Comparing Mixture Estimates by Parametric Bootstrapping Likelihood Ratios

Joel H. Reynolds and William D. Templin
Gene Conservation Laboratory
Alaska Department of Fish and Game
333 Raspberry Road
Anchorage, AK 99518-1599

Author Footnote

Joel H. Reynolds is Biometrician and William D. Templin is Fisheries Geneticist with the Gene Conservation Laboratory of the Commercial Fisheries Division of Alaska Department of Fish and Game, 333 Raspberry Road, Anchorage, AK 99518-1599 (E-mail: Joel_Reynolds@fishgame.state.ak.us).

ABSTRACT

Estimating the relative contributions of distinct populations in a mixture of organisms is a common task for fisheries and wildlife managers and researchers. There is increasing interest in comparing these mixture contributions across time or space. Researchers regularly compare mixtures by checking for overlap in the interval estimates for each population contribution from each mixture. This method of comparison is subject to inflated Type I error rates; done carefully, the technique has limited power due to its focus on marginal comparisons. More fundamentally, the method implicitly employs an inappropriate measure of mixture difference. A more powerful approach is to compare mixtures using a likelihood ratio test. In applications where the standard asymptotic theory does not hold, the null reference distribution can be obtained through parametric bootstrapping. Using the likelihood ratio to test competing mixture models encourages modeling the change in mixture contributions as a function of covariates in addition to testing simple hypotheses. The method is demonstrated with an analysis of potential sampling bias in the estimation of population contributions to the commercial sockeye salmon (*Oncorhynchus nerka*) fishery in Upper Cook Inlet, Alaska.

Keywords: discrete mixture analysis, genetic stock identification, mixed stock analysis, mixture difference, compositional data, simultaneous inference.

1. INTRODUCTION

Mixed stock analysis (MSA) estimates the relative contributions of distinct populations in a mixture of organisms. MSA is an important tool in fisheries management and research (Begg, Friedland, and Pearce 1999; Shaklee, Beacham, Seeb, and White 1999), marine mammal research (Pella and Masuda 2001), and wildlife management and conservation (Pearce et al. 2000). MSA has also been used as an introgression index to calculate the percentage of genes from source or parental populations (Planes and Doherty 1997). While methods for MSA estimation have appeared in the fisheries literature for many years (Grant, Milner, Krasnowski, and Utter 1980; Fournier, Beacham, Ridell, and Busack 1984; Millar 1987; Pella and Milner 1987), and much longer in the statistics literature (see reviews in Redner and Walker 1984; Titterington, Smith, and Makov 1985), new applications continually require methodological extensions.

Recently fisheries researchers have begun investigating spatial or temporal homogeneity in mixtures by comparing mixture estimates from two or more independent samples. Differences between samples are assessed by looking across the samples for overlap of the confidence intervals for a given population's contribution (e.g., Wilmot, Kondzela, Guthrie, and Masuda 1998; McParland, Ferguson, and Liskauskas 1999; Shaklee et al. 1999; Ruzzante, Taggart, Lang, and Cook 2000). This approach is subject to both inflated Type I error rates due to multiple testing and inflated Type II error rates due to focusing on marginal, rather than joint, summary statistics. More fundamentally, this approach implicitly employs an inappropriate measure of mixture difference that ignores the dependence among contribution estimates due to the constraint that they sum to one (DISCUSSION).

This paper extends the maximum likelihood framework commonly employed in MSA estimation to compare competing mixture models using the likelihood ratio. Mixture homogeneity across independent samples is assessed by a likelihood ratio test of the null model, in which all samples come from a common mixture, versus the alternative model, in which each sample comes from a potentially different mixture. Asymptotic theory, Monte Carlo simulation, or parametric bootstrapping can provide approximate *P* values for the test. Adopting a likelihood ratio framework encourages researchers to begin explicitly modeling mixtures as functions of covariates in addition to testing simple hypotheses. The method has been implemented in the latest release of the freeware SPAM: Statistical Package for Analyzing Mixtures (version 3.5; Reynolds 2001, available online at http://www.cf.adfg.state.ak.us/geninfo/research/genetics/Software/SpamPage.htm).

We introduce the basic finite mixture model, derive the likelihood ratio test of M-sample mixture homogeneity, and present three approaches to approximating the null reference distribution (METHODS). The method is illustrated with an example from the sockeye salmon (*Oncorhynchus nerka*) commercial fishery in Upper Cook Inlet, Alaska (APPLICATION). Parametric bootstrapping is used to derive the null reference distribution. We compare the performance of the likelihood ratio method and the confidence interval method both in terms of the current application and in general (DISCUSSION). Marginal measures of 'mixture difference' appropriate to compositional data are briefly discussed. The finite mixture model is extended to two-stage sampling (APPENDIX).

2. METHODS

2.1 The Finite Mixture Model

A friend goes into a candy store. Two jars contain strawberry candies and licorice candies, but the jars differ in the proportions of each flavor. She randomly grabs handfuls of candy from each well-mixed jar (*the baseline populations*), combines the handfuls into a single bag (*the mixture*), pays for it, walks out of the store and hands it to you. She tells you the original proportions in each jar, then says you may have some candy if you can tell her what portion of the mixture came from each jar. This is a mixture problem. More precisely, it is a finite mixture problem as only two jars contributed to the mixture.

Identifiability of the mixture requires that the probability density functions of the characteristic (e.g., *flavor*) differ across the contributing populations (e.g., *jars*) (Redner and Walker 1984). Characteristics commonly used in fisheries include parasite assemblages (Moles, and Jensen 2000; Urawa, Nagasawa, Margolis, and Moles 1998), scale patterns (Marshall et al. 1987), morphometrics and meristics (Fournier et al. 1984), artificial tags such as thermal marks, coded wire tags, or fin clips (Ihssen et al. 1981), and, increasingly, genetic markers (Seeb and Crane 1999; Ruzzante et al. 2000). Although discrete characteristics are not essential (Millar 1987), they are assumed in the following presentation. The model holds for continuous characteristics as well.

Let n items be randomly sampled from a mixture of J populations. Let the j^{th} population contribute an unknown proportion $\theta_j >= 0$ to the mixture, $\Sigma\theta_j = 1$; $\Theta = (\theta_1, ..., \theta_J)$. If the characteristic measured on the i^{th} sample observation is denoted by x_i, then the probability of observing the sample $\mathbf{X} = \{x_1, x_2, ...,x_n\}$ is:

$$\Pr(\mathbf{X}|\ \Theta, \Phi) = \prod_{i=1}^{n} \Pr(x_i \mid \Theta, \Phi\ ) = \prod_{i=1}^{n} \left\{ \sum_{j=1}^{J} \theta_j * \Pr_j(x_i \mid \phi_j\ ) \right\} \tag{1}$$

where ϕ_j is the probability density function of the characteristic in population j, reiterated in the notation $\Pr_j(\ )$, and $\Phi = (\phi_1, ..., \phi_J)$. The model, and its extension below, assumes that all potentially contributing populations are included in the set {Pop. 1, Pop. 2, ..., Pop. J} (see Smouse, Waples, and Tworek 1990). Multivariate characteristics are easily incorporated by appropriate expansion of the $\Pr_j(x_i|\phi_j)$ terms (Millar 1987).

Estimation.

Estimating the mixture proportions, Θ, requires information regarding the (possibly multivariate) characteristic probability density function, ϕ_j, for each contributing population. This is generally available in the form of a sample from each baseline population. The mixture and baseline samples can be used with the expectation-maximization algorithm (EM, Dempster, Laird, and Rubin 1977) to solve the unconditional maximum likelihood problem (Redner and Walker 1984). In most fisheries applications, however, researchers fix the nuisance parameters, ϕ_j, at their estimates from only the baseline samples, $\hat{\phi}_j$ (Millar 1987). Maximum likelihood is then used to estimate the unknown Θ conditional on $\phi_j = \hat{\phi}_j$. This conditioning is justified by the fact that, relative to the baseline sample, there is generally little information on ϕ_j in the mixture sample (Milner, Teel, Utter, and Burley 1981).

Uncertainty in the estimates of mixture proportions, $\hat{\Theta}$, arises from sampling uncertainty in both the mixture and the population baseline samples. In practice, these sampling uncertainties are accounted for by nonparametric bootstrap resampling from the mixture sample and parametric bootstrap resampling from the baseline characteristic distributions, $\hat{\phi}_j$. The bootstrap mixture estimates are then used to construct confidence intervals for each unknown baseline population contribution, θ_j (ADF&G 2000).

The conditional maximum likelihood estimation (CMLE) method is implemented in the SPAM software package (see Debevec et al. 2000). SPAM uses the EM algorithm, a conjugate gradient search algorithm, and/or iteratively reweighted least squares to numerically solve the CMLE problem (for algorithm implementation details see Pella, Masuda, and Nelson 1996).

The CMLE method can produce biased estimates if contributing populations are missing from the baseline or, in the case of discrete characters, if the characteristic distribution estimate assigns zero probability to values that actually do occur in a baseline population but were not observed in the sample, that is, sampling zeros (Smouse, Waples, and Tworek 1990). Methods have been developed to account for missing baseline populations by applying the EM algorithm to estimate the missing ϕ_j along with Θ (Pella and Milner 1987; Smouse et al. 1990). The problem of sampling zeros also can be addressed by use of the EM algorithm (Smouse et al. 1990) or via a Bayesian analysis using shrinkage estimators (Pella and Masuda 2001).

2.2 Extension to Two Mixture Samples

The basic mixture model is easily extended to two (or more) independent samples. Let m index the M independent simple random samples from possibly different mixtures of the same baseline populations, $\Theta^{\mathbf{1}}, \Theta^{\mathbf{2}}, ..., \Theta^{\mathbf{M}}$. E.g., m could index samples taken through time or space. Following the previous notation, the general mixture model for the sequence of samples, $\{\mathbf{X^1} = \{x^1_1, x^1_2, ..., x^1_{n\_1}\}, ..., \mathbf{X^M} = \{x^M_1, x^M_2, ..., x^M_{n\_M}\}\}$, allowing each sample to come from a different mixture, $\Theta^{\mathbf{1}}, \Theta^{\mathbf{2}}, ..., \Theta^{\mathbf{M}}$, is:

$$\Pr(\{\mathbf{X^1}, \mathbf{X^2},...,\mathbf{X^M}\} \mid \Theta^1, \Theta^2, ..., \Theta^M, \Phi) = \prod_{m=1}^{M} \prod_{i=1}^{n\_m} \left\{ \sum_{j=1}^{J} \theta_j^m * \Pr_j(x_i^m \mid \phi_j) \right\}. \quad (2)$$

Note that the model assumes the characteristic distribution function for each population, ϕ_j, is constant with regard to the index m (e.g., population characteristics do not change through time). We revisit this point in DISCUSSION.

Estimation.

The general model, in which the M independent samples potentially come from M different mixtures, can be fit by estimating each mixture independently of the others using the CMLE method described above. The constrained null model, in which the M samples come from a common mixture, Θ^0, can be fit by combining the mixture samples into a single sample and again using the CMLE approach described above. Both cases follow from the likelihood under (2). Unconditional estimation is considered in DISCUSSION.

2.3 Testing Mixture Equality

Suppose one has samples from M potentially different mixtures, each mixture consisting of contributions from a known set of baseline populations. A likelihood ratio test of equality of the M mixture proportions, H_o: $\Theta^{\mathbf{m}} = \Theta^0$ for m = 1, ..., M, versus the general inequality alternative, follows directly from model (2). The ratio of the likelihood under the general model to the likelihood under the constrained null model, conditional on $\phi_j = \hat{\phi}_j$, reduces to:

$$LR = L(\{\Theta^1,\Theta^2,...,\Theta^M\} \mid \{\mathbf{X^1}, \mathbf{X^2},...,\mathbf{X^M}\}, \hat{\Phi}) \Big/ L(\{\Theta^0,\Theta^0,...,\Theta^0\} \mid \{\mathbf{X^1}, \mathbf{X^2},...,\mathbf{X^M}\}, \hat{\Phi}) =$$

$$\prod_{m=1}^{M} \prod_{i=1}^{n\_m} \left\{ \sum_{j=1}^{J} \theta_j^m * \Pr_j(x_i^m \mid \hat{\phi}_j) \right\} \Big/ \prod_{m=1}^{M} \prod_{i=1}^{n\_m} \left\{ \sum_{j=1}^{J} \theta_j^0 * \Pr_j(x_i^m \mid \hat{\phi}_j) \right\}. \quad (3)$$

Null Reference Distribution Method 1: Asymptotic Theory.

The null hypothesis of mixture equality can be tested by comparing –2 * ln(LR) to its asymptotic distribution under the null model, a χ^2 with degrees of freedom d = (J-1)*(m-1) (Stuart, Ord, and Arnold 1999). However, the approximations underlying this asymptotic result break down when any of the mixture parameters take values near the boundary of the parameter space (Stuart et al. 1999), that is, when one or more populations contribute little or nothing to the mixture. As this is quite often the case in genetic stock identification problems (Millar 1987), the asymptotic results are frequently unreliable. Although the appropriate family of asymptotic distributions is known for tests on the boundary of the parameter space (Self and Liang 1987), it is not simple to employ this theoretical result.

Null Reference Distribution Method 2: Monte Carlo Simulation (Θ^0 Known).

The null distribution can be approximated by Monte Carlo simulation if the specific value of Θ^0 is known *a priori* (Davison and Hinkley 1997). For r = 1, ..., R, iterations:

1. Simulate N observations from model (1) using the known null mixture proportions Θ^0 and the baseline population characteristic densities $\hat{\phi}_j$. Here N = Σn_m, where n_m is the number of observations in mixture sample m, m = 1, ..., M ;

2. Fit the single mixture model (1) to the N simulated observations, giving an estimate $\hat{\Theta}^{\mathbf{0,*r}}$;

3. Randomly assign the N simulated observations to M simulated mixture samples of size {n_1, n_2, …, n_M};

4. Fit the general M-mixture model (2) to the simulated observations by estimating the M different sets of mixture proportions, $\hat{\Theta}^{\mathbf{m,*r}}$, m = 1,..., M;

5. Using the M simulated mixture samples and the estimates $\hat{\boldsymbol{\Theta}}^{\mathbf{0,*r}}$ and $\{\hat{\boldsymbol{\Theta}}^{\mathbf{m,*r}}\}$, calculate and record the likelihood ratio (3), LR^{*r}.

This process gives a sample of size R, $\{LR^{*r}: r = 1, ..., R\}$, from the unknown null reference distribution. Calculate the observed likelihood ratio, LR^{obs}, by fitting the general and restricted models as described in the previous section and plugging the estimates into (3). An approximate *P* value for the test is then given by $\left(1+\sum_r I\left(LR^{*r} \geq LR^{obs}\right)\right)/(1+R)$ (Davison and Hinkley 1997), where the indicator function I( ) has value one when the argument is true and zero otherwise. Generally, R in the range 1000 – 10000 will guarantee very little loss of power due to finite simulation (Davison and Hinkley 1997, sec. 4.2.5).

Null Reference Distribution Method 3: Parametric Bootstrapping (Θ^0 Unknown).

The appropriate value of Θ^0 will generally not be known prior to analysis of the data. In this case, we first estimate Θ^0 then perform parametric bootstrapping (Davison and Hinkley 1997) to approximate the null reference distribution:

1. Estimate Θ^0 from the M observed mixture samples by combining samples and fitting model (1);

2. Follow steps 1 – 5 outlined above, simulating from the estimated null mixture $\hat{\Theta}^0$.

Uncertainty in the conditional values of the nuisance parameters, $\hat{\phi}_j$, can be incorporated into either simulation approach by parametric bootstrap resampling from each $\hat{\phi}_j$ before constructing the null mixture during each of the R simulation rounds.

If a significant difference is detected, one could continue the model selection process by fitting less-constrained null models. For example, models in which subsets of the M samples come from identical mixtures or the M samples differ in the contributions from only a subset of the baseline populations. The software package SPAM (Reynolds 2001) currently allows the former investigation, but not the latter.

3. APPLICATION: COMPARING SALMON HARVEST MIXTURES

The sockeye salmon fishery in Upper Cook Inlet, Alaska (Figure 1) is important to the local economy. Over the last ten years, the total annual value of commercial harvests in the region ranged from US$8.8 to $111.1 million, with sockeye salmon comprising 80% - 97% of the annual value (Ruesch and Fox 1999). The fishing fleet is very efficient; the approximately 600 drift gillnet vessels can harvest as much as 70% of the available fish in a single 12-hr opening (Seeb et al. 2000).

Most sockeye salmon home with precision, returning from the ocean to their natal habitats to spawn and then die (Burgner 1991). Among the Pacific salmonids, the sockeye salmon life cycle generally places the greatest emphasis on early life use of a lake. Although the adults may spawn in many diverse environments (i.e. rivers, sloughs, lake shores), survival of their offspring generally depends on the offspring finding a rearing lake shortly after emergence, though there are types that emigrate directly to estuaries or oceans (Burgner 1991). Consequently, low rates of straying (spawning in a location other than the natal habitat) and the demands of different spawning environments can lead, over time, to significant genetic, morphometric and behavioral differences within a relatively small geographic area (e.g. Woody, Olsen, Reynolds, and Bentzen 2000).

To maintain genetic diversity and future productivity in the face of more immediate demands for economic returns by highly efficient fishers, fishery managers must accurately identify the harvest contributions of the major Upper Cook Inlet sockeye salmon stocks. Sustainable management will be very difficult or unachievable otherwise. Seeb et al. (2000) identified 44 genetically distinct populations, or stocks, within the major sockeye salmon-producing areas in Upper Cook Inlet. Overharvesting any of these stocks will affect the genetic diversity within the region; loss of a population means loss of unique combinations of genetic characters. This will also affect the economic value of the fishery as lost stocks are generally not replaceable and will no longer produce salmon for future harvests.

Mixed stock analysis has been conducted previously on Upper Cook Inlet sockeye using a number of different characteristics: scale patterns (Marshall et al. 1987), parasites (Waltemeyer, Tarbox, and Brannian 1993), and genetic markers (Grant et al. 1980; Seeb et al. 2000). Of these, genetic stock identification is best able to identify mixture proportions with the accuracy and precision required by managers (Seeb et al. 2000).

In their study, Seeb et al. (2000) collected samples of spawning salmon from each of the 44 baseline populations (Figure 1, Table 1). A target sample size of 100 individuals was selected to give acceptably precise allele frequency estimates (Allendorf and Phelps 1981; Waples 1990). Allozyme electrophoresis provided each individual's genotype at 27 discriminating unlinked loci (see Seeb et al. 2000). For management purposes, the contributions from these baseline populations are aggregated into six regions determined by geography and genetic diversity (West Cook Inlet, Susitna/Yentna, Knik, Northeast Cook Inlet, Kenai, and Kasilof). Most sockeye salmon come from four of these regions, all of which contain major river drainages (Figure 1): the Kenai River drainage, the Kasilof River drainage, the Susitna River drainage (Susitna/Yentna Region), and the Crescent River drainage (West Cook Inlet Region) (Tobias and Tarbox 1999).

3.1 The Problem

The mixture of interest is the sockeye harvest in the Central District fishery during a 12-hour opening (Figure 1). Each boat delivers its catch to one of eleven processors. Traditionally, the harvest was sampled only at the largest processor, Wards Cove (Seeb et al. 2000). To ascertain whether this procedure produces biased mixture estimates, replicate samples from a second processor, Salamatof Seafoods, Inc., were collected on four openings during the 1997 and 1998 seasons (Table 2). The equality of the mixture estimates from the two processors was tested.

The mixture sample at each processor was obtained by two-stage sampling: boats were randomly sampled from the incoming sequence of deliveries, and a random sample of sockeye salmon were selected from each boat's catch. Forty boats were sampled at Wards Cove at a rate of 10 fish per boat, for a target sample size of 400 fish. Salamatof Seafoods Inc., the smaller processor, serves a fleet of 20 – 30 boats. In 1997, the goal was to sample 400 fish per period, so between 10 and 15 fish were sampled per boat depending on the number of boats returning. In 1998, the goal was revised to 10 fish per boat for a total of 200 fish per period. Model (2) is extended to handle the two-stage sampling design in the APPENDIX. The resulting likelihood ratio is identical to (3), so the details of simulating the null reference distribution remain as given above.

Parametric bootstrapping was used to test the null hypothesis that the two processor samples came from the same mixture (R = 5000 simulations). All mixture simulations and model fitting were done in SPAM using CMLE; final analysis of the simulation results was conducted in S-Plus 2000 (Insightful, Inc., Seattle, WA). Before generating each null mixture simulation, the allele frequency estimates for each baseline population, ϕ_j, were parametrically bootstrapped to incorporate uncertainty in their values into the null reference distribution.

3.2 Results

For three of the four openings, the likelihood ratio test revealed no evidence against the null hypothesis that processors sampled a common harvest mixture (Table 2). Ninety percent bootstrap confidence intervals were calculated for each opening, both for comparison with other published mixture comparison methods (Figure 2) and for *a posteriori* insight when mixtures estimates were found to differ (see DISCUSSION). Note that the question of mixture equality and the associated likelihood ratio test focus on baseline populations, not management regions. However, results are generally presented and published as regional estimates. Therefore bootstrap confidence intervals of the total contribution from each of the six management regions were calculated for each processor-specific estimate.

Intervals were calculated using Efron's percentile method (B=1000 resamples) (Davison and Hinkley 1997). Two sets of bootstrap confidence intervals were calculated. (i) For the processor-specific estimates of the total contribution from each region (Table 2, Figure 2); published assessments of mixture equality generally focus on whether the confidence intervals for each region overlap across mixtures (e.g., processors). (ii) For the difference in processor-specific estimates of the total contribution from each region; that is θ_i^A - θ_i^B (Table 2); this is a 'natural' extension of (i). Neither confidence interval approach is fully recommended due to the lack of power and inappropriate handling of the dependence among region estimates (see DISCUSSION).

In the first opening the boat from which each fish was sampled was not recorded, making it impossible to replicate the two-stage sampling in the bootstrap confidence interval calculations. All interval estimates (Table 2, Figure 2) therefore assume simple random sampling and hence may underestimate the true variance. Confidence intervals incorporated parametric resampling of the allele frequencies from each baseline population and nonparametric resampling of each mixture sample, following ADF&G (2000).

4. DISCUSSION

4.1 Upper Cook Inlet Sockeye Salmon

Processor-specific mixture differences may arise from a combination of spatial heterogeneity in the harvestable Central District sockeye salmon mixture and clustering during harvest among boats that deliver to a specific processor. If such clustering regularly occurs, then the current harvest-sampling plan may need to be revised. One possibility would be to sample every processor and develop a weighted average, across processors, of the mixture estimates, with weights proportional to each processor's portion of the total harvest.

4.2 Method Comparison

The mixture equality problem is often assessed by checking, for each contributing region, the overlap among confidence intervals from the different mixture samples (Seeb et al. 1999; Wilmot et al. 1998, McParland et al. 1999, Shaklee et al. 1999, Ruzzante et al. 2000) (e.g., Figure 2). This is a very poor approach, fraught with statistical deficiencies both obvious and subtle. It suffers from both (i) inflated Type I error rates arising from the simultaneous inferences, and (ii) inflated Type II error rates arising from the use of marginal (region-specific) measures of mixture difference.

One can control the Type I error rate when comparing M-independent $(1-\alpha)*100\%$ confidence intervals for overlap by enlarging each interval's level to $(1-\alpha)^{(1/M)}*100\%$, producing a simultaneous confidence level of $(1-\alpha)*100\%$ across the set of M intervals (Hsu 1996). However, the inflation arising from repeating this 'overlap check' across J-1 sets of intervals remains.

More importantly, these marginal comparisons are much less powerful than a single omnibus test of the difference between mixture compositions. The overlap method fails to suggest any marginal difference between processor estimates (Figure 2, Table 2), while the likelihood ratio test reveals a very significant difference on the 14 July 1997 sampling event (Figure 3, Table 2).

The general loss of power inherent in marginal comparisons is magnified in the context of mixtures because of the dependence among mixture contributions; mixtures are constrained to lie on the simplex, $\theta_I >= 0$, $\Sigma\theta_i = 1$. A change in one region contribution necessitates a change in at least one other region contribution. The overlap method and its extension - looking at the marginal difference in region contributions $\Theta^A - \Theta^B = (\theta^A{}_1 - \theta^B{}_1, \ldots, \theta^A{}_J - \theta^B{}_J)$, ignore this dependence (Table 2). For example, the mixture difference on the 14 July 1997 sampling event is driven by simultaneous shifts in the contributions from West Cook Inlet, Susitna / Yentna, and Kenai regions (Table 2, Figure 3); the unadjusted marginal confidence intervals ($\theta^{WC}{}_{S/Y} - \theta^{Sal}{}_{S/Y}$, Table 2) only detects the shift in the Sustina / Yenta contribution.

There is ongoing research in the development of an appropriate measure of mixture difference, one that captures this dependence among region contributions (e.g., Aitchison 1982, 1986, 1992; Billheimer, Guttorp and Fagan 2001). The (inverse) addition operator for compositions (Aitchison 1986, 1992) has been used to develop a measure of distance between compositions (Billheimer et al. 2001). In conjunction with the logistic normal distribution (Aitchison 1982, 1986), this provides an alternative means of testing mixture equality. Unfortunately, the operator and distance measure both assume non-zero contributions, limiting practical implementation. Also, as the originators acknowledge, the operator and measure are difficult to interpret (Billheimer et al. 2001). Even visual display of composition data presents methodological and implementation difficulties (e.g., Figure 3 and the visual compression of distances near the boundaries) (Billheimer et al. 2001).

A more subtle criticism of the confidence interval overlap method is that it often is used to examine mixture equality not at the scale of the baseline population contributions but at the scale of regional aggregates of populations. Comparison of regional aggregates may obscure differences at the level of the baseline populations. For example, two populations in the same region may tradeoff in their contributions to two mixtures, producing an apparently constant regional contribution to each mixture but by means of differing population contributions.

Researchers must use caution investigating mixture differences and interpreting contribution confidence intervals. The likelihood ratio approach controls both Type I and Type II error rates and provides a test of mixture difference that recognizes the constraints of mixture (that is, composition) data. Furthermore, if one’s level of interest is the regional aggregates or any smooth function of the baseline population contributions, the likelihood ratio test remains applicable as it is invariant to transformation of the parameters (Stuart et al. 1999). Such cases may require more care in fitting the null model. Most importantly, the likelihood ratio method provides a paradigm for model development and selection. This encourages researchers to begin modeling mixture variation across time or space rather than simply testing hypotheses of equality.

4.3 Conditioning and Model Extensions

Small baseline population samples, relative to the mixture samples, may warrant unconditional maximum likelihood estimation (Smouse et al. 1990). In this case the likelihood ratio test remains applicable but the details of simulating the null reference distribution change. Fitting the general model cannot be broken down into M separate estimation problems as each mixture sample potentially contains information regarding each ϕ_j. The EM-fitting algorithm of Smouse et al. (1990) can be extended to handle both this general M-mixture model and the constrained M-mixture model. However, with many baseline populations unconditional fitting can encounter numerical problems overcoming local optima in the likelihood surface (Jerry Pella, personal communication, 12 October 2000).

The M-mixture model can also be extended to allow the characteristic density for each baseline population, ϕ_j, to potentially change with the mixture index m. Whereas this requires more baseline samples and estimation of many more parameters, the likelihood ratio test of equality remains applicable.

5. CONCLUSIONS

Mixed stock analysis, especially using genetic markers, is an increasingly important tool in fisheries and wildlife management. Advances in genetics continue to simplify the collection and analysis of field samples, allowing managers and researchers to develop extensive baseline population databases as well as sample mixtures through space and time. Unfortunately, the methods commonly employed to compare mixtures through space and time are fraught with statistical deficiencies. The likelihood ratio test presented here provides a statistically sound method for comparing these mixture samples. Currently employed confidence interval methods may give some insight into the structure detected by the test, but researchers must use caution in interpreting the results as the implicit measures of marginal difference are inappropriate and the methods suffer from very low power. More appropriate confidence interval methods await development of more appropriate, and readily interpretable, measures of mixture difference.

The likelihood ratio approach can be used to develop more refined models of mixture variation, providing greater insight into wild populations subject to research and management. Such efforts can provide insight into the adequacy of mixture sampling protocols (illustrated here), investigation of marine migration patterns (Seeb and Crane 1999), and temporal and spatial stability of scientifically or economically important mixtures (Ruzzante et al. 2000).

ACKNOWLEDGMENTS

We wish to thank the following people for helpful discussions on the sampling protocol question and/or their helpful reviews of earlier drafts – Eric Anderson, Ed Debevec, David Evans, Rich Hinrichsen, and two anonymous reviewers. A comment by one of the reviewers led, in a roundabout way, to a greater awareness of the fundamental problem of measuring differences in mixtures; thanks are due Dean Billheimer for conversations expanding on that theme.

The genetics data for the population baselines and the Wards Cove mixture are available in Seeb et al. (2000). The genetics data from the Salamatof Seafoods, Inc. mixtures are available by contacting Dr. Lisa Seeb at the Gene Conservation Laboratory: lisa_seeb@fishgame.state.ak.us.

Contribution PP-204 of the Alaska Department of Fish and Game, Commercial Fisheries Division, Juneau, Alaska, USA.

APPENDIX: TWO-STAGE SAMPLING M-MIXTURE MODEL

In the sockeye harvest application, mixture samples were obtained by a two-stage sampling scheme. Model (2) is easily extended to this situation. Let k index the sequence of K^m primary sampling units randomly selected from the m^th^ of M independent mixtures. Let i_k index the sequence of $n^m{}_k$ secondary sampling units randomly selected from the k^th^ primary unit from the m^th^ mixture. The possibly multivariate characteristic observed on the secondary sampling unit i_k in the m^th^ mixture is denoted $x^m{}_{i\_k}$. Following the text, $\theta^m{}_j$ is the unknown proportion of the m^th^ mixture contributed by population j (out of J contributing populations), $\sum_j \theta_j^m = 1$ for each m, and ϕ_j is the probability density of characteristics in population j. The resulting likelihood ratio for testing H_o: $\Theta^{\mathbf{m}} = \Theta^0$ for m = 1, ..., M, versus the general inequality alternative, is:

$$\prod_{m=1}^{M}\left\{\prod_{k=1}^{K^m}\prod_{i\_k=1}^{n_k^m}\left\{\sum_{j=1}^{J}\theta_j^m * \Pr_j(x_{i\_k}^m \mid \phi_j)\right\}\right\} \Bigg/ \prod_{m=1}^{M}\left\{\prod_{k=1}^{K^m}\prod_{i\_k=1}^{n_k^m}\left\{\sum_{j=1}^{J}\theta_j^0 * \Pr_j(x_{i\_k}^m \mid \phi_j)\right\}\right\} =$$

$$\prod_{m=1}^{M}\left\{\prod_{i=1}^{\Sigma n_k^m}\left\{\sum_{j=1}^{J}\theta_j^m * \Pr j(x_i^m \mid \phi j)\right\}\right\} \Bigg/ \prod_{m=1}^{M}\left\{\prod_{i=1}^{\Sigma n_k^m}\left\{\sum_{j=1}^{J}\theta_j^0 * \Pr j(x_i^m \mid \phi j)\right\}\right\}. \quad \text{(A.1)}$$

Because each mixture is assumed homogeneous across its associated primary sampling units, the likelihood ratio under two-stage sampling, (A.1), reduces to that for simple random sampling (3).

REFERENCES

ADF&G (2000), "SPAM version 3.2 User's Guide: Statistics Program for Analyzing Mixtures," Special Publication No. 15, Alaska Department of Fish and Game, Division of Commercial Fisheries, Anchorage, AK, http://www.cf.adfg.state.ak.us/geninfo/research/genetics/Software/SpamPage.html.

Aitchison, J. (1982), "The statistical analysis of compositional data (with discussion)," *Journal of the Royal Statistical Society, Ser. B,* 44, 139 - 177.

Aitchison, J. (1986), *The Statistical Analysis of Compositional Data*, New York: Chapman and Hall.

Aitchison, J. (1992), "On criteria for measures of compositional difference," *Mathematical Geology*, 24, 365 – 379.

Allendorf, F. W., and Phelps, S. R. (1981), "Use of allelic frequencies to describe population structure," *Canadian Journal of Fisheries and Aquatic Sciences*, 38, 1507-1514.

Begg, G. A., Friedland, K. D., and Pearce J. B. (1999), "Stock identification – its role in stock assessment and fisheries management: a selection of papers presented at a symposium of the 128th annual meeting of the American Fisheries Society in Hartford, Connecticut, USA, 23-27 August 1998," *Fisheries Research*, 43, 1-3.

Billheimer, D., Guttorp, P., and Fagan W. F. 2001. "Statistical interpretation of species composition," *Journal of the American Statistical Association*, 96, 1205 – 1214.

Burgner, R. L. (1991), "Life history of sockeye salmon (*Oncorhynchus nerka*)," in *Pacific Salmon Life Histories*, eds. C. Groot and L. Margolis, Vancouver, BC: University of British Columbia Press, pp 3-117.

Davison, A. C., and Hinkley, D. V. (1997), *Bootstrap Methods and their Application*, Cambridge, UK: Cambridge University Press.

Debevec, E. M., Gates, R. B., Masuda, M., Pella, J., Reynolds, J. H., and Seeb, L. W. (2000), "SPAM (version 3.2): Statistics Program for Analyzing Mixtures," *Journal of Heredity*, 91, 509 – 511.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Ser. B*, 39, 1-38.

Fournier, D. A., Beacham, T. D., Ridell, B. E., and Busack, C. A. (1984), "Estimating stock composition in mixed stock fisheries using morphometric, meristic, and electrophoretic characteristics," *Canadian Journal of Fisheries and Aquatic Sciences*, 41, 400-408.

Grant, W. S., Milner, G. B., Krasnowski, P., and Utter, F. M. (1980), "Use of biochemical genetic variants for identification of sockeye salmon (*Onchorhynchus nerka*) stocks in Cook Inlet, Alaska," *Canadian Journal of Fisheries and Aquatic Sciences*, 37, 1236-1247.

Hsu, J. C (1996), *Multiple Comparisons: Theory and Methods*, London: Chapman and Hall.

Ihssen, P. E., Booke, H. E., Casselman, J. M., McGlade, J. M., Payne, N. R., and Utter, F. M. (1981), "Stock identification: materials and methods," *Canadian Journal of Fisheries and Aquatic Sciences*, 38, 1838-1855.

Marshall, S., Bernard, D., Conrad, R., Cross, B., McBride, D., McGregor, A., McPherson, S., Oliver, G., Sharr, S., and Van Allen, B. (1987), "Application of scale patterns analysis to the management of Alaska's sockeye salmon (*Onchorhynchus nerka*) fisheries," in *Sockeye Salmon (Oncorhynchus nerka) Population Biology and Future Management*, eds. H. D. Smith, L. Margolis, and C. C. Wood, Canadian Special Publication on Fisheries and Aquatic Science 96, pp. 207-326.

McParland, T. L., Ferguson, M. M., and Liskauskas, A. P. (1999), "Genetic population structure and mixed-stock analysis of walleyes in the Lake Erie-Lake Huron corridor using allozyme and mitochondrial DNA markers," *Transactions of the American Fisheries Society*, 128, 1055-1067.

Moles, A., and Jensen, K. (2000), "Prevalence of the sockeye salmon brain parasite Myxobolus arcticus in selected Alaska streams," *Alaska Fisheries Research Bulletin*, 6, 85-93.

Millar, R. B (1987), "Maximum likelihood estimation of mixed stock fishery composition," *Canadian Journal of Fisheries and Aquatic Sciences*, 44, 583-590.

Milner, G. B., Teel, D. J., Utter, F. M., and Burley, C. L. (1981), "Columbia River stock identification study: Validation of genetic method," Unpublished manuscript (Final report of research (FY80) financed by Bonneville Power Administration Contract DE-A179-80BP18488), National Marine Fisheries Service, Northwest and Alaska Fisheries Center, Seattle, WA.

Pearce, J. M., Pierson, B. J., Talbot, S. L., Derksen, D. V., Kraege, D., and Scribner, K. T. (2000), "A genetic evaluation of morphology used to identify harvested Canada geese," *Journal of Wildlife Management*, 64, 863-874.

Pella, J. J., and Masuda, M. (2001), "Bayesian methods for stock-mixture analysis from genetic characters," *Fishery Bulletin*, 99, 151-167.

Pella, J. J., Masuda, M., and Nelson, S. (1996), "Search algorithms for computing stock composition of a mixture from traits of individuals by maximum likelihood," NOAA/NMFS Technical Memo NMFS-AFSC-61, National Marine Fisheries Service, Northwest and Alaska Fisheries Center, Seattle, WA.

Pella, J. J., and Milner, G. B. (1987), "Use of genetic marks in stock composition analysis," in *Population Genetics and Fishery Management*, eds. N. Ryman and F. Utter, Seattle, WA: Washington Sea Grant Program, pp. 247-276.

Planes S. and Doherty P. J. (1997), "Genetic and color interactions at a contact zone of *Acanthochromis polyacanthus:* a marine fish lacking pelagic larvae," *Evolution*, 51, 1232-1243.

Redner, R. A., and Walker, H. F. (1984), "Mixture densities, maximum likelihood and the EM algorithm," *Society for Industrial and Applied Mathematics Review*, 26, 195-239.

Reynolds, J. H. (2001), "SPAM (Statistics Program for Analyzing Mixtures) Version 3.5: User's Guide Addendum" Addendum to Special Publication No. 15, Alaska Dept. of Fish and Game, Division of Commercial Fisheries, Gene Conservation Laboratory, Anchorage, AK, http://www.cf.adfg.state.ak.us/geninfo/research/genetics/Software/SpamPage.htm

Ruesch, P. H., and Fox, J. (1999), "Upper Cook Inlet commercial fisheries annual management report, 1998," Regional Information Report No. 2A99-21, Alaska Department of Fish and Game, Division of Commercial Fisheries, Anchorage, AK.

Ruzzante, D. E., Taggart, C. T., Lang, S., and Cook, D. (2000), "Mixed-stock analysis of Atlantic cod near the Gulf of St. Lawrence based on microsatellite DNA," *Ecological Applications*, 10, 1090-1109.

Seeb, L. W., and Crane, P. A. (1999), "Allozymes and mitochondrial DNA discriminate Asian and North American populations of chum salmon in mixed-stock fisheries along the south coast of the Alaska Peninsula," *Transactions of the American Fisheries Society*, 128, 88-103.

Seeb, L. W., Habicht, C., Templin, W. D., Tarbox, K. E., Davis, R. Z., Brannian, L. K., Seeb, J. E. (2000), "Genetic diversity of sockeye salmon of Cook Inlet, Alaska, and its application to management of populations affected by the *Exxon Valdez* Oil Spill," *Transactions of the American Fisheries Society*, 129, 1223-1249.

Self, S. G., and Liang, K. Y. (1987), "Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions," *Journal of the American Statistical Association*, 82, 605-610.

Shaklee, J. B., Beacham, T. D., Seeb, L., and White, B. A. (1999), "Managing fisheries using genetic data: case studies from four species of Pacific salmon," *Fisheries Research*, 43, 45-78.

Smouse, P. E., Waples, R. S., and Tworek, J. A. (1990), "A genetic mixture analysis for use with incomplete source population data," *Canadian Journal of Fisheries and Aquatic Science*s, 47, 620-634.

Stuart, A., Ord, J. K., and Arnold, S. (1999), *Kendall's Advanced Theory of Statistics Vol 2A: Classical Inference and the Linear Model* (6th ed.), New York: Oxford University Press.

Titterington, D. M., Smith, A. F. M., and Makov, U. E. (1985), *Statistical Analysis of Finite Mixture Distributions*, New York: Wiley and Sons.

Tobias, T., and Tarbox, K. E. (1999), "An estimate of total return of sockeye salmon to Upper Cook Inlet, Alaska 1976-1998," Regional Information Report 2A99-11, Alaska Department of Fish and Game, Division of Commercial Fisheries, Anchorage, AK.

Urawa, S., Nagasawa, K., Margolis, L., and Moles, A. (1998), "Stock identification of chinook salmon (*Onchorhynchus tshawytscha*) in the North Pacific Ocean and Bering Sea by parasite tags," *North Pacific Anadromous Fish Commission Bulletin*, 1, 199-204.

Waltemeyer, D. L., Tarbox, K. E., and Brannian, L. K. (1993), "Presence of the parasite *Philonema oncorhynchic* in sockeye salmon return to Upper Cook Inlet, Alaska in 1991," Regional Information Report No. 2A93-24, Alaska Department of Fish and Game, Division of Commercial Fisheries, Anchorage, AK.

Waples, R. S. (1990), "Temporal changes of allele frequency in Pacific salmon – Implications for mixed-stock fishery analysis," *Canadian Journal of Fisheries and Aquatic Sciences*, 47, 968-976.

Wilmot, R. L., Kondzela, C. M., Guthrie, C. M., and Masuda, M. M. (1998), "Genetic stock identification of chum salmon harvested incidentally in the 1994 and 1995 Bering Sea trawl fishery," *North Pacific Anadromous Fish Commission Bulletin*, 1, 285-299.

Woody, C. A., Olsen, J., Reynolds, J., and Bentzen, P. (2000), "Temporal variation in phenotypic and genotypic traits in two sockeye salmon populations, Tustumena Lake, Alaska," *Transactions of the American Fisheries Society*, 129, 1031-1043.

Table 1. Baseline populations and associated reporting regions for the mixture analysis of commercially harvested sockeye salmon in Upper Cook Inlet, Alaska from Seeb et al. (2000). Numbers refer to labels in Figure 1. Abbreviations: Ck. – Creek, Lk. – Lake, R. – River.

Region	Population
West Cook Inlet	1 - Chilligan R., 2 – Crescent Lk., 3 – Wolverine Ck., 4 – McArthur R., 5 – Packers Lk., 6 – Coal Ck.
Susitna / Yenta	7 – Yentna R., 8 – Shell Lk., 9 – Hewitt / Whiskey Lks., 10 – Trinity / Movie Lks., 11 – Judd Lk., 12 – Chelatna Lk., 13 – Byers Lk., 14 – Susitna R., 15 – Mama & Papa Bear Lks., 16 – Larson Lk., 17 – Talkeetna R., 18 – Stephan Lk., 19 – Birch Ck., 20 – Red Shirt Lk.
Knik	21 – Nancy Lk., 22 – Cottonwood Lk., 23 – Fish Ck., 24 – Jim Ck., 25 – Sixmile Ck.
Northeast Cook Inlet	26 – Daniels Lk., 27 – Bishop Ck., 28 – Swanson R.
Kenai	29 – Skilak Lk. Outlet, 30 – Hidden Ck., 31 – Between Kenai and Skilak Lk., 32 – Upper Russian R., 33 – Tern Lk., 34 – Quartz Ck., 35 – Moose Ck., 36 – Johnson Ck., 37 – Railroad Ck., 38 – Ptarmigan Ck.
Kasilof	39 – Nikolai Ck., 40 – Tustumena Lk., 41 – Bear Ck., 42 – Moose Ck., 43 – Glacier Ck., 44 – Seepage Ck.

Table 2. Opening dates, collection site, sample sizes (N), and conditional maximum likelihood mixture estimates for sockeye salmon sampled from the commercial harvest in Upper Cook Inlet, Alaska. Contributions from each of the 44 baseline populations were estimated and then summed to the six management regions for display (see Figure 1 and Table 1). Ninety percent bootstrap confidence intervals (1000 replicates, Efron's percentile method, Davison and Hinkley 1997) are given for both the processor-specific region contribution estimates (Figure 2) and the marginal difference in processor-specific region contribution estimates. These intervals are commonly used to compare mixture equality: do the processor-specific intervals overlap? Do the marginal difference intervals contain zero? Both approaches suffer from poor power to detect mixture differences as they ignore the inherent dependence among region contributions. Approximate *P* values were calculated from the parametric bootstrap likelihood ratio test of Ho: Identical mixtures, and Ha: Mixtures differ with processor (R = 5000 resamples). Note that the *P* value is testing for equality of baseline population contributions, not region contributions. Processors: WC – Wards Cove, Sal. – Salamatof Seafoods, Inc.

Opening	Processor	N	West Cook Inlet	Susitna/ Yentna	Knik	Northeast Cook Inlet	Kenai	Kasilof	*P* value
14 July 1997	WC	394	0.00 (0, 0.03)	0.16 (0.07, 0.22)	0.02 (0, 0.06)	0.00 (0, 0.01)	0.79 (0.70, 0.87)	0.03 (0, 0.91)	0.001
	Sal.	391	0.06 (0, 0.12)	0.05 (0, 0.12)	0.02 (0, 0.07)	0.00 (0, 0.01)	0.84 (0.75, 0.91)	0.03 (0, 0.08)	
	90% CI WC - Sal		(-0.12, 0.03)	(0.00, 0.18)	(-0.05, 0.05)	(-0.01, 0.01)	(-0.16, 0.07)	(-0.06, 0.07)	
21 July 1997	WC	398	0.02 (0, 0.06)	0.07 (0.01, 0.12)	0 (0, 0.03)	0 (0, 0.02)	0.90 (0.80, 0.96)	0.02 (0, 0.08)	0.171
	Sal.	394	0.00 (0, 0.04)	0.05 (0.01, 0.13)	0.04 (0.01, 0.08)	0.01 (0, 0.02)	0.85 (0.76, 0.92)	0.05 (0, 0.08)	
	90% CI WC - Sal		(-0.03, 0.05)	(-0.08, 0.08)	(-0.08, 0.00)	(-0.02, 0.01)	(-0.06, 0.15)	(-0.07, 0.05)	

10 July 1998	WC	394	0.06 (0.04, 0.15)	0.31 (0.21, 0.40)	0.08 (0.03, 0.14)	0.01 (0, 0.03)	0.38 (0.28, 0.45)	0.16 (0.08, 0.23)	0.230
	Sal.	159	0.11 (0.01, 0.25)	0.25 (0.11, 0.38)	0.01 (0, 0.13)	0.01 (0, 0.05)	0.51 (0.35, 0.64)	0.11 (0, 0.23)	
	90% CI WC - Sal		(-0.18, 0.09)	(-0.10, 0.24)	(-0.06, 0.14)	(-0.04, 0.02)	(-0.30, 0.05)	(-0.11, 0.17)	
17 July 1998	WC	398	0.05 (0, 0.16)	0.37 (0.22, 0.43)	0.04 (0.01, 0.12)	0.00 (0, 0.01)	0.53 (0.42, 0.63)	0.01 (0, 0.06)	0.689
	Sal.	197	0.01 (0, 0.11)	0.27 (0.16, 0.39)	0.09 (0.03, 0.20)	0.00 (0, 0.02)	0.56 (0.42, 0.66)	0.05 (0, 0.11)	
	90% CI WC - Sal		(-0.07, 0.15)	(-0.11, 0.22)	(-0.16,0.05)	(-0.02, 0.01)	(-0.17, 0.17)	(-0.10, 0.03)	

FIGURE LEGENDS

Figure 1. Upper Cook Inlet, Alaska. Numbers refer to locations of baseline populations listed in Table 1 (from Seeb et al. 2000). Commercial harvests occurred in the Central District.

Figure 2. A common method of investigating mixture equality using ninety percent confidence intervals, demonstrated with sockeye salmon harvests sampled at two different processors in Upper Cook Inlet, Alaska. This overlap method suggests no processor differences at any of the four sampling events, though there is a significant difference on 14 July 1997 (Table 2, Figure 3). Processor-specific intervals are labeled for the West Cook Inlet region in each panel to show ordering (WC – Wards Cove, top interval; Sal – Salamatof Seafoods, Inc, bottom interval).

Figure 3. Sub-mixture projections of the processor-specific nonparametric bootstrap mixture estimates for the 14 July 1997 collections. The process-specific mixture estimates only differed at three of six regions (Table 2), so results are displayed for the four-component mixture (West Cook Inlet = WCI, Susitna / Yentna = S/Y, Kenai = K, All Others = O). Four-component mixture data inhabits a triangular pyramid; we display the four three-component projections of this data space. The plots were created as follows. Consider placing a bright light at the WCI vertex of the data space pyramid and marking the shadows cast on the far wall by the data points – these shadows are the projection of the (WCI, S/Y, K, O) data points to the (S/Y, K, O) sub-mixture; the projection is obtained by dropping the WCI contribution and renormalizing the remaining contributions. Repeat at each vertex, then slice the pyramid along the sides and folding down the walls to give the two-dimensional display shown: Wards Cove resamples (left); Salamatof Seafoods, Inc. resamples (right). Each triangle, or ternary diagram, should be read as follows: the closer a point is to a vertex, the greater the contribution of that component to

the mixture. That is, points on a vertex are mixtures consisting of 100% of that component; points along a side are mixtures consisting of two components in proportions equal to the relative distance from the opposing vertex (closer to S/Y, then more S/Y contribution); points in the interior are mixtures of all three components. Ternary diagrams greatly compress distances between mixtures that fall near the boundaries (Billheimer et al. 2001) and so tend to visually underplay substantial mixture differences. The Wards Cove sample mixture differs significantly from the Salamatof Seafoods, Inc. sample mixture (Table 2), having less Kenai and West Cook Inlet contributions and more Susitna / Yenta contributions (left vrs right figures).

Figure 1

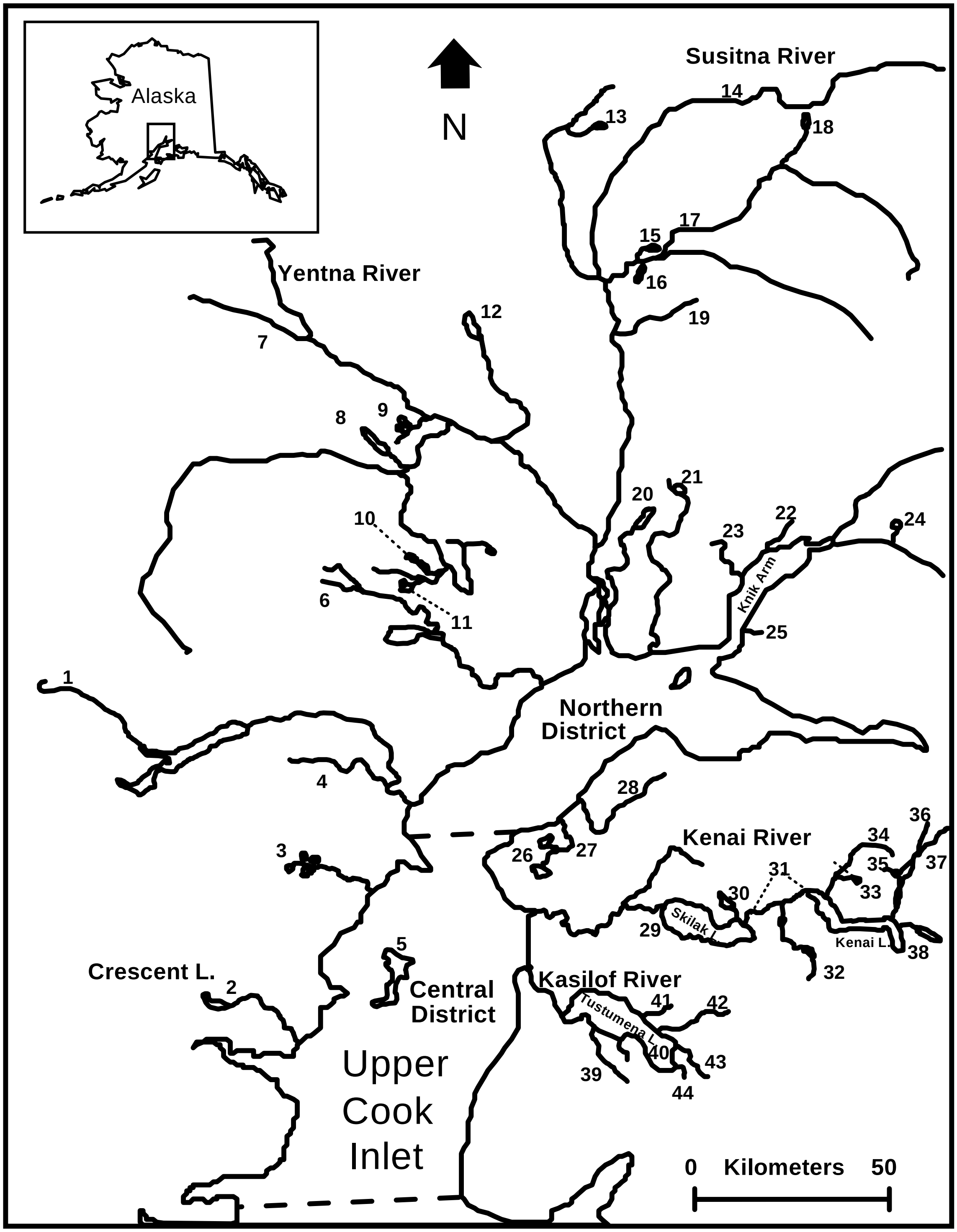

Alaska
N
Susitna River
Yentna River
Northern District
Kenai River
Skilak L.
Kenai L.
Knik Arm
Crescent L.
Central District
Kasilof River
Tustumena L.
Upper Cook Inlet
0 Kilometers 50
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44

Figure 2

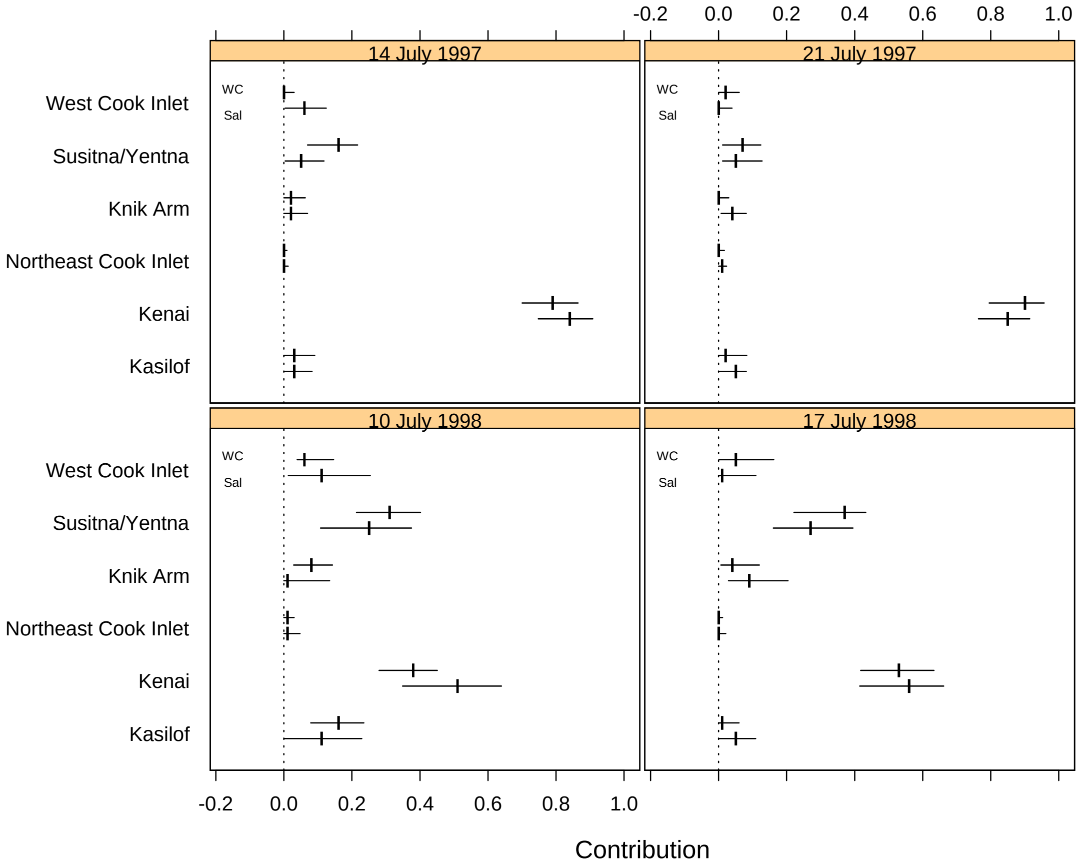
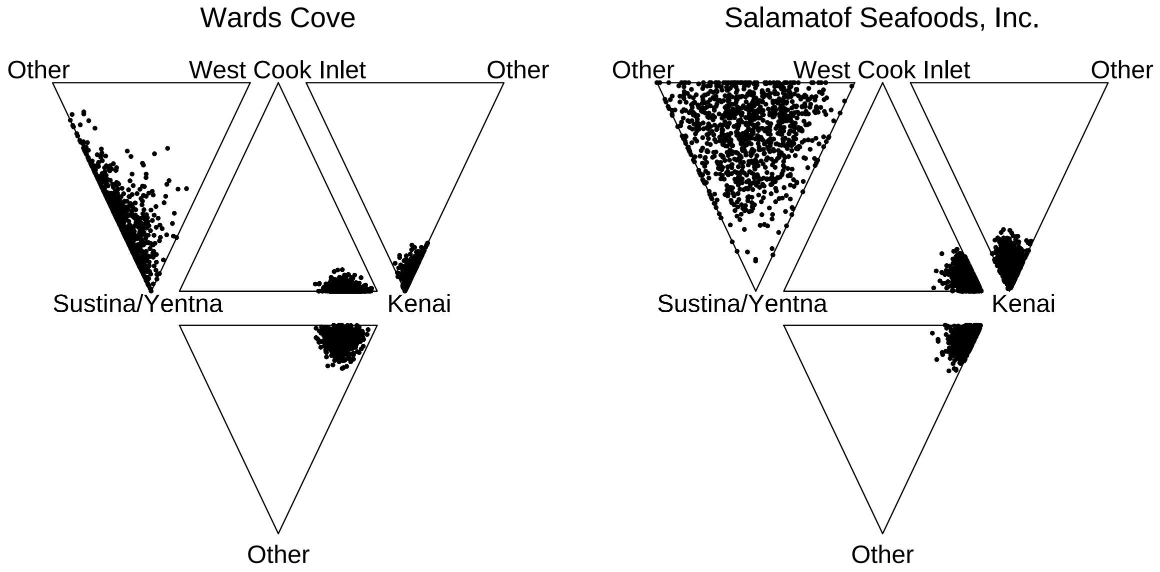

-0.2
0.0
0.2
0.4
0.6
0.8
1.0
14 July 1997
21 July 1997
10 July 1998
17 July 1998
WC
Sal
West Cook Inlet
Susitna/Yentna
Knik Arm
Northeast Cook Inlet
Kenai
Kasilof
Contribution

Figure 3



Wards Cove

Salamatof Seafoods, Inc.