Overview of the seven approaches used to assign fish provenance from otolith $^{87}\text{Sr}/^{86}\text{Sr}$ profiles. Panels (A–B) show user-defined models (A: median-based; B: feature extraction). Panels (C–D) show automated discrete models (C: median-based; D: feature extraction). Panels (E–F) show time series models (E: discrete time series; F: DTW + KNN). Panel (G) illustrates the ensemble model that integrates outputs across multiple approaches. Solid and dashed lines show example profiles for late and early migrants, respectively, measured from 100 and 320 μm from the core. In panel (A), green and blue shading indicate natal river and downstream river water $^{87}\text{Sr}/^{86}\text{Sr}$ values, respectively. In panels (A) and (B), blue and red dashed boxes indicate user-defined natal regions identified by two independent readers. While readers assign similar natal regions for late migrants, they diverge on early migrants, with User 2 selecting a later portion of the profile than User 1. This highlights the subjectivity and inconsistency inherent in conventional manual inspection approaches. See Materials and Methods for detailed descriptions of each approach. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

To standardize inputs, we fit spline models to interpolate $^{87}\text{Sr}/^{86}\text{Sr}$ values every 10 μm between 100 and 320 μm from the otolith core. Spline degrees of freedom were adjusted to avoid oversmoothing. We excluded the first 100 μm due to potential overpolishing, which can expose material deposited later in life. Data beyond 320 μm were excluded to ensure that only the natal portion of the otolith was analyzed, as this otolith radius corresponds to when juveniles begin emigrating from their natal river and transition toward non-natal habitats (Cordoleani et al., 2021; Willmes et al., 2024).

2.2.1. User-defined models

2.2.1.1. User-defined median.

This traditional approach relies on expert visual inspection to identify the natal region—typically the plateau following yolk depletion, often guided by the exogenous feeding check (Fig. 3A). The median $^{87}\text{Sr}/^{86}\text{Sr}$ value within this region, plus the mean exogenous feeding check score, were used as predictors in a random forest (RF) classifier. While flexible, this approach is time-intensive and subjective, as each profile must be interpreted manually.

2.2.1.2. User-defined feature extraction.

To account for early downstream movements of fish—before their otoliths fully reflect the natal $^{87}\text{Sr}/^{86}\text{Sr}$ signature—this model builds upon the user-defined median approach by incorporating additional time series features derived from prior biological knowledge of Chinook salmon in this system (Fig. 3B). Although this feature extraction approach is not a time series method in the formal sense, it attempts to extract biologically meaningful time series features informed by expert knowledge. Specifically, we calculated the minimum, maximum, and range of $^{87}\text{Sr}/^{86}\text{Sr}$ values between 160 and 320 μm , based on the premise that these features could capture movements across habitats with distinct $^{87}\text{Sr}/^{86}\text{Sr}$ signatures unique to each source. We also included the $^{87}\text{Sr}/^{86}\text{Sr}$ value measured at 160 μm from the core, as this point often precedes the onset of early dispersal,

and thus all individuals from a given source tend to share similar isotopic values at this size. These summary statistics were used alongside the user-defined median $^{87}\text{Sr}/^{86}\text{Sr}$ and the mean exogenous feeding check score in RF classification.

2.2.2. Automated discrete models

To reduce subjectivity and increase efficiency, we developed automated variants of the above two models using fixed rules to define the natal region and extract summary statistics.

2.2.2.1. Automated median.

This model defines a fixed natal region (245–320 μm) for all samples assumed to represent otolith growth post-yolk absorption but prior to outmigration (Fig. 3C; Barnett-Johnson et al., 2008). The median $^{87}\text{Sr}/^{86}\text{Sr}$ value within this region and the mean exogenous feeding check score were used as RF inputs.

2.2.2.2. Automated feature extraction.

This model mirrors the automated median approach but includes additional features to capture variability in the $^{87}\text{Sr}/^{86}\text{Sr}$ profile (Fig. 3D). In addition to the median $^{87}\text{Sr}/^{86}\text{Sr}$ value and mean exogenous feeding check score, we calculated the minimum, maximum, and range of $^{87}\text{Sr}/^{86}\text{Sr}$ values between 160 and 320 μm , as well as the $^{87}\text{Sr}/^{86}\text{Sr}$ value measured at 160 μm from the core and used as RF inputs.

2.2.3. Time series models

Moving beyond summary statistics, the next set of models retain the temporal structure of the otolith $^{87}\text{Sr}/^{86}\text{Sr}$ profiles.

2.2.3.1. Discrete time series.

This model treats the otolith $^{87}\text{Sr}/^{86}\text{Sr}$ profile as a set of interpolated values sampled at discrete intervals (Fig. 3E). Specifically, 23 $^{87}\text{Sr}/^{86}\text{Sr}$ values spaced every 10 μm from 100 to 320 μm were used as individual predictors in RF, alongside the mean exogenous feeding check score. Although this automated approach does

not require a user-defined natal region and leverages regularly spaced values from a time series, it does not explicitly model the sequential nature of the data.

2.2.3.2. Dynamic time warping with k -nearest neighbor classification (DTW + KNN). To explicitly treat the otolith $^{87}\text{Sr}/^{86}\text{Sr}$ profile as a continuous time series and to classify samples based on the overall structure and sequence of the profile, we developed a model that combines dynamic time warping (DTW) with k -nearest neighbor (KNN) classification (Wang et al., 2013; Fig. 3F). DTW aligns sequences by minimizing time-warping cost and enables comparison of profiles with similar patterns but varying lengths (Aghabozorgi et al., 2015; Hegg and Kennedy, 2021). While DTW has been applied to otolith chemistry data in an unsupervised context to identify distinct migration patterns (Arai et al., 2024; Hegg and Kennedy, 2021; Takai et al., 2024; Teichert et al., 2024), our approach represents a novel application of DTW for supervised classification by pairing it with KNN. This allows us to directly classify unknown samples based on their temporal structure, leveraging the full profile information rather than relying on summary statistics or discrete points. While deep learning approaches have increasingly been used for large-scale time series classification (Fawaz et al., 2019), DTW + KNN provides an interpretable, robust approach for comparing temporal profiles when sample sizes are moderate such as ours ($n = 255$; Bagnall et al., 2017).

We computed the DTW distance between the $^{87}\text{Sr}/^{86}\text{Sr}$ profile of each unknown-origin sample and all known-origin samples in the training set using the *dtw* R package (Giorgino, 2009). DTW was implemented with an *asymmetric* step pattern and a Sakoe-Chiba window (*window.size* = 5 L), with both *open.begin* and *open.end* set to "TRUE". Classification was then performed using KNN ($k = 5$; see below), where the sample was assigned to the most common class among its five nearest neighbors based on DTW distance. Note that this method did not incorporate information about the mean exogenous feeding check score, and assignments were based solely on DTW distance.

2.2.4. Ensemble model

Finally, to capitalize on the complementary strengths of different modeling strategies, we developed an ensemble classifier that integrates predictions from three automated approaches: the automated feature extraction, discrete time series, and DTW + KNN models (Fig. 3G). Each method independently generated a class prediction, and the final assigned natal origin was determined by majority vote among the three predictions.

2.2.5. Model evaluation

We included a "null" model that assigns all samples to the most frequent class to benchmark the classification performance of all other classifiers. Natal assignments were performed in R (R Core Team, 2025) using *tidymodels* (Kuhn and Wickham, 2020), *ranger* (Wright and Ziegler, 2017), and *manymodelsR* (Gonzabato, 2025). Hyperparameters for RF (*mtry* and *min_n*) and KNN (k) were tuned using grid search with 10-fold cross-validation. The full set of optimal hyperparameter combinations is provided in Table S3.

Model performance was evaluated using 10-fold cross-validation repeated 10 times. Because the dataset consisted of balanced multi-class data ($n = 15$ samples per source), we evaluated model performance using multiple complementary metrics: overall accuracy, macro-averaged F1 score, macro precision, and macro recall (Akosa, 2017; Diallo et al., 2025). To further examine classification errors and identify natal sources most frequently mis-assigned, we also generated confusion matrices for the top-performing models (Dunnigan et al., 2023; Sigurðardóttir et al., 2023).

To provide additional context for model performance, we also compared classification accuracy between the RF and traditional

statistical classifiers such as linear and quadratic discriminant analysis, using the automated median approach as a common baseline (Fig. 3C). Detailed methods and results are presented in Supplementary Text S3.

Lastly, to evaluate subjectivity of natal region selection in user-defined approaches and how these factors might influence subsequent natal assignment results, we conducted a visual identification trial in which three trained readers independently selected natal regions based on otolith $^{87}\text{Sr}/^{86}\text{Sr}$ profiles. Detailed methods and results are presented in Supplementary Text S4.

2.3. Natal assignment of unknown-origin juveniles

To evaluate the performance of the natal assignment model in a more realistic context, we assigned natal origins to wild, randomly selected unknown-origin juveniles ($n = 50$) collected at Chippis Island from 2014 to 2021, the downstream terminus of the Delta through which all natal river sources must pass to reach the ocean (Fig. 2; Table S1). These fish were collected under scientific permit SC-12811.

2.4. Natal assignment of known-origin adult returns

To assess the potential benefit of including simulated synthetic early migrants in the training data, we tested model performance using an independent set of otolith $^{87}\text{Sr}/^{86}\text{Sr}$ profile data from post-spawned adults sampled from the Yuba (YUB), American (AME), and Stanislaus (STA) Rivers ($n = 30$; Fig. 2; Table S1). These adults were assumed to be known-origin, early migrants based on high homing rates and chemical transitions between natal rivers and subsequent habitats along their migration path. We specifically compared the classification accuracy of models trained with and without synthetic early migrants to assess the impact of this additional training data.

3. Results

3.1. Otolith $^{87}\text{Sr}/^{86}\text{Sr}$ of Central Valley Chinook salmon

The California Central Valley's geologically diverse watershed produced unique otolith $^{87}\text{Sr}/^{86}\text{Sr}$ profiles for Chinook salmon from 17 natal sources (Fig. 4, Fig. S1). Both natural-origin and hatchery fish showed $^{87}\text{Sr}/^{86}\text{Sr}$ values close to maternal marine values near the otolith core, gradually transitioning to natal river signatures as yolk reserves were exhausted. The steepness of this transition reflected the degree of difference between ocean and natal water chemistry: sources with highly distinct natal signatures (e.g., BAT, $^{87}\text{Sr}/^{86}\text{Sr} \sim 0.704$) showed steep transitions, while those closer to marine values (e.g., MEH, ~ 0.7085) exhibited more gradual shifts.

Differences in otolith $^{87}\text{Sr}/^{86}\text{Sr}$ profiles between early and late migrants also varied by source. In rivers with natal signatures distinct from the downstream Delta (~ 0.706), early migrants exhibited pronounced shifts compared to late migrants (e.g., AME). In contrast, sources with values similar to the Delta showed minimal differences (e.g., FEA).

3.2. Classification performance among different models

Ten-fold cross-validation, repeated 10 times on known-origin juvenile samples, showed that the ensemble (mean $\pm$ SD: 0.628 ± 0.009), automated feature extraction (0.622 ± 0.011), and user-defined feature extraction (0.619 ± 0.016) models had the highest overall classification accuracy (Fig. 4). These time series approaches substantially outperformed the traditional user-defined median method based on summary statistics (0.476 ± 0.011). All models performed far above the "null" model (0.039), indicating they captured meaningful structure in the data beyond random chance. Model rankings across other evaluation metrics (macro F1, macro precision, and macro recall) were consistent with those based on accuracy (Fig. S2).

Accuracy varied by natal source, with each model showing unique

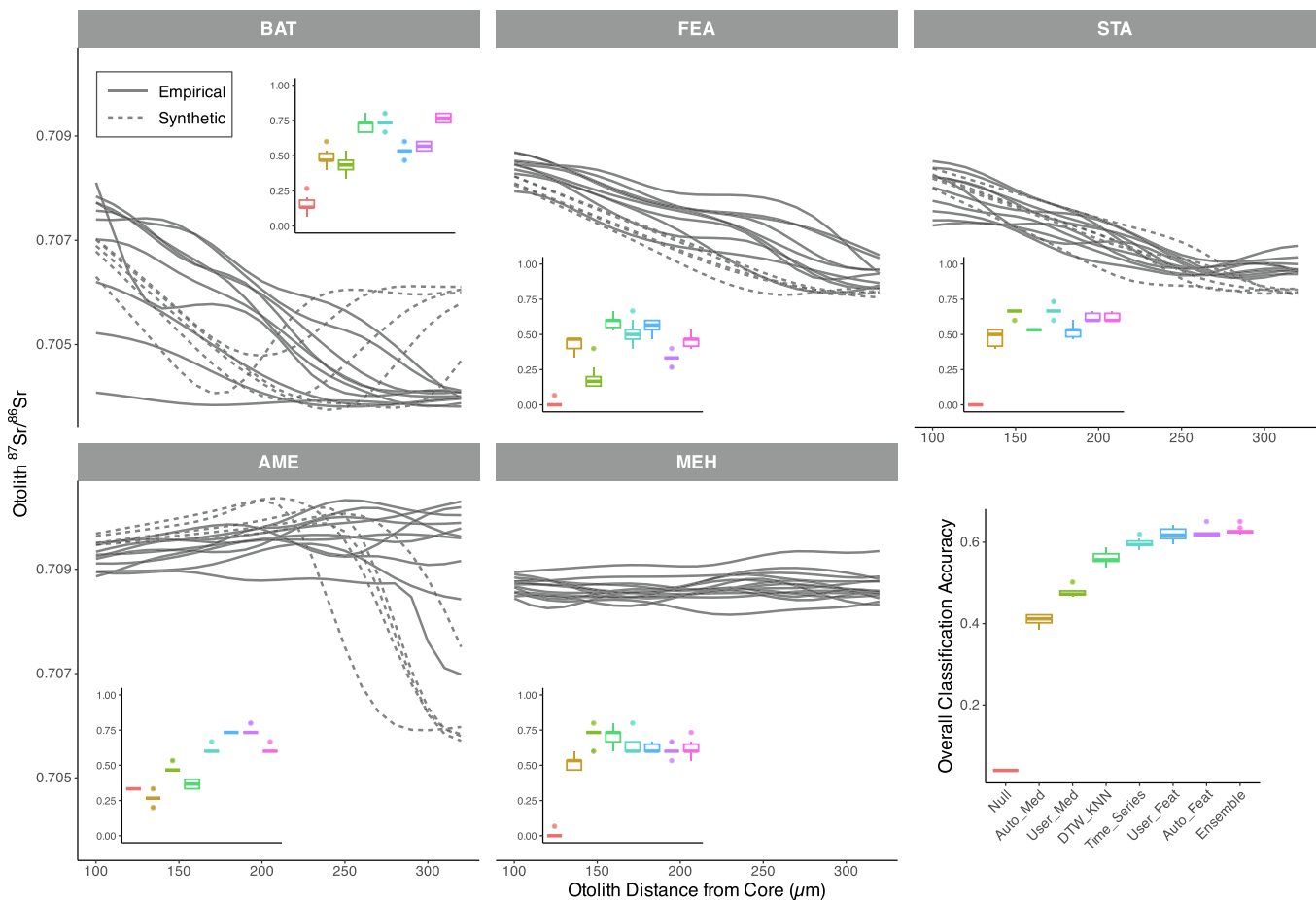


Fig. 4. Classification accuracy of natal assignment models evaluated using repeated 10-fold cross-validation ($n = 255$ known-origin Chinook salmon). Panels show otolith $^{87}\text{Sr}/^{86}\text{Sr}$ profiles by natal source, with classification accuracy indicated in insets. Solid and dashed lines represent empirical and synthetic fish, respectively. Overall classification accuracy across all 17 sources is shown in the bottom right panel. The strip text in each panel corresponds to the natal source code; these abbreviations are defined in the map in Fig. 2. This figure displays a subset of natal sources (5 out of 17) chosen to illustrate a range of distinct otolith $^{87}\text{Sr}/^{86}\text{Sr}$ patterns. Results for all 17 sources are provided in Fig. S1. Model abbreviations: “Null” = non-informative model; “Auto_Med” / “User_Med” = automated / user-defined median; “DTW_KNN” = dynamic time warping with k -nearest neighbors; “Time_Series” = discrete time series; “User_Feat” / “Auto_Feat” = user-defined / automated feature extraction; “Ensemble” = ensemble model.

strengths (Fig. 4, Fig. S1). The median-based approaches were generally outperformed, but performed moderately well for hatchery sources with distinct $^{87}\text{Sr}/^{86}\text{Sr}$ values between 245 and 320 μm from the otolith core (e.g., MEH), or where early and late migrants had similar $^{87}\text{Sr}/^{86}\text{Sr}$ profiles (e.g., STA). The automated feature extraction performed well across most sources, particularly where early migrants differed markedly from late migrants within the same source (e.g., AME). The discrete time series and DTW + KNN approaches excelled where profiles lacked distinct values at fixed points but had unique overall “profile shapes” (e.g., FEA, MEH). The ensemble model, which combined predictions from the automated feature extraction, discrete time series, and DTW + KNN approaches, consistently performed well across sources.

Confusion matrices revealed that misclassifications most frequently occurred among geographically adjacent sources (e.g., BAT, MIL, and DEE; Fig. S3), or among sources with similar natal $^{87}\text{Sr}/^{86}\text{Sr}$ signatures despite geographic separation (e.g., FEA and STA; SAC and BUT). Misassignments were also common between natural and hatchery fish from the same river system (e.g., AME and NIH; MER and MEH).

3.3. Natal assignment of unknown-origin juveniles

Applying different natal assignment models to unknown-origin juveniles collected at Chipps Island (the Delta terminus where all juvenile sources pass to reach the ocean; Fig. 2) from 2014 to 2021 revealed a

diverse array of sources contributing to the stock complex (Fig. 5; Fig. S4). Among-model assignment agreement varied from 0.34 to 0.92 (Fig. S5). The DTW + KNN, discrete time series, user-defined feature extraction, automated feature extraction, and ensemble approaches generally produced consistent assignments (agreement: 0.70–0.92). In contrast, the lowest agreement (0.34–0.56) occurred between median-based models (e.g., automated or user-defined median) and time series based approaches (e.g., discrete time series, DTW + KNN, automated feature extraction).

3.4. Natal assignment of known-origin adult returns

The user-defined feature extraction approach yielded the highest overall classification accuracy for known-origin adult returns that emigrated early (“early migrants”), performing well both with (accuracy = 0.73) and without (0.57) synthetic early migrants in the training data (Fig. 6; Fig. S6). Accuracy varied by natal sources, with AME showing the highest (0.69), and YUB (0.21) and STA (0.28) the lowest. Removing synthetic early migrants generally reduced accuracy across models, with the largest declines in the automated feature extraction (–42.1%) and discrete time series (–29.4%) models. The effect varied by natal source: accuracy for YUB and AME declined sharply—up to 80% for YUB in the discrete time series approach—while STA showed little to no change. Notably, the user-defined feature extraction model

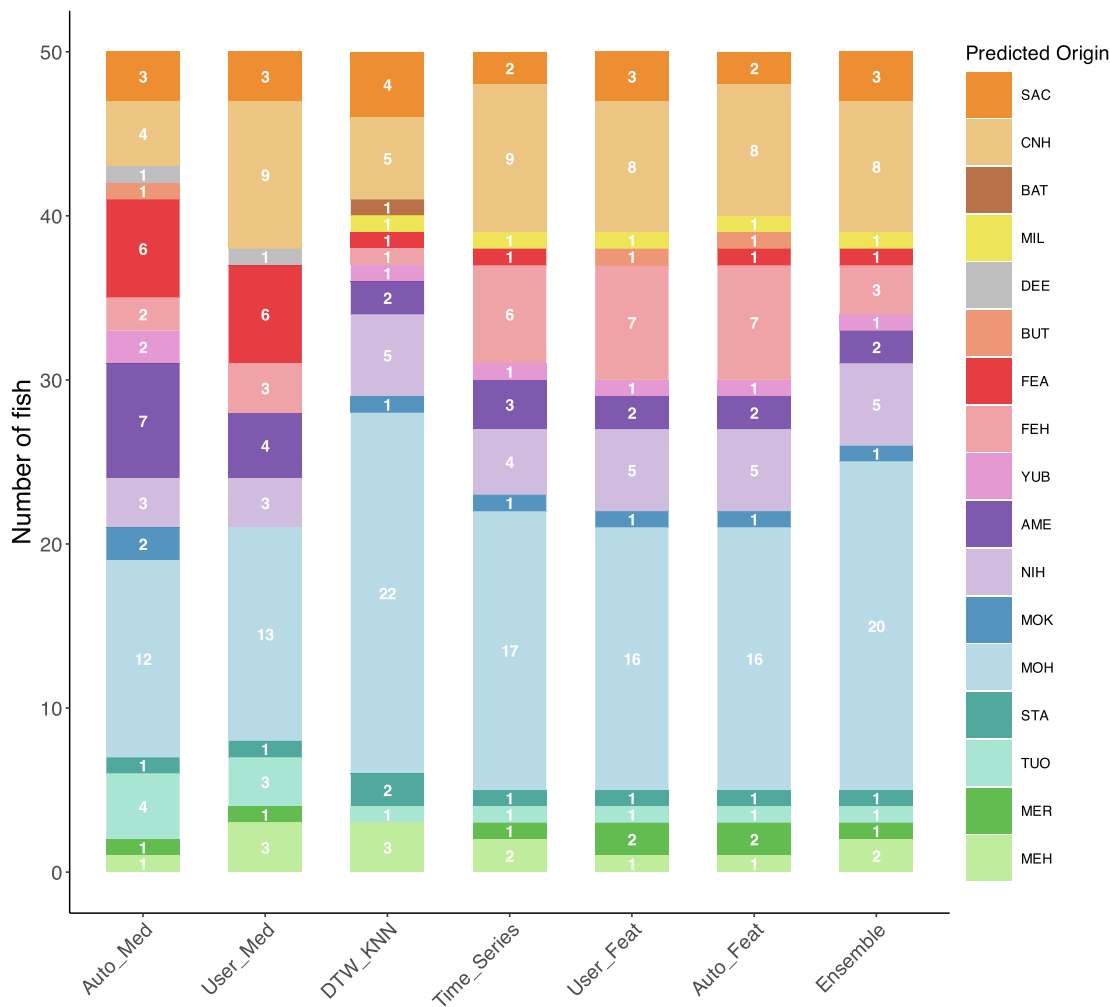


Fig. 5. Natal assignment results using different modeling approaches, applied to unknown-origin juvenile Chinook salmon ($n = 50$) collected at Chipps Island. The numbers in each bar represent the sample size assigned to each predicted natal source. These samples were randomly selected from fish collected between 2014 and 2021 and do not reflect the true population composition at Chipps Island in any specific year, as the primary purpose was to evaluate model performance with real-world samples. Natal sources (and abbreviations) are defined in Fig. 2 and are ordered from north to south according to their geographic location. Assignment model abbreviations are defined in Fig. 4.

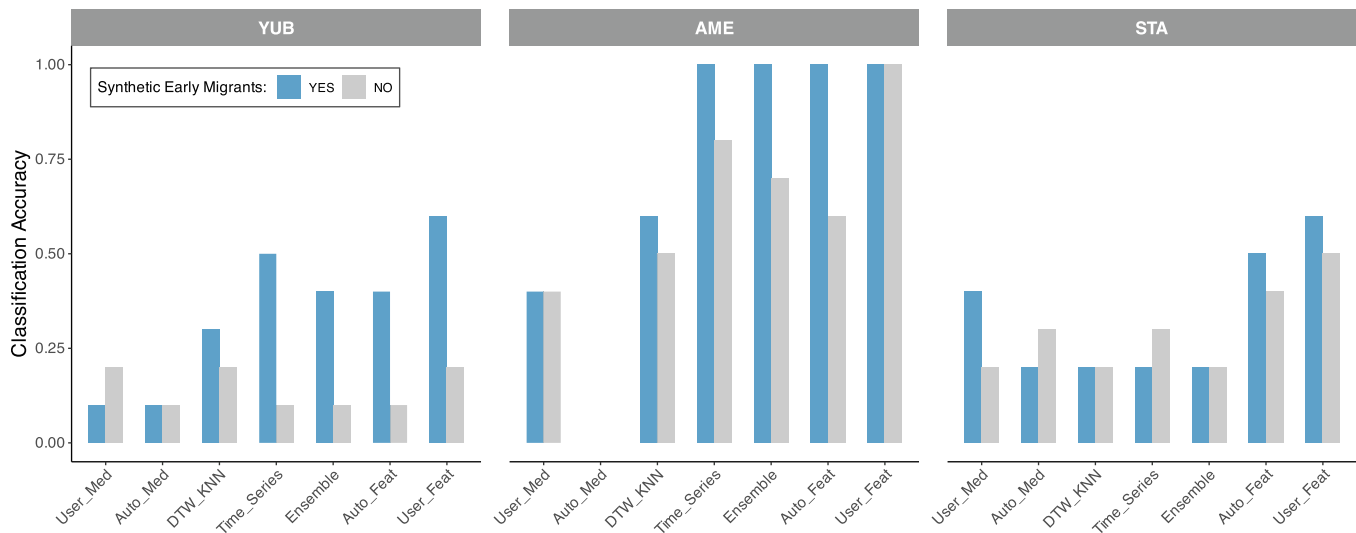


Fig. 6. Classification accuracy of known-origin adult returns that emigrated early as fry from the Yuba (YUB), American (AME), and Stanislaus (STA) Rivers, evaluated using different natal assignment models. Results compare models trained with and without synthetic early migrants in the training dataset. Model abbreviations are defined in Fig. 4.

remained robust for AME and STA, showing almost no loss in accuracy after removing synthetic early migrants.

4. Discussion

4.1. Advancing archival tissue provenance assignment with machine learning and time series approaches

By integrating time series methods with machine learning, our approach addresses key biological and technical challenges in provenance assignment. Unlike traditional methods that rely on static or single-point measurements, treating chemical data as a time series allowed us to capture temporal structure, improving classification accuracy.

Models that incorporated time series information, either through the discrete time series approach or by including time series-derived features (i.e., user-defined or automated feature extraction approaches), were especially effective where $^{87}\text{Sr}/^{86}\text{Sr}$ profiles exhibited steep transitions from marine to natal values, resulting in larger maternal overprinting (e.g., BAT, SAC; Fig. 4, Fig. S1). Even when profiles did not explicitly reflect natal signatures due to early dispersal, these approaches leveraged the shape and steepness of the initial transition to accurately assign origin, outperforming traditional methods based solely on summary statistics (median-based approaches). Likewise, for sources exhibiting large differences between the natal stream and downstream habitats, resulting in steep transitions in the later part of the profile for early migrants (e.g., BAT, AME), time series approaches better captured within-source variation than traditional median-based models. This demonstrates the advantage of time series approaches in accounting for early dispersal and maternal influences in natal origin assignment.

While differences in classification accuracy among models were modest, the feature extraction approaches consistently achieved high performance (Fig. 4). These models incorporate time series structure by extracting summary features (median, minimum, maximum, and range of $^{87}\text{Sr}/^{86}\text{Sr}$ values) rather than preserving the full temporal sequence. By reducing dimensionality, these feature-extraction methods filter out less informative segments of the signal, effectively enhancing the signal-to-noise ratio. Unlike explicit time series approaches (e.g., DTW + KNN), which generally requires temporal alignment of key chronological features (e.g., yolk depletion or downstream movement) across individuals to maximize assignment accuracy, the feature extraction approach accommodates greater individual variability in the timing of chemical changes. Additionally, the feature extraction approach implicitly incorporates expert knowledge by focusing on specific regions of the profile (e.g., 245–320 μm) where Chinook salmon natal signatures are most likely to be found (Barnett-Johnson et al., 2008).

Time series feature extraction may be especially suitable for species where environmental signals are overlaid onto consistent ontogenetic patterns, a phenomenon commonly observed in otolith chemical records linked to individual physiology (Sturrock et al., 2015). It also suits systems with predictable environmental transitions (e.g., seasonal flows, hypoxia, pollution) that produce characteristic “motifs” in chemical profiles. For example, this approach can be valuable in studies of transboundary migratory species, where chemical records reflect transitions across environmental gradients that occur at consistent developmental stages (Pérez and Hobson, 2007; Walther and Limburg, 2012). The feature extraction approach may also be relevant in studies of partial migration, where individuals from the same source diverge in movement strategy or timing (Chapman et al., 2011). In such cases, extracting features from targeted segments of the time series may outperform analyses using the full profile.

This feature-extraction strategy pairs well with machine learning classifiers such as random forests (RF), which handle large predictor sets, tolerate collinearity, and capture nonlinear relationships more effectively than traditional methods like linear (LDA) or quadratic (QDA) discriminant function analysis (Arai et al., 2023; Jones et al.,

2017; Mercier et al., 2011). In our dataset, RF outperformed LDA and QDA and was more adaptable to complex inputs, which limited the application of traditional statistical models (Supplementary Text S3). Machine learning models offer the flexibility to incorporate multiple features and capture complex, nonlinear interactions (Hastie et al., 2009). This includes not only multiple chemical measurements from a single tissue but also chemical data from multiple tissue types within the same individual (e.g., otoliths, eye lenses, muscle tissue; Rosinski et al., 2023; Walther and Torrance, 2024). These models can also integrate complementary data beyond tissue chemistry, such as otolith microstructure and early growth patterns (Barnett-Johnson et al., 2007; Kerr et al., 2025), which we included in our classification models (Supplementary Text S2), as well as genetic signatures (Brophy et al., 2020; Makhoul et al., 2025).

Explicit time series classification methods, such as DTW + KNN, were particularly effective in distinguishing sources with distinct overall profile “shapes” rather than point-specific markers. This makes DTW + KNN especially powerful in systems with unconstrained chemical baselines and variable movement timing, often shaped by dynamic environmental conditions (Teichert et al., 2024). For example, marine species inhabiting oceanographically dynamic environments may experience highly individualistic exposure histories to changes in temperature, salinity, or prey availability (Secor, 2015). In such cases, methods that retain the integrity of the entire temporal profile—rather than summarizing it through extracted features—can better capture ecologically meaningful variation. Additionally, DTW can readily handle multivariate time series (Aghabozorgi et al., 2015; Arai et al., 2024), making it well suited for cases where multi-element/isotope data are available for each individual, especially in systems where no single marker reliably reflects provenance. Although DTW can be more sensitive to noise and misalignments, they may outperform traditional feature-extraction methods when group differences manifest as patterns rather than as consistent absolute values.

Our findings highlight the advantages of ensemble methods. By combining relatively uncorrelated models (automated feature extraction, discrete time series, DTW + KNN), the ensemble model consistently outperformed individual approaches. This reflects a key advantage of ensemble methods: when component classifiers are relatively independent, as in our case, aggregating their predictions can reduce variance and improve generalization (Dietterich, 2000). Ensemble methods are especially attractive for their resistance to overfitting and ability to maintain reliable performance across varying conditions. This makes them well-suited for real-world applications like archival tissue chemistry, where reference datasets may evolve over time (e.g., Barnett-Johnson et al., 2008; Brophy et al., 2020; Rooper et al., 2008). By aggregating diverse classifiers, ensemble models provide a robust solution to temporal shifts in environmental baselines or analytical protocols, helping ensure consistent performance as data structures change.

Another major advantage of our framework is the automation and consistency it offers. Manual chemical profile inspection is time-intensive and subjective (e.g., Brennan et al., 2015; Sturrock et al., 2020; Willmes et al., 2018). Despite all readers being trained and experienced, our reader trial demonstrated notable variability in natal region identification, with the starting position of the natal region differing significantly among readers (Supplementary Text S4; Table S4). These differences resulted in up to 40% disagreement in natal assignments among analysts. Such subjectivity introduces uncertainty that can obscure biological signals and undermine reproducibility, particularly when multiple analysts or laboratories are involved. In contrast, automated approaches delivered consistent, reproducible assignments with equal or better accuracy. This standardization facilitates scaling up to larger datasets or long-term monitoring applications where manual approaches become impractical and inconsistent.

4.2. Real-world performance and management implications

Accurate provenance estimates are essential for spatial management, conservation planning, and food traceability across taxa including birds, mammals, and fishes (Bataille et al., 2020; Coelho et al., 2017; Hobson, 1999; Johnson et al., 2016). This is particularly critical for Central Valley Chinook salmon, which exhibit diverse life history strategies—spatially distinct spawning regions and variable juvenile migration timing—that buffer the populations against perturbations through the “portfolio effect” (Schindler et al., 2010). Reliable natal origin and population composition estimates are therefore key to understanding how this diversity supports persistence of this highly impacted stock under stressors like drought and altered flows (Sturrock et al., 2020).

Applying a suite of assignment models to unknown-origin juveniles revealed substantial differences in population composition estimates among methods (Fig. 5; Fig. S4). Traditional approaches, like the user-defined median, tended to assign hatchery-origin MOH samples to a natural source (FEA), reflecting its poor classification accuracy for MOH compared to time series-based methods. Such misclassifications could bias management by underestimating hatchery contributions and overestimating natural sources, potentially misguiding restoration priorities, hatchery release strategies, and viability assessments (Johnson et al., 2012).

Analysis of known-origin adult returns further highlight the value of including synthetic early migrants in the training data (Fig. 6). Simulated data especially benefited automated approaches, such as the automated feature extraction and discrete time series approach, showing substantial improvements in classification accuracy. In contrast, user-defined approaches benefited less, likely because expert readers can visually interpret chemical profiles and accurately assign natal origin, even when specific patterns are missing from the empirical reference. Thus, synthetic data designed to capture underrepresented life history variation in empirical samples are especially valuable when expert curation is impractical, or when scaling to larger, more complex datasets.

Incorporating synthetic early migrants also improved population composition estimates (Fig. S6). Excluding them from the training set led to the underrepresentation of early-migrant sources (e.g., AME) and the overrepresentation of others (e.g., FEA) that overlap isotopically with Delta habitats used by early migrants. This mismatch suggests that critical contributors, such as early migrants from AME, may be overlooked if reference datasets fail to capture the full spectrum of life-history strategies. Such omissions could have important consequences for habitat restoration efforts, as they risk excluding ecologically meaningful but underrepresented contributors to the population. In this context, generating synthetic data based on well-established biological principles and local ecological knowledge can be essential for building more complete baselines (but see Neubauer et al., 2013 for a Bayesian alternative).

While our approach substantially improved classification performance compared to traditional methods, accuracy remained relatively low for some sources, particularly among geographically adjacent rivers (e.g., BAT, MIL, DEE) or between natural and hatchery-origin fish within the same river system (e.g., AME–NIH; MER–MEH; Fig. S3), reflecting overlap in $^{87}\text{Sr}/^{86}\text{Sr}$ signatures. Incorporating additional otolith chemical tracers (e.g., Sr/Ca, Ba/Ca, $\delta^{18}\text{O}$; Reis-Santos et al., 2023) or complementary tissue chemistry (e.g., eye lens $\delta^{34}\text{S}$; Bell-Tilcock et al., 2021) could help better distinguish such sources. Furthermore, when translating individual-level assignments into population-level composition estimates (Fig. 5), it is critical to propagate both individual- and model-level uncertainty. Future work could address this through mixture model frameworks that explicitly account for uncertainty and ensure population estimates reflect underlying model uncertainty (Pella and Masuda, 2001; Smith and Campana, 2010). Finally, combining our nominal assignment approach with continuous “isoscapes” (Bowen, 2010; Brennan et al., 2025) would enable provenance assignment along

a spatial continuum rather than among discrete sources, and expand applications from single time-point assignments to lifetime movement reconstructions (Brennan et al., 2019a).

5. Conclusion

By integrating machine learning, time series classification, and ensemble modeling, we developed a novel framework for provenance assignment that advances beyond traditional expert interpretation and static summary metrics. Our combination of time series feature extraction, dynamic time warping with supervised classification, ensemble approaches, and synthetic training data represents a significant step forward in applying archival tissue chemistry to ecological questions.

Although no single method performed best across all natal sources, time series feature extraction and ensemble approaches consistently showed strong, robust performance. Optimal method choice will depend on study goals, data structure, and biological context. Together, our findings provide a practical roadmap for selecting and applying provenance assignment strategies and demonstrate the potential for machine learning and time series tools to advance ecological inference from chemical chronologies.

CRediT authorship contribution statement

Kohma Arai: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Malte Willmes:** Writing – review & editing, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Rachel C. Johnson:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Anna M. Sturrock:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Funding

Funding for analyses and personnel was provided by the California Department of Fish and Wildlife Proposition 1 Grant (Ecosystem Restoration Program, Water Quality, Supply, and Infrastructure Improvement Act of 2014, CWC §79,707[g]) [Contract number: P1596028] and by a UK Research and Innovation Future Leaders Fellowship (MR/V023578/1) awarded to AMS. Additional salary support for KA was provided by the California State Water Resources Control Board (Agreement Number: 20–036–300-2).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank the many individuals and organizations who contributed samples, data, or support for this study.

Reference juvenile and adult otoliths and associated chemical data

We are grateful to Flora Cordoleani (UC Santa Cruz), Carson Jeffres (UC Davis), George Whitman (UC Davis), Justin Glessner (UC Davis Interdisciplinary Center for Plasma Mass Spectrometry), Pedro Morais (UC Berkeley); the California Department of Fish and Wildlife (CDFW; Rob Titus, Tim Heyne, Doug Killam, Clint Garman, Tracy McReynolds,

Grantton Henley, Matt Johnson, Ryan Revnak); Pacific States Marine Fisheries Commission (PSMFC); US Fish and Wildlife Service (USFWS; Laurie Earley, Kevin Niemela, Kevin Offill); California Department of Water Resources (DWR; Virginia Afentoulis, Ryon Kurth, Jason Kindopp); Merced Fish Facility (Mary Serr); East Bay Municipal Utility District (EBMUD; Michelle Workman, Ed Ribble); Nimbus Hatchery (Jason Julienne, Gary Novak); FishBio (Jason Guignard, Andrea Fuller); Natalie Stauffer-Olsen (Trout Unlimited); and Cramer Fish Sciences (Joe Merz, Kirsten Sellheim).

Chippis Island juvenile fish.

We thank the Interagency Ecological Program Delta Juvenile Fish Monitoring Program—USFWS, CDFW, and DWR—for collecting these samples (Jack Ingram, Cory Graham, Jeff McClain, Denise Barnard, Pat Brandes).

Adult otoliths from the Yuba, Stanislaus, and American Rivers.

For the American River analyses, we thank Cramer Fish Sciences (Joe Merz, Kirsten Sellheim, Jamie Sweeney) and the Sacramento Water Forum. For the Stanislaus River analyses, we acknowledge CDFW (Tim Heyne, Steve Tsao, Crystal Sinclair, Gretchen Murphey, Shelly Schubert, Dom Giudice) and the Central Valley Project Improvement Act and USFWS (contract no. 81332-08-G017 to RCJ) under US Bureau of Reclamation Agreement R09AC20043. For the Yuba River analyses, we thank PSMFC (Duane Massa, Colin Laubach) and the Yuba River Management Team.

Otolith extraction, preparation, and analysis.

We appreciate the many people who supported otolith extraction, preparation, and analysis, including Keiko Mertz, Mollie Ogaz, Laura Coleman, Kelly Neal, Dana Myers, Sierra Schluep, and Bradyn O'Connor.

Additional contributions.

We thank Jonathan Walter, Clement Bataille, and the two anonymous reviewers for their insightful comments on the initial draft, which greatly improved this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ecoinf.2026.103604>.

Data availability

Data and code available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.3bk3j9kz3>

References

- Aghabozorgi, S., Shirkhorshidi, A.S., Wah, T.Y., 2015. Time-series clustering – a decade review. *Inf. Syst.* 53, 16–38. <https://doi.org/10.1016/j.is.2015.04.007>.
- Akosa, J., 2017. Predictive accuracy: a misleading performance measure for highly imbalanced data. In: *Proceedings of the SAS Global Forum*, 12.
- Arai, K., Castonguay, M., Secor, D.H., 2021. Multi-decadal trends in contingent mixing of Atlantic mackerel (*Scomber scombrus*) in the Northwest Atlantic from otolith stable isotopes. *Sci. Rep.* 11, 6667. <https://doi.org/10.1038/s41598-021-86116-2>.
- Arai, K., Castonguay, M., Lyubchich, V., Secor, D.H., 2023. Integrating machine learning with otolith isoscapes: reconstructing connectivity of a marine fish over four decades. *PloS One* 18, e0285702. <https://doi.org/10.1371/journal.pone.0285702>.
- Arai, K., Best, J., Craig, C., Lyubchich, V., Miller, N., Secor, D., 2024. Early growth and environmental conditions control partial migration of an estuarine-dependent fish. *Mar. Ecol. Prog. Ser.* 732, 149–166. <https://doi.org/10.3354/meps14551>.
- Artetxe-Arrate, I., Fraile, I., Lastra-Luque, P., Farley, J., Clear, N., Shahid, U., Razzaque, S.A., Ahusan, M., Vidot, A., Parker, D., Marsac, F., Murua, H., Merino, G., Zudaire, I., 2025. Otolith stable isotopes highlight the importance of local nursery areas as the origin of recruits to yellowfin tuna (*Thunnus albacares*) fisheries in the western Indian Ocean. *Fish. Res.* 281, 107241. <https://doi.org/10.1016/j.fishres.2024.107241>.
- Avigliano, 2022. Optimizing the methodological design in fish stock delineation from otolith chemistry: review of spatio-temporal analysis scales. *Rev. Fisher. Sci. Aquac.* 30 (3), 330–345. <https://doi.org/10.1080/23308249.2021.1961679>.
- Bagnall, A., Lines, J., Bostrom, A., Large, J., Keogh, E., 2017. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Min. Knowl. Disc.* 31, 606–660. <https://doi.org/10.1007/s10618-016-0483-9>.
- Barnett-Johnson, R., Grimes, C.B., Royer, C.F., Donohoe, C.J., 2007. Identifying the contribution of wild and hatchery Chinook salmon (*Oncorhynchus tshawytscha*) to the ocean fishery using otolith microstructure as natural tags. *Can. J. Fish. Aquat. Sci.* 64, 1683–1692. <https://doi.org/10.1139/f07-129>.
- Barnett-Johnson, R., Pearson, T.E., Ramos, F.C., Grimes, C.B., Bruce MacFarlane, R., 2008. Tracking natal origins of salmon using isotopes, otoliths, and landscape geology. *Limnol. Oceanogr.* 53, 1633–1642. <https://doi.org/10.4319/lo.2008.53.4.1633>.
- Bataille, C.P., Bowen, G.J., 2012. Mapping $^{87}\text{Sr}/^{86}\text{Sr}$ variations in bedrock and water for large scale provenance studies. *Chem. Geol.* 304–305, 39–52. <https://doi.org/10.1016/j.chemgeo.2012.01.028>.
- Bataille, C.P., Crowley, B.E., Wooller, M.J., Bowen, G.J., 2020. Advances in global bioavailable strontium isoscapes. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 555, 109849. <https://doi.org/10.1016/j.palaeo.2020.109849>.
- Bell-Tilcock, M., Jeffres, C.A., Rypel, A.L., Sommer, T.R., Katz, J.V.E., Whitman, G., Johnson, R.C., 2021. Advancing diet reconstruction in fish eye lenses. *Methods Ecol. Evol.* 12, 449–457. <https://doi.org/10.1111/2041-210X.13543>.
- Bowen, G.J., 2010. Isoscapes: spatial pattern in isotopic biogeochemistry. *Annu. Rev. Earth Planet. Sci.* 38, 161–187. <https://doi.org/10.1146/annurev-earth-040809-152429>.
- Brennan, S.R., Zimmerman, C.E., Fernandez, D.P., Cerling, T.E., McPhee, M.V., Wooller, M.J., 2015. Strontium isotopes delineate fine-scale natal origins and migration histories of Pacific salmon. *Sci. Adv.* 1, e1400124. <https://doi.org/10.1126/sciadv.1400124>.
- Brennan, S.R., Cline, T.J., Schindler, D.E., 2019a. Quantifying habitat use of migratory fish across riverscapes using space-time isotope models. *Methods Ecol. Evol.* 10, 1036–1047. <https://doi.org/10.1111/2041-210X.13191>.
- Brennan, S.R., Schindler, D.E., Cline, T.J., Walsworth, T.E., Buck, G., Fernandez, D.P., 2019b. Shifting habitat mosaics and fish production across river basins. *Science* 364, 783–786. <https://doi.org/10.1126/science.aav4313>.
- Brennan, K.G., Brennan, S.R., Cline, T., Bowen, G.J., 2025. Delineating population structure of resilient sea/river-type sockeye salmon. *Limnol. Oceanogr. Letters* 10, 223–233. <https://doi.org/10.1002/lo.12437>.
- Brophy, D., Rodríguez-Ezpeleta, N., Fraile, I., Arrizabalaga, H., 2020. Combining genetic markers with stable isotopes in otoliths reveals complexity in the stock structure of Atlantic bluefin tuna (*Thunnus thynnus*). *Sci. Rep.* 10, 14675. <https://doi.org/10.1038/s41598-020-71355-6>.
- Carter, J.F., Chesson, L.A., 2017. *Food Forensics: Stable Isotopes as a Guide to Authenticity and Origin*. CRC Press, USA.
- Chapman, B.B., Brönmark, C., Nilsson, J.-Å., Hansson, L.-A., 2011. The ecology and evolution of partial migration. *Oikos* 120, 1764–1775. <https://doi.org/10.1111/j.1600-0706.2011.20131.x>.
- Chatfield, C., Xing, H., 2019. *The Analysis of Time Series: An Introduction with R*, Seventh edition. CRC Press/Taylor and Francis Group, Boca Raton. Chapman & Hall/CRC texts in statistical science series.
- Chittaro, P., Hegg, J., Fuhrman, A., Robichaux, D., Doung, R., Beckman, B., Fisher, C., Vervoort, J., Kennedy, B., 2023. The incorporation of environmentally derived $^{87}\text{Sr}/^{86}\text{Sr}$ and Sr/ca in early otolith formation of Chinook salmon. *Ecol. Freshw. Fish* 32, 107–119. <https://doi.org/10.1111/eff.12673>.
- Coelho, I., Castanheira, I., Bordado, J.M., Donard, O., Silva, J.A.L., 2017. Recent developments and trends in the application of strontium and its isotopes in biological related fields. *Trends Anal. Chem.* 90, 45–61. <https://doi.org/10.1016/j.trac.2017.02.005>.
- Cordoleani, F., Phillis, C.C., Sturrock, A.M., FitzGerald, A.M., Malkasian, A., Whitman, G.E., Weber, P.K., Johnson, R.C., 2021. Threatened salmon rely on a rare life history strategy in a warming landscape. *Nat. Clim. Chang.* 11, 982–988. <https://doi.org/10.1038/s41558-021-01186-4>.
- Cowen, R.K., Sponaugle, S., 2009. Larval dispersal and marine population connectivity. *Ann. Rev. Mar. Sci.* 1, 443–466. <https://doi.org/10.1146/annurev.marine.010908.163757>.
- Diallo, R., Edalo, C., Awe, O.O., 2025. Machine learning evaluation of imbalanced health data: a comparative analysis of balanced accuracy, MCC, and F1 score. In: Awe, O.O., Vance, A. (Eds.), *Practical Statistical Learning and Data Science Methods*. Springer, Cham, STEAM-H: Science, Technology, Engineering, Agriculture, Mathematics & Health.
- Dieterich, T.G., 2000. Ensemble methods in machine learning. In: *Multiple Classifier Systems. Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–15. https://doi.org/10.1007/3-540-45014-9_1.
- Doubleday, Z.A., Martino, J.C., Trueman, C., 2022. Harnessing universal chemical markers to trace the provenance of marine animals. *Ecol. Indic.* 144, 109481. <https://doi.org/10.1016/j.ecolind.2022.109481>.
- Douglas, T.A., Blum, J.D., Guo, L., Keller, K., Gleason, J.D., 2013. Hydrogeochemistry of seasonal flow regimes in the Chena River, a subarctic watershed draining discontinuous permafrost in interior Alaska (USA). *Chem. Geol.* 335, 48–62. <https://doi.org/10.1016/j.chemgeo.2012.10.045>.
- Dunnigan, J.L., Linley, T.J., Janak, J.M., Nims, M.K., Garavelli, L., McMichael, G.A., 2023. Otolith $^{87}\text{Sr}/^{86}\text{Sr}$ identifies natal origin, movement and life history of Burbot *Lota lota* in the Kootenai River following 45 years of impoundment. *Ecol. Freshw. Fish* 32, 618–632. <https://doi.org/10.1111/eff.12712>.
- Elsdon, T.S., Wells, B.K., Campana, S.E., Gillanders, B.M., Jones, C.M., Limburg, K.E., Secor, D.H., Thorrold, S.R., Walther, B.D., 2008. Otolith chemistry to describe movements and life-history parameters of fishes: hypotheses, assumptions, limitations and inferences. *Oceanogr. Mar. Biol. Ann. Rev.* 46, 297–330.
- Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., Muller, P.-A., 2019. Deep learning for time series classification: a review. *Data Min. Knowl. Disc.* 33, 917–963. <https://doi.org/10.1007/s10618-019-00619-1>.

- Giorgino, T., 2009. Computing and visualizing dynamic time warping alignments in R: the dtw package. *J. Stat. Soft.* 31. <https://doi.org/10.18637/jss.v031.i07>.
- Gonzabato, N., 2025. *manymodelr*: build and tune several models. R package version 0.3.9. <https://doi.org/10.32614/CRAN.package.manymodelr>.
- Hanson, N.N., Wurster, C.M., Eimf Todd, C.D., 2010. Comparison of secondary ion mass spectrometry and micromilling/continuous flow isotope ratio mass spectrometry techniques used to acquire intra-otolith $\delta^{18}\text{O}$ values of wild Atlantic salmon (*Salmo salar*). *Rapid Commun. Mass Spectrom.* 24, 2491–2498. <https://doi.org/10.1002/rcm.4646>.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media.
- Hegg, J.C., Kennedy, B.P., 2021. Let's do the time warp again: non-linear time series matching as a tool for sequentially structured data in ecology. *Ecosphere* 12. <https://doi.org/10.1002/ecs2.3742>.
- Hegg, J.C., Kennedy, B.P., Chittaro, P., 2019. What did you say about my mother? The complexities of maternally derived chemical signatures in otoliths. *Can. J. Fish. Aquat. Sci.* 76, 81–94. <https://doi.org/10.1139/cjfas-2017-0341>.
- Hobson, K.A., 1999. Tracing origins and migration of wildlife using stable isotopes: a review. *Oecologia* 120, 314–326. <https://doi.org/10.1007/s004420050865>.
- Huber, E.R., Carlson, S.M., 2015. Temporal trends in hatchery releases of fall-run Chinook salmon in California's Central Valley. *San Franc. Estuary. Watershed. Sci.* 13. <https://doi.org/10.15447/sfews.2015v13iss2art3>.
- Hüssy, K., Limburg, K.E., de Pontual, H., Thomas, O.R.B., Cook, P.K., Heimbrand, Y., Blass, M., Sturrock, A.M., 2020. Trace element patterns in otoliths: the role of biomineralization. *Rev. Fish. Sci. Aquac.* <https://doi.org/10.1080/23308249.2020.1760204>.
- Johnson, R.C., Weber, P.K., Wikert, J.D., Workman, M.L., MacFarlane, R.B., Grove, M.J., Schmitt, A.K., 2012. Managed metapopulations: do salmon hatchery 'sources' lead to in-river 'sinks' in conservation? *PLoS One* 7, e28880. <https://doi.org/10.1371/journal.pone.0028880>.
- Johnson, R., Garza, J., MacFarlane, R., Grimes, C., Phillis, C., Koch, P., Weber, P., Carr, M., 2016. Isotopes and genes reveal freshwater origins of Chinook salmon *Oncorhynchus tshawytscha* aggregations in California's coastal ocean. *Mar. Ecol. Prog. Ser.* 548, 181–196. <https://doi.org/10.3354/meps11623>.
- Jones, C.M., Palmer, M., Schaffler, J.J., 2017. Beyond Zar: the use and abuse of classification statistics for otolith chemistry: otolith chemistry classification statistics. *J. Fish Biol.* 90, 492–504. <https://doi.org/10.1111/jfb.13051>.
- Kerr, L.A., Whitener, Z., Becker, S., Goethel, D., Zemeckis, D.R., Sherwood, G., Kovach, A.I., Cadrin, S.X., 2025. Stock identification of sympatric Atlantic cod populations in the Gulf of Maine and mixed stock fishery analysis using otolith-based techniques. *Can. J. Fish. Aquat. Sci.* 82, 1–18. <https://doi.org/10.1139/cjfas-2023-0159>.
- Kuhn, M., Wickham, H., 2020. *TidyModels: A Collection of Packages for Modeling and Machine Learning Using Tidyverse Principles*.
- Lewis, L.S., Willmes, M., Denney, C., Cavole, L., Xie, W., Fichman, R.A., Hung, T.-C., Ellison, L., Stevenson, T., Schultz, A., Hammock, B., Teh, S., Hobbs, J.A., 2025. Validating estuarine movements from otoliths for an imperiled fish. *Aquat. Biol.* 34, 1–13. <https://doi.org/10.3354/ab00773>.
- Makhlof, B., Cline, T.J., Fernandez, D., Seeb, L., Lee, E., Gilk-Baumer, S., Whited, D., Zimmerman, C.E., Schindler, D.E., 2025. Combining genetic and isotope frameworks improves reconstruction of fish provenance across riverscapes. *Limnol. Oceanogr. Lett.* <https://doi.org/10.1002/lol2.70025>.
- Mercier, L., Darnaude, A.M., Bruguier, O., Vasconcelos, R.P., Cabral, H.N., Costa, M.J., Lara, M., Jones, D.L., Mouillot, D., 2011. Selecting statistical models and variable combinations for optimal classification using otolith microchemistry. *Ecol. Appl.* 21, 1352–1364. <https://doi.org/10.1890/09-1887.1>.
- Neubauer, P., Shima, J.S., Sweater, S.E., 2013. Inferring dispersal and migrations from incomplete geochemical baselines: analysis of population structure using Bayesian infinite mixture models. *Methods Ecol. Evol.* 4, 836–845. <https://doi.org/10.1111/2041-210X.12076>.
- Newsome, S.D., Clementz, M.T., Koch, P.L., 2010. Using stable isotope biogeochemistry to study marine mammal ecology. *Mar. Mamm. Sci.* <https://doi.org/10.1111/j.1748-7692.2009.00354.x>.
- Pella, J., Masuda, M., 2001. Bayesian methods for analysis of stock mixtures from genetic characters. *Fish. Bull.* 99, 151–167.
- Pérez, G.E., Hobson, K.A., 2007. Feather deuterium measurements reveal origins of migratory western loggerhead shrikes (*Lanius ludovicianus excubitorides*) wintering in Mexico. *Divers. Distrib.* 13, 166–171. <https://doi.org/10.1111/j.1472-4642.2006.00306.x>.
- R Core Team, 2025. *R: A Language and Environment for Statistical Computing*.
- Reich, M.S., Ghouri, S., Zabudsky, S., Hu, L., Le Corre, M., Ng'iru, I., Benyamini, D., Shipilina, D., Collins, S.C., Martins, D.J., Vila, R., Talavera, G., Bataille, C.P., 2024. Trans-Saharan migratory patterns in *Vanessa cardui* and evidence for a southward leapfrog migration. *iScience* 27, 111342. <https://doi.org/10.1016/j.isci.2024.111342>.
- Reis-Santos, P., Gillanders, B.M., Sturrock, A.M., Izzo, C., Oxman, D.S., Lueders-Dumont, J.A., Hüssy, K., Tanner, S.E., Rogers, T., Doubleday, Z.A., Andrews, A.H., Trueman, C., Brophy, D., Thiem, J.D., Baumgartner, L.J., Willmes, M., Chung, M.-T., Charapata, P., Johnson, R.C., Trumble, S., Heimbrand, Y., Limburg, K.E., Walther, B.D., 2023. Reading the biomineralized book of life: expanding otolith biogeochemical research and applications for fisheries and ecosystem-based management. *Rev. Fish Biol. Fish.* 33, 411–449. <https://doi.org/10.1007/s11160-022-09720-z>.
- Rooker, J., Secor, D., DeMetrio, G., Kaufman, A., Belmonte Ríos, A., Ticina, V., 2008. Evidence of trans-Atlantic movement and natal homing of bluefin tuna from stable isotopes in otoliths. *Mar. Ecol. Prog. Ser.* 368, 231–239. <https://doi.org/10.3354/meps07602>.
- Rooker, J., Arrizabalaga, H., Fraile, I., Secor, D., Dettman, D., Abid, N., Addis, P., Deguara, S., Karakulak, F., Kimoto, A., Sakai, O., Macías, D., Santos, M., 2014. Crossing the line: migratory and homing behaviors of Atlantic bluefin tuna. *Mar. Ecol. Prog. Ser.* 504, 265–276. <https://doi.org/10.3354/meps10781>.
- Rosinski, C.L., Glaid, J., Hahn, M., Fetzer, W.W., 2023. Natal origin differentiation using eye lens stable isotope analysis. *N. Am. J. Fish Manag.* 43, 547–555. <https://doi.org/10.1002/nafm.10875>.
- Schindler, D.E., Hilborn, R., Chasco, B., Boatright, C.P., Quinn, T.P., Rogers, L.A., Webster, M.S., 2010. Population diversity and the portfolio effect in an exploited species. *Nature* 465, 609–612. <https://doi.org/10.1038/nature09060>.
- Secor, D.H., 1999. Specifying divergent migrations in the concept of stock: the contingent hypothesis. *Fish. Res.* 43, 13–34. [https://doi.org/10.1016/S0165-7836\(99\)00064-8](https://doi.org/10.1016/S0165-7836(99)00064-8).
- Secor, D.H., 2015. *Migration Ecology of Marine Fishes*. Johns Hopkins University Press, Baltimore (Md.).
- Sigurðardóttir, A.R., Sverrisson, P., Jónsdóttir, A., Guðjónsdóttir, M., Elvarsson, B.P., Einarsson, H., 2023. Otolith age determination with a simple computer vision based few-shot learning method. *Ecol. Inform.* 76, 102046. <https://doi.org/10.1016/j.ecoinf.2023.102046>.
- Smith, S.J., Campana, S.E., 2010. Integrated stock mixture analysis for continuous and categorical data, with application to genetic-otolith combinations. *Can. J. Fish. Aquat. Sci.* 67, 1533–1548. <https://doi.org/10.1139/F10-078>.
- Sturrock, A.M., Hunter, E., Milton, J.A., Johnson, R.C., Waring, C.P., Trueman, C.N., 2015. Quantifying physiological influences on otolith microchemistry. *Methods Ecol. Evol.* 6, 806–816. <https://doi.org/10.1111/2041-210X.12381>.
- Sturrock, A.M., Satterthwaite, W.H., Cervantes-Yoshida, K.M., Huber, E.R., Sturrock, H.J. W., Nusslé, S., Carlson, S.M., 2019. Eight decades of hatchery salmon releases in the California Central Valley: factors influencing straying and resilience. *Fisheries* 44, 433–444. <https://doi.org/10.1002/fsh.10267>.
- Sturrock, A.M., Carlson, S.M., Wikert, J.D., Heyne, T., Nusslé, S., Merz, J.E., Sturrock, H.J.W., Johnson, R.C., 2020. Unnatural selection of salmon life histories in a modified riverscape. *Glob. Chang. Biol.* 26, 1235–1247. <https://doi.org/10.1111/gcb.14896>.
- Takai, K., Kuroki, M., Shirai, K., Manabu, K., Murakami, H., Terashima, Y., Mitamura, H., Yamashita, Y., 2024. River-sea migration patterns of juvenile Japanese seabass influence migrations in subsequent stages. *Estuar. Coast. Shelf Sci.* 307, 108921. <https://doi.org/10.1016/j.ecss.2024.108921>.
- Teichert, N., Tabouret, H., Lizé, A., Davaert, F., Acou, A., Trancart, T., Virag, L.-S., Pécheyran, C., Feunteun, E., Carpentier, A., 2024. Quantifying larval dispersal portfolio in seabass nurseries using otolith chemical signatures. *Mar. Environ. Res.* 196, 106426. <https://doi.org/10.1016/j.marenvres.2024.106426>.
- Quaek, 2017. *Stable isotope analysis of fish eye lenses: reconstruction of ontogenetic trends in spatial and trophic ecology of elasmobranchs and deep water teleosts*. PhD thesis. University of Southampton.
- van Dijk, J.G.B., Meissner, W., Klaassen, M., 2014. Improving provenance studies in migratory birds when using feather hydrogen stable isotopes. *J. Avian Biol.* 45, 103–108. <https://doi.org/10.1111/j.1600-048X.2013.00232.x>.
- Vander Zanden, H.B., Nelson, D.M., Wunder, M.B., Conkling, T.J., Katzner, T., 2018. Application of isoscapes to determine geographic origin of terrestrial wildlife for conservation and management. *Biol. Conserv.* 228, 268–280. <https://doi.org/10.1016/j.biocon.2018.10.019>.
- Vignon, M., 2015. Extracting environmental histories from sclerochronological structures — recursive partitioning as a mean to explore multi-elemental composition of fish otolith. *Ecol. Inform.* 30, 159–169. <https://doi.org/10.1016/j.ecoinf.2015.10.002>.
- Voss, B.M., Peucker-Ehrenbrink, B., Eglinton, T.I., Fiske, G., Wang, Z.A., Hoering, K.A., Montluçon, D.B., LeCroy, C., Pal, S., Marsh, S., Gillies, S.L., Janmaat, A., Bennett, M., Downey, B., Fanslau, J., Fraser, H., Macklam-Harron, G., Martinec, M., Wiebe, B., 2014. Tracing river chemistry in space and time: dissolved inorganic constituents of the Fraser River, Canada. *Geochim. Cosmochim. Acta* 124, 283–308. <https://doi.org/10.1016/j.gca.2013.09.006>.
- Walther, B.D., Limburg, K.E., 2012. The use of otolith chemistry to characterize diadromous migrations. *J. Fish Biol.* 81, 796–825. <https://doi.org/10.1111/j.1095-8649.2012.03371.x>.
- Walther, B.D., Torrance, L.E., 2024. Quantifying euryhaline histories in red drum *Sciaenops ocellatus*: otolith chemistry and muscle isotope ratios. *J. Fish Biol.* 105, 1389–1405. <https://doi.org/10.1111/jfb.15173>.
- Wang, X., Mueen, A., Ding, H., Trajcevski, G., Scheuermann, P., Keogh, E., 2013. Experimental comparison of representation methods and distance measures for time series data. *Data Min. Knowl. Disc.* 26, 275–309. <https://doi.org/10.1007/s10618-012-0250-5>.
- Williams, J.G., 2006. Central Valley salmon: a perspective on Chinook and steelhead in the Central Valley of California. *San Franc. Estuary Watershed Sci.* 4. <https://doi.org/10.15447/sfews.2006v4iss3art2>.
- Willmes, M., Hobbs, J.A., Sturrock, A.M., Bess, Z., Lewis, L.S., Glessner, J.J.G., Johnson, R.C., Kurth, R., Kindopp, J., 2018. Fishery collapse, recovery, and the cryptic decline of wild salmon on a major California river. *Can. J. Fish. Aquat. Sci.* 75, 1836–1848. <https://doi.org/10.1139/cjfas-2017-0273>.
- Willmes, M., Sturrock, A.M., Cordoleani, F., Hugentobler, S., Meek, M.H., Whitman, G., Evans, K., Palkovacs, E.P., Stauffer-Olsen, N.J., Johnson, R.C., 2024. Integrating otolith and genetic tools to reveal intraspecific biodiversity in a highly impacted salmon population. *J. Fish Biol.* 105, 412–430. <https://doi.org/10.1111/jfb.15847>.
- Wright, M.N., Ziegler, A., 2017. Ranger: a fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Soft.* 77. <https://doi.org/10.18637/jss.v077.i01>.

- Wunder, M.B., 2012. Determining geographic patterns of migration and dispersal using stable isotopes in keratins. *J. Mammal.* 93, 360–367. <https://doi.org/10.1644/11-MAMM-S-182.1>.
- Yang, D., Bowen, G.J., Uno, K.T., Podkovyrov, K., Carpenter, N.A., Fernandez, D.P., Cerling, T.E., 2023. BITS : a Bayesian isotope turnover and sampling model for strontium isotopes in proboscideans and its potential utility in movement ecology. *Methods Ecol. Evol.* 14, 2800–2813. <https://doi.org/10.1111/2041-210X.14218>.
- Yoshiyama, R.M., Fisher, F.W., Moyle, P.B., 1998. Historical abundance and decline of Chinook salmon in the Central Valley region of California. *N. Am. J. Fish. Manag.* 18, 487–521. [https://doi.org/10.1577/1548-8675\(1998\)018%253C0487:HAADOC%253E2.0.CO;2](https://doi.org/10.1577/1548-8675(1998)018%253C0487:HAADOC%253E2.0.CO;2).