

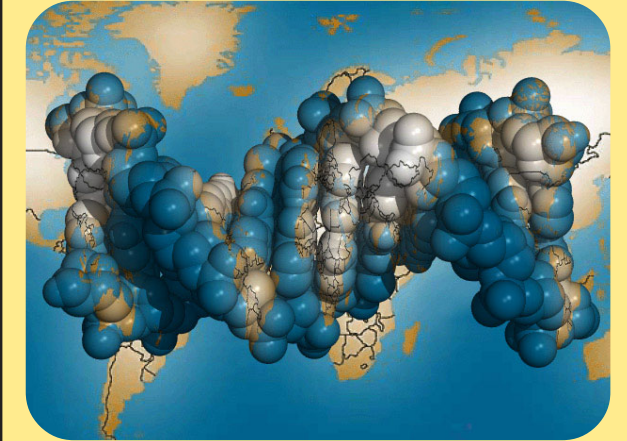


# Mixed Stock Analysis of Sockeye Salmon Allows for Detection of Small Contributions with Adequate Power: Bayesian Approach versus Maximum Likelihood

Anton Antonovich<sup>1</sup>, William Templin<sup>1</sup>, Christopher Habicht<sup>1</sup>, Joel Reynolds<sup>1,2</sup>, and James Seeb<sup>1</sup>

<sup>1</sup>Alaska Department of Fish and Game, 333 Raspberry Road, Anchorage, Alaska 99518, USA

<sup>2</sup>U.S. Fish and Wildlife Service, 1011 E. Tudor Road, Anchorage, Alaska, 99503, USA



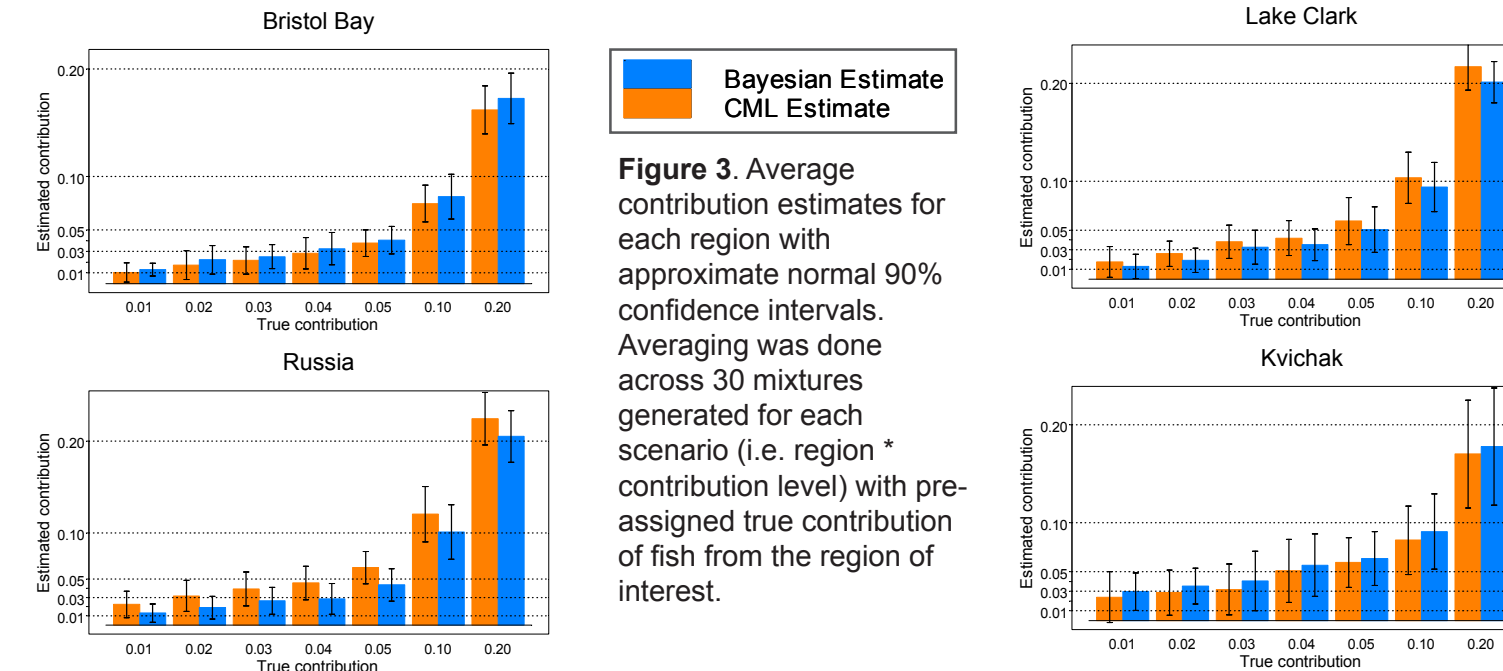
## Abstract

Genetic data have proven extremely useful in identifying stock groups in complex mixtures. In this study we used eight microsatellite loci to serve as genetic markers from 51 Bristol Bay and 12 Russian collections of sockeye salmon. We investigated the Bayesian approach to Mixed Stock Analysis (MSA) and compared its detection power to the power of the Conditional Maximum Likelihood (CML) method. For this purpose a simulation study was designed to directly compare performance of the Bayesian and CML approaches. Sensitivity of both methods to detection of small contributions in a mixture was evaluated for a wide range of known mixtures. Detection was defined as a non-zero lower limit of the 90% credibility/confidence interval of the contribution estimate. We found that the Bayesian method is capable of detecting small contributions of selected stock groups with adequate statistical power. We also found that the CML method produces similar results with a higher power at low contribution levels. However, this increase in power was attributed to the limitations of the bootstrap method used to obtain confidence intervals for CML estimates, rather than to its better performance over the Bayesian approach. At moderate and high contribution levels (> 10%) both methods have maximum power of 1.0. Our results also indicate that posterior means serve as reliable proxies for the Bayesian stock contribution estimates. They are generally less biased than the CML estimates and have smaller variability.

## Simulation Study

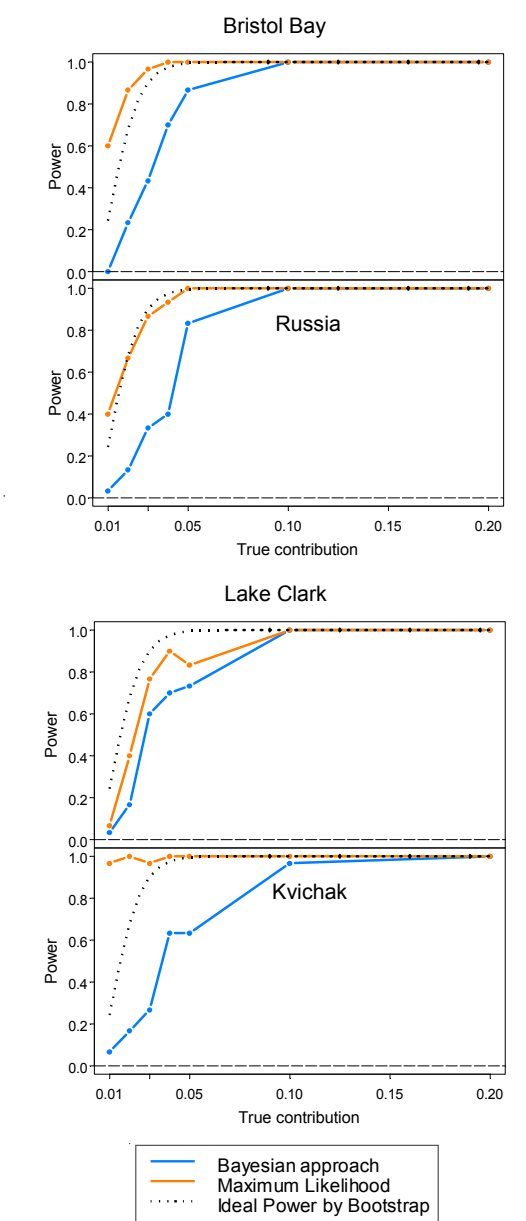
- For each region, we generated 30 random mixtures with a pre-assigned contribution of fish from that region. Seven contribution levels were considered – 1%, 2%, 3%, 4%, 5%, 10%, and 20%. The mixtures were then analyzed by both the Bayesian and Conditional Maximum Likelihood methods.
- Contribution estimates for a region of interest along with their 90% credibility/confidence intervals (CI) were obtained for all mixtures by both methods. In the Bayesian approach, posterior means were used as proxies for contribution estimates. To obtain confidence intervals for the CML estimates we used the symmetric percentile bootstrap method.
- The statistical power to detect a group of populations was defined as the proportion of times (out of 30 mixtures) that a group has been successfully detected. Detection of a group is equivalent to a hypothesis testing for a non-zero contribution:  $H_0: \theta = 0$  vs.  $H_a: \theta > 0$ . A decision to reject the null hypothesis was made if the lower limit of the 90% CI for  $\theta$  was not zero when rounded to two significant digits. With the sample size of 200 fish in the mixture, this would be equivalent to having detected just a single

## Simulation Results



**Figure 3.** Average contribution estimates for each region with approximate normal 90% confidence intervals. Averaging was done across 30 mixtures generated for each scenario (i.e. region \* contribution level) with pre-assigned true contribution of fish from the region of interest.

## Detection Power



**Figure 5.** Detection power of the Bayesian and CML methods for each region. Power was defined as proportion of times a group was successfully detected in mixtures with known composition. The dotted line is the ideal power based on the lower limit of 90% bootstrap CI for a perfectly identifiable group and a mixture of size 200.

## Baseline Development and Management Interests

Genetic data of eight microsatellite loci from 63 baseline populations of sockeye salmon (*Oncorhynchus nerka*) were used in this study; 51 of them were collected in Bristol Bay drainages (see **Figure 1**) and 12 from Russian rivers in the Kamchatka Peninsula. These populations were further grouped into 12 reporting regions according to their genetic similarities.

From a management perspective, the following two major questions stimulated this study:

- 1) How sensitive are the modern methods of MSA to detecting small contributions of Bristol Bay (vs. Russian) salmon in composite mixtures?
- 2) On a smaller scale, is it possible to detect small contributions of salmon originated from Lake Clark alone and Iliamna and Lake Clark combined (hereafter called Kvichak region) in mixtures typical for the Naknek harvest area?

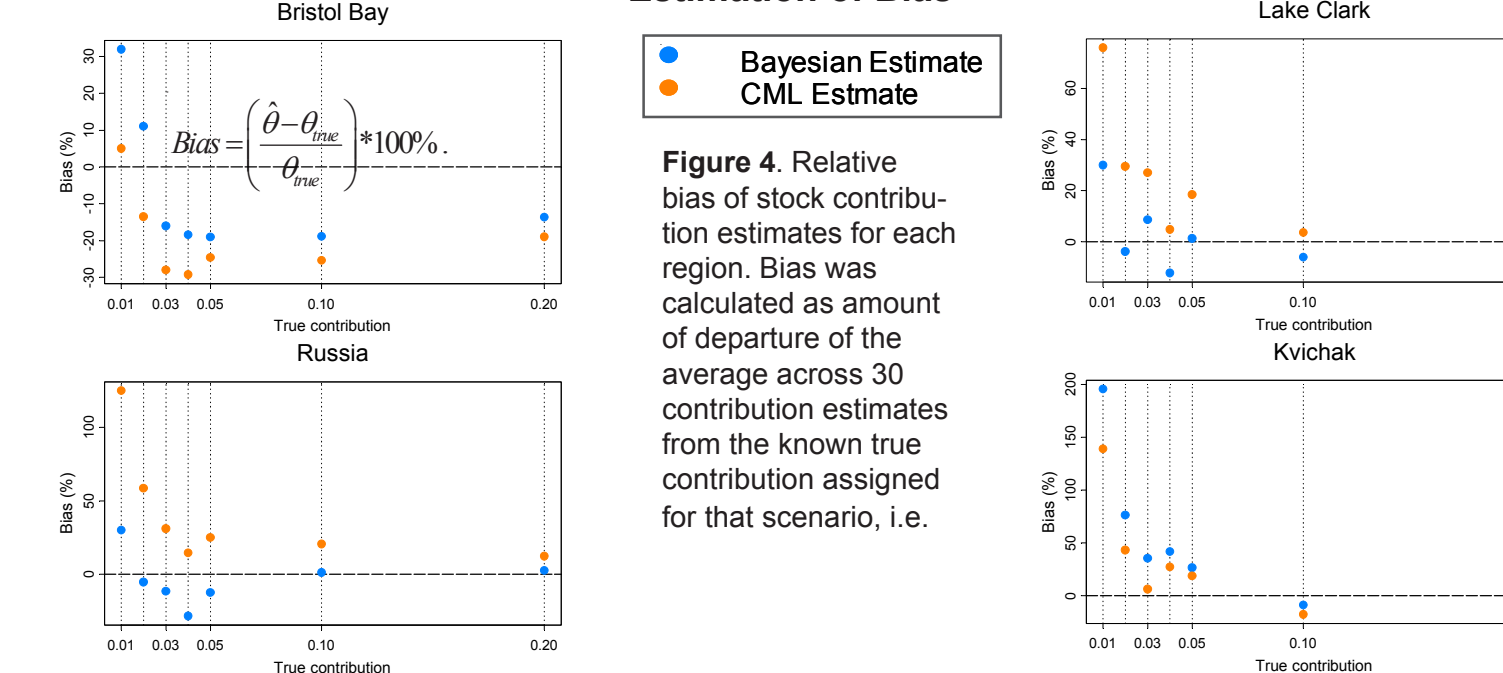
Thus, we concentrated our attention on the following four groups: Bristol Bay and Russian stocks to answer the first question, Lake Clark and Kvichak stocks for the second (**Figure 2**).



**Figure 1.** Map of the region with main Bristol Bay drainages.

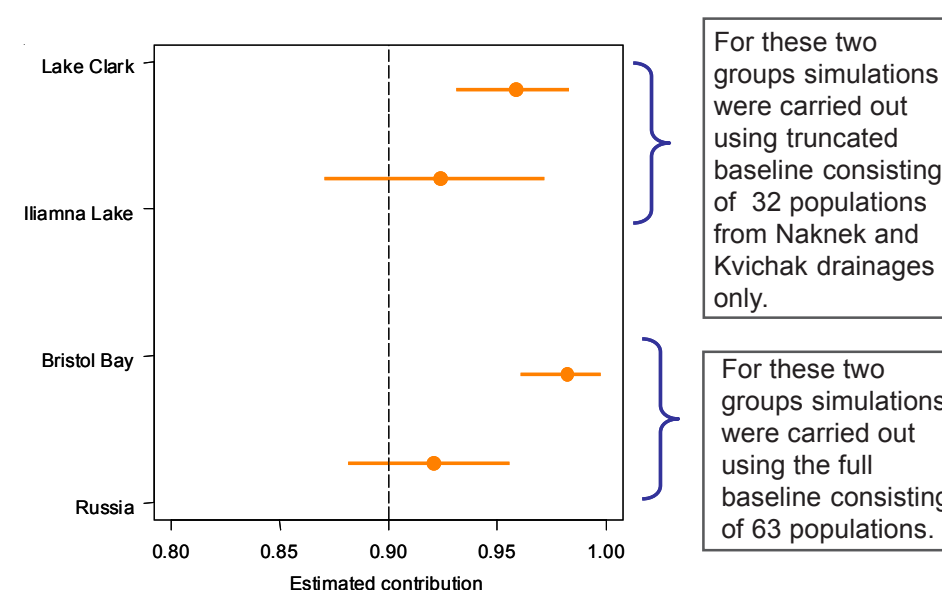


## Estimation of Bias

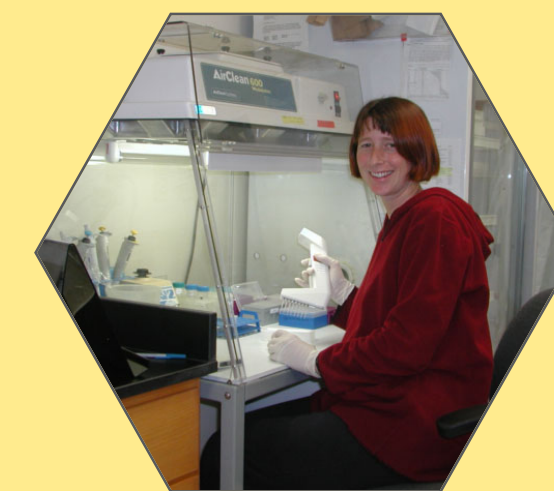


**Figure 4.** Relative bias of stock contribution estimates for each region. Bias was calculated as amount of departure of the average across 30 contribution estimates from the known true contribution assigned for that scenario, i.e.

## Simulation results with 100% contribution from a region



**Figure 2.** Mixture estimates for each region from 100% simulations with corresponding 90% bootstrap CIs. All estimates are greater than 0.9 confirming the genetic similarity within groups.



## Results

- ◆ Bayesian estimates are generally less biased and have slightly smaller variability than their CML counterparts (**Figure 3**). Relative bias tends to decrease for both methods with increase in contribution level (**Figure 4**).
- ◆ Bayesian method is capable of detecting small contributions of selected stock groups with adequate statistical power (**Figure 5**). Its detection power decreases progressively with the decline in contribution level.
- ◆ At lower contribution levels, the CML method demonstrated higher power than the Bayesian approach. At higher levels (>= 10%) both methods had maximum power of 1.
- ◆ The increase in power shown by the CML method at low contributions was attributed to biased bootstrap CIs. Since the stock detection was based on the lower limit of bootstrap CI, biased-high confidence intervals led to artificially high detection power. This phenomenon was especially pronounced for a group of Kvichak stocks, where power remained unusually high for all contribution levels (**Figure 5**).

## Conclusions

- Both Bayesian and CML approaches give similar results with Bayesian estimates being slightly less biased and variable.
- The symmetric bootstrap method (used in CML) can lead to the biased-high confidence intervals, thus artificially inflating the detection power.

## Software

- For CML simulations we used SPAM version 3.7 (Alaska Department of Fish and Game, 2003).
- For Bayesian simulations we used BAYES (Auke Bay Laboratory, Alaska Fisheries Science Center).