# Methods

## Additional context about the data (probably will be added to intro)

We evaluate the effectiveness of several statistical methods for processing otolith isotope data acquired through Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LA-ICPMS). Otolith data contains inherent noise from multiple sources: analytical instrument error, signal blurring due to laser spot geometry, and biological factors related to non-instantaneous isotope incorporation in calcified tissues. These factors create time series with underlying latent trends that can be difficult to interpret from raw data alone. Signal quality is further influenced by the concentration of strontium at any measurement point, with higher concentrations enabling better measurement precision. For this reason, when analyzing ⁸⁷Sr/⁸⁶Sr ratios, researchers typically collect simultaneous profiles of both isotopic ratios and ⁸⁸Sr, the latter serving as a proxy for total strontium concentration.

The primary objective of applying statistical tools to otolith LA-ICPMS data is to reveal underlying true states obscured by multiple noise sources. Researchers generally aim to either: 1) smooth the data to reveal patterns in underlying ⁸⁷Sr/⁸⁶Sr ratios, or 2) employ time series models to detect and extract discrete states in the data. The choice of statistical approach may significantly affect ecological interpretation; however, rigorous methodological testing or quantitative comparison of how different analytical tools influence data interpretation is often lacking.

### Raw and Simulated Data Structure.

The dataset for this research comprises LA-ICPMS measurements from Chinook salmon (Oncorhynchus tshawytscha) otoliths collected over several years from the Yukon, Kuskokwim, and Nushagak river basins in Alaska. These sequentially ordered time series contain approximately 800-2000 data points with varying signal-to-noise ratios. The measurements span the entire freshwater residence period and extend into the marine phase. The underlying states typically include a "core" state representing natal development within the egg, followed by several sequential habitat states, ultimately concluding with a marine value of 0.7092, which is consistent across all individuals. However, in these empirical data, both the number and isotopic values of true states remain unknown, with potentially extensive combinations of environmental signatures.

To overcome this limitation and rigorously test multiple analytical methods, we simulated data that mimics the noise characteristics and patterns of our real otolith measurements while maintaining known underlying states. Our simulation framework incorporated three key components: (1) creation of a base signal with five distinct states (core, three intermediate freshwater states, and marine) connected by sigmoid transitions; (2) addition of both white noise and autocorrelated noise to simulate analytical variation and natural biological processes; and (3) implementation of a machine blurring effect to account for instrumental signal integration and washout effects. This approach provided a controlled environment with known ground truth A graph with blue lines and dots

AI-generated content may be incorrect.against which different analytical methods could be evaluated

[ Show a real sample]

### Models

Our analysis applied four statistical approaches to process otolith LA-ICPMS data based on examples found in the literature (cite, cite, cite) as well as methods which are well suited to addressing these research questions.

### Smoothing Models

To reveal underlying patterns by reducing noise while preserving essential signal features, we implemented two distinct smoothing approaches:

***Generalized Additive Models (GAMs)****:*

We employed Generalized Additive Models (GAMs) to characterize the non-parametric patterns in strontium isotope ratios across otolith transects. GAMs are well-suited for this application as they can capture complex, non-linear relationships without assuming a specific functional form. The general model structure is represented as:

*E[Sr87/86t] = Bo + f(Distancet) + et*

Where *Bo* is the intercept, *f(Distancet)* is a smooth function of distance from the otolith core, and *et* represents the error term.

All models were constructed using the 'mgcv' package in R. To determine optimal model parameters, we conducted a systematic assessment of spline types and smoothing parameters. We evaluated four different spline bases: thin plate regression splines ("tp"), cubic regression splines ("cr"), P-splines ("ps"), and adaptive splines ("ad"). For each spline type, we tested a range of knot values (k) from 10 to 200 to examine how different levels of flexibility affected model fit. Model performance was qualitatively evaluated against simulated data.

A graph of different values

AI-generated content may be incorrect.

**State Space Models with Kalman Smoothing** provide an approach for extracting underlying trends by treating the observed isotope ratios as noisy measurements of a true, unobserved state. We implemented a local level model using the Kalman filter and smoother, which produces optimal estimates of the underlying signal trajectory. Unlike simple moving averages or splines, this method accounts for temporal dependence in the data while effectively removing measurement noise. The Kalman smoother specifically produces a smoothed line through the data by using both past and future observations to estimate each state value, making it particularly effective for LA-ICPMS data where analytical noise can obscure important transitions. We used the 'KFAS' package in R (Helske, 2017), with smoothing parameters selected to balance signal preservation and noise reduction based on the estimated noise characteristics of our empirical otolith data.

### Automatic Detection of Habitat States

To objectively identify discrete habitat states and transition points without manual interpretation, we implemented two statistical approaches:

**Hidden Markov Models (HMMs)** provide a probabilistic framework for identifying distinct states in time series data when the true states are not directly observable. HMMs characterize the otolith isotope profile as a sequence of discrete habitat states, each with its own strontium isotope distribution. The model simultaneously estimates the number of distinct habitats, their characteristic isotopic signatures, and the most likely sequence of habitat occupancy throughout the fish's life history. This approach is particularly well-suited for otolith data as it aligns with the ecological understanding that fish typically remain in discrete habitats for extended periods before transitioning to new environments, rather than constantly moving between habitats.

**Changepoint Analysis** identifies locations in the time series where statistical properties fundamentally shift, indicating potential habitat transitions. Standard changepoint detection methods identify points where the mean isotope value or variance changes significantly, representing abrupt movements between chemically distinct habitats. We also implemented Behavioral Changepoint Analysis (BCPA), which incorporates autocorrelation structure and can identify more subtle transitions, including gradual shifts that might indicate movement between chemically similar but connected waterways. These methods quantitatively determine the optimal number and location of habitat transitions, providing an objective alternative to visual interpretation of smoothed profiles.

### Approach

To systematically evaluate these statistical approaches, we will generate 100 simulated otolith time series with varying characteristics. These simulations will incorporate different combinations of state values, transition patterns, noise levels, and signal quality to represent the diversity of real-world samples. We will apply each of our four statistical methods—GAMs, State Space Models, Hidden Markov Models, and Changepoint Analysis—to all simulated time series.

Performance assessment will follow two tracks. For smoothing methods (GAMs and State Space Models), we will qualitatively evaluate how effectively they reveal the underlying pattern while reducing noise, comparing the smoothed output against the known true states in our simulations. For automatic detection methods (HMMs and Changepoint Analysis), we will quantitatively assess accuracy by measuring how correctly they identify the number of states and the timing of transitions. Metrics will include classification accuracy, precision in locating transition points, and sensitivity to subtle state changes. Following the simulation study, we will apply the smoothing methods to our real otolith data from Alaskan watersheds to demonstrate their practical utility in field applications and to compare how findings from the controlled simulations translate to actual biological samples.