# Methods

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

We evaluate the effectiveness of several statistical tools for processing otolith isotope data acquired through Laser Ablation Inductively Coupled Mass Spectrometry (LA-ICPMS). Otolith data contains inherent noise from multiple sources: analytical instrument error, signal blurring due to laser geometry, and biological factors related to non-instantaneous isotope turnover in tissues. These factors create time series with underlying latent trends that can be difficult to interpret from raw data alone. Signal quality within a sample is further influenced by the concentration of bulk strontium at any measurement point. Higher strontium availability enables better measurement resolution. When analyzing ^87^Sr/^86^Sr ratios, researchers typically collect a secondary signal of ^88^Sr—the most abundant strontium isotope—which serves as an index of total strontium concentration at each sampling point.

The primary objective of applying statistical tools to otolith LA-ICPMS data is to reveal underlying true states obscured by noise. These states represent a continuous chronological record of isotopic and chemical signals from early development through marine migration. Statistical approaches generally serve two purposes: (1) smoothing data for qualitative analysis or manual state selection, and (2) automatic detection of states for provenance determination or migration-pattern research. However, statistical model selection across applications and research groups remains inconsistent and often lacks rigorous methodological justification. This highlights the need to systematically evaluate which methods perform best under various analytical scenarios.

### Raw and Simulated Data Structure.

Our dataset consists of raw otolith LA-ICPMS measurements collected over several years from the three largest watersheds in western Alaska (Yukon, Kuskokwim, and Nushagak), representing the complete freshwater residence period of individual fish. These raw time series contain between 800-2000 data points with varying signal-to-noise ratios. The profiles typically exhibit a "core" state representing natal development in the egg, followed by several sequential habitat states, and ultimately concluding with a consistent "marine" value of 0.7092. However, the underlying true states in field-collected samples remain unknown and the potential combinations of environmental states are extensive.

To address this limitation, we simulated data that mimics the noise characteristics and patterns of real otolith measurements while maintaining known underlying states. The simulation incorporated several key components: (1) creation of a base signal with five distinct states (core, three intermediate states, and marine) connected by sigmoid transitions; (2) addition of both white noise and autocorrelated noise to mimic analytical variation and natural processes and; (3) implementation of a machine blurring effect to simulate instrumental signal integration.

[ Show a real sample alongside a simulated sample]

### Models

Our analysis will apply four statistical approaches to process otolith LA-ICPMS data. For manual extraction of states, we will use Generalized Additive Models (GAMs) that create smooth curves through the noisy data points, helping researchers visualize underlying patterns. These models will account for both the isotope ratios and strontium concentration to improve accuracy where signal quality varies. State Space Models will separate the data into meaningful components—the underlying trend representing habitat shifts, regular patterns that might occur within habitats, and random noise. This approach allows the model to adjust for changing levels of measurement error that occur along different portions of the otolith.

For automatic detection of state changes, we will implement Hidden Markov Models (HMMs) that identify distinct habitat states and transitions without human intervention. These models assume fish stay in the same habitat for periods before moving and can determine both the number of different habitats and when transitions occurred. Changepoint Analysis will identify specific points where the isotope pattern shows significant shifts, marking potential habitat changes. We'll also use a variation called Behavioral Changepoint Analysis that can detect more subtle transitions, including cases where isotope values change gradually rather than abruptly, which might indicate movements between connected waterways with similar chemical signatures.

### 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.