The Estimation of Stock Composition in Mixed Stock Fisheries Using the Program SPAYK

By

Jeffrey F. Bromaghin
&
Dana A. Bruden

Regional Information Report[1] No. 3A99-01
Alaska Department of Fish and Game
Division of Commercial Fisheries
Arctic-Yukon-Kuskokwim Region
333 Raspberry Road, Anchorage, Alaska 99518

January 1999

[1] The Regional Information Report Series was established in 1987 to provide an information access system for all unpublished divisional reports. These reports frequently serve diverse ad hoc informational purposes or archive basic uninterpreted data. To accommodate timely reporting of recently collected information, reports in this series undergo only limited internal review and may contain preliminary data; this information may be subsequently finalized and published in the formal literature. Consequently, these reports should not be cited without prior approval of the author or the Division of Commercial Fisheries.

AUTHORS

Jeffrey F. Bromaghin is the Arctic-Yukon-Kuskokwim Regional Biometrician for the Alaska Department of Fish and Game, Commercial Fisheries Division, 333 Raspberry Road, Anchorage, Alaska 99518.

Dana A. Bruden is an Arctic-Yukon-Kuskokwim Assistant Regional Biometrician for the Alaska Department of Fish and Game, Commercial Fisheries Division, 333 Raspberry Road, Anchorage, Alaska 99518.

SPONSORSHIP

This investigation was partially funded by U.S./Canada salmon research Cooperative Agreement Award number NA76FP0208.

OEO/ADA STATEMENT

The Alaska Department of Fish and Game administers all programs and activities free from discrimination on the basis of sex, color, race, religion, national origin, age, marital status, pregnancy, parenthood or disability. For information on alternative formats available for this and other department publications, contact the department ADA Coordinator at (voice) 907-465-4120, or (TDD) 907-465-3646. Any person who believes s/he has been discriminated against should write to: ADF&G, P.O. Box 25526, Juneau, AK 99802-5526; or O.E.O., U.S. Department of the Interior, Washington, D.C. 20240.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDICES

INTRODUCTION

Researchers and managers often need information about salmon populations when fish from multiple stocks are present simultaneously. In such cases, knowledge of the stock composition of the mixture is often of primary importance. Stock composition, or stock identification, studies are common in salmon fishery management and research, and can be based on many types of information. For example, radio-telemetry, or one of several types of physical marks or tags, have been useful in certain applications (e.g., Shaul and Clark, 1990; Barton, 1992). Genetic information has become widely used to study population structure and estimate stock composition in recent years (e.g., Wilmot et al., 1994; Beacham et al., 1996; Crane et al., 1996). Some studies have utilized parasite information (Moles et al., 1990), and others have combined a variety of data types (Fournier et al, 1984). Historically, a large number of studies have been based on growth information from fish scales (e.g., Cook and Lord, 1978; Schneiderhan, 1997)

Substantial research on estimation methods, which is partially independent of the type of information available, has been conducted. The two most commonly applied methods are classification estimators, i.e., estimators based on classifying individual fish to stocks, and maximum likelihood estimators based on a probability model. Both of these techniques require information characterizing each stock, termed stock standards, to be obtained by sampling stocks when they are segregated. Each stock standard is assumed to characterize individuals from that stock present in mixtures. Stock standards can be used in conjunction with several types of stock composition estimators. Studies based upon scale measurement data have tended to use classification estimators (e.g., Cook and Lord, 1978; Schneiderhan, 1997). Hand (1997) provides an overview of various classification methods. Maximum likelihood methods (Stuart and Ord, 1991) have been used in the majority of genetic studies. Millar (1990) compares several stock composition estimators, and Pella et al. (1996) provides a comparison of algorithms for obtaining stock composition estimates under a maximum likelihood model.

The Alaska Department of Fish and Game has conducted a stock identification project for Yukon River chinook salmon annually since 1980 (McBride and marshall, 1983). The project's objective is to estimate the stock composition of all harvests within the Yukon River drainage. Schneiderhan (1997) summarizes the history of the project and the estimation methods used.

In the spring of 1998, the Alaska Department of Fish and Game initiated an effort to streamline and automate the analytical methods used in the Yukon River chinook salmon stock identification project. A second objective was to improve statistical methods to the degree feasible. Several computer programs that implemented the analytical methods were replaced with a single computer program. The new computer program was named SPAYK; a combination of the commonly used reference for the prior methods, scale pattern analysis (SPA), and the acronym for the Arctic-Yukon-Kuskokwim (AYK) Region. Implementation of all methods within a single computer program substantially streamlines the analysis process. The need for subjective input and manual manipulation of intermediate results was largely eliminated. A maximum likelihood estimator was implemented in SPAYK, which Bromaghin and Bruden (1998) found to be superior to the classification estimator used previously. A number of other more minor changes to the methods were also made. The purpose of this report is to document the methods implemented in the computer program SPAYK, and to serve as a guide to users of the program. Although SPAYK was designed for the Yukon River chinook salmon application, it is sufficiently general that it may be useful in other applications using the same type of scale growth measurements.

METHODS

Overview

The central feature of SPAYK is a maximum likelihood estimator of stock composition based on measurements of the distances between scale circuli. Salmon aggregations may contain individuals from various stocks and of various age classes, and growth may differ as a result of either genetic or environmental factors that vary among stock groups and years, or age classes. Consequently, stock standard data for each stock group and each age class must be obtained at a time or location when stocks are separated, and the age of each fish included in the sample must be determined. Because scale growth is thought to be primarily influenced by environmental conditions, stock standard samples would need to be collected annually in most applications.

Some age classes may be relatively rare, and it may be difficult to obtain stock standard samples of a size sufficient to support a maximum likelihood estimation procedure for all age classes. Those age classes for which sufficiently large stock standard samples are available are termed major ages. The stock composition of major ages is estimated using maximum likelihood techniques. The stock composition of less abundant age classes, termed minor ages, is estimated using an ad hoc procedure based on stock composition estimates for major ages and ratios of age composition estimates (Schneiderhan, 1997).

For each major age, stock standard data, in the form of files containing scale growth measurements created by a computer-controlled digitizing system, are read and processed by SPAYK. A large number of variables are computed from the basic measurement data, and a variable selection algorithm is used to select a subset of the variables that best differentiate the stock groups. These data are assumed to characterize fish of each major age from each of the stock groups, and form the basis of the maximum likelihood estimation procedure.

The stock composition of harvests may be estimated in one of three ways. If digitized scale data are available from a harvest sample, the stock compositions of the major ages are estimated using maximum likelihood methods, and the stock composition of minor ages are estimated using the method described by Schneiderhan (1997). If digitized scale data are not available from a harvest, the stock composition may be estimated using the estimated stock composition from one or more harvests presumed to have similar stock composition. The entire harvest may also be assigned to a particular stock group based on the geographic location of the harvest. In these later two cases, the stock composition can be estimated within or across age classes, depending on whether or not an age composition estimate of the harvest is available. Age-specific stock composition estimates are multiplied by the size of the harvest, resulting in the estimated number of fish harvested by stock group and age class.

SPAYK is designed to read and process stock standard data for each major age, select a subset of variables for each major age that best distinguish between the stocks, and estimate the stock composition of all harvests in one program execution. All input controlling program execution is obtained from one ASCII file, termed the "control file", which also serves to document the process. Appendix A contains an example control file from the Yukon River chinook salmon stock identification program. Program results are written to an ASCII file as they are obtained, an example of which is given in Appendix B. Although the output style is primitive, it allows the user to select the output of interest and format it as necessary for final utilization.

Computer Program

The program SPAYK is written in the FORTRAN programming language (Metcalf and Reid, 1996). The program was compiled using the Digital Equipment Corporation Visual FORTRAN Professional Edition compiler[1], version 5.0.C. Some components of the code utilize routines in the Visual Numeric[1] IMSL library bundled with the compiler.

[1] Use of product names does not constitute endorsement by the Alaska Department of Fish and Game.

Age Class Definitions

The control file must contain a list of all possible age classes, in European notation. Following each age class is a major-age indicator variable and a surrogate age class, also in European notation. The major-age indicator variable can have a value of one or zero, with a one indicating the age class is a major age and a zero indicating the age class is a minor age. A surrogate age class is a major age class used in the estimation of stock composition of minor age classes, and major age classes when the sample size from a particular harvest is very small or not available (see Estimating Stock Composition, below). This information is located in Section 1 of a control file (Appendix A).

Digitized Scale Data Files

SPAYK is designed to utilize scale measurement data in one or two freshwater zones, the freshwater-plus growth zone, and as many as the first three ocean growth zones. Data files must be in the format created by the scale digitizing software SCALE3, custom-developed by the Alaska Department of Fish and Game (Beverly Cross, Alaska Department of Fish and Game, personal communication). An example of such a data file is given in Appendix C. Integer "keys" precede each measurement in a data file, and indicate the zone in which a measurement was made. The keys used for each of the scale zones must be the same in all data files, both stock standard and harvest sample files, that are to be used in any one execution of SPAYK. The keys for each of the zones are defined in Section 2 of a control file (Appendix A). In the Yukon River chinook salmon application, integer keys 1-5 are typically used for the first freshwater zone, the freshwater-plus zone, and the first three ocean zones, respectively.

Stock Standards

Section 3 of a control file must contain a name of each stock group for which stock composition estimates are to be computed. The names are primarily for use in the program output. The Yukon River chinook salmon application usually generates stock composition estimates for three stock groups, termed Lower, Middle, and Upper (Schneiderhan, 1997).

There must be at least one file of stock standard data for each major age class from each stock group. However, multiple stock standard samples may be obtained for a given stock group and major age class. For example, in the Yukon River chinook salmon application, it is common to obtain samples from the Chena and Salcha Rivers, both of which are in the Middle River stock group. The population associated with a single stock standard file will be referred to as a "stock", and an aggregate of one or more stocks will be referred to as a stock group. For a given major age, SPAYK computes maximum likelihood estimates for each stock, i.e., each stock standard file listed in the control file. These estimates, which are kept internal to the program and are not written to the output file, are subsequently pooled to obtain a stock composition estimate for each of the stock groups listed in a control file (Section 3). The number of stock standard files listed may be different for each major age. The stock standard file names are listed in Section 4 of a control file (Appendix A).

The user must decide whether to combine data from multiple stocks within a stock group into a single file, or to keep sample data in separate files. Results of hypothesis tests that the samples were drawn from identical populations (Anderson, 1984) might be used as a guide. However, many such tests are sensitive to assumption violations, or find differences of little practical significance to be of substantial statistical significance. For that reason, no comparative tests were implemented in SPAYK, and the decision must be made based on less rigorous considerations. If multiple samples are to be pooled, the data files must be combined prior to running SPAYK, and the name of the combined data file must be listed in the control file.

There are several issues to consider when deciding whether or not to pool stocks within stock groups. One guideline is that samples should not be pooled if there is a reasonable expectation that the scale growth characteristics of the populations being sampled are likely to differ. Sample size is another primary concern. It is difficult to give specific recommendations on sample size to guide making decisions on whether or not to pool samples that would be applicable across a range of situations. However, small sample sizes, perhaps less than 30 to 50 fish, should be avoided if possible. The use of very small samples, perhaps less than 20 fish, may cause the program to fail or, even worse, cause highly spurious results because of the poor estimation of population parameters based on small samples. Larger sample sizes are often necessary to accurately estimate population parameters when a large number of variables are selected for use in maximum likelihood estimation. A general consideration is that using more samples is likely to reduce bias and modestly increase variability, while the inappropriate pooling of samples may lead to increased bias.

Variable Selection

As many as 113 variables are computed from the basic scale measurements observed from each fish scale. Table 1 contains a complete listing of the variables that are computed. The variables are divided into "sets" of variables that are expected to reflect similar growth information. The set each variable belongs to is indicated in Table 1. For each major age, a variable selection strategy is implemented to select a subset of the variables for inclusion in the maximum likelihood probability model. The variables are evaluated based upon their ability to distinguish between the stocks in the model.

The variable selection strategy consists of two stages. The first stage is primarily designed to select one variable from each variable set for further consideration. The user specifies the maximum allowable proportion of missing observations for a variable to be given consideration (Section 6 of a control file). If the proportion of missing observations for a variable exceeds the maximum allowable proportion, it is excluded from further consideration. A one-way analysis of variance (Montgomery, 1984) is performed with each of the remaining variables within each variable set (Table 1). The variable having the largest F statistic between the stocks within each variable set is selected for further consideration. To maintain consistency between methods previously used in the Yukon River chinook salmon application (Schneiderhan, 1997), variables having a large negative correlation with those variables selected based on analysis of variance results are also retained in the first stage. The user specifies the minimum required negative correlation in Section 6 of a control file. Note that if the user does not wish to consider variables having a large negative correlation, a minimum necessary negative correlation less than -1.0, say -1.1, can be specified in a control file.

In the second stage of the variable selection strategy, a stepwise variable selection algorithm is used to select a final subset of variables from the variables remaining under consideration. Before detailing the stepwise variable selection routine, some notation is needed. Note that although the following notation is not age specific, the methods are applied separately for each of the major age classes. Let

S = the number of stock standard samples,

n_i = the number of fish in the ith stock standard sample,

$n. = \sum_{i=1}^{S} n_i$,

k = the number of variables previously selected,

$S_{(k)i}$ = estimated dispersion matrix, based on k-dimensional data vectors, for the ith stock standard sample,

$S_{(k)C}$ = estimated dispersion matrix, based on k-dimensional data vectors, for all stock standard samples combined, and

$D_{(k)}(y)$ = a k-dimensional diagonal matrix having the constant y on the diagonal.

Given that notation, the total sum of squares between individual data vectors and the overall mean across all stock standard samples is

$$T_{(k)} = D_{(k)}(n. - 1)S_{(k)C} , \quad (1)$$

and the sum of squares within stock standard samples is

$$E_{(k)} = \sum_{i=1}^{S} D_{(k)}(n_i - 1)S_{(k)i} . \quad (2)$$

Wilks' likelihood ratio statistic (Seber, 1984), based on k variables, can be expressed as

$$\Lambda_{(k)} = \frac{|E_{(k)}|}{|T_{(k)}|} , \quad (3)$$

where $|M|$ indicates the determinant of the matrix M, and $\Lambda_{(0)}$ is defined to be 1.

Each iteration of the variable selection process consists of a selection step and an elimination step. At the beginning of an iteration, k variables have been selected. Each of the remaining variables are considered in turn, and the test statistic

$$\Psi_{(k+1)} = \frac{n. - S - k}{S - 1}\left(\frac{\Lambda_{(k)}}{\Lambda_{(k+1)}} - 1\right) \quad (4)$$

is computed for each. Assuming that the stock standard samples have multivariate normal distributions with equal dispersion matrices, $\Psi_{(k+1)}$ has an F distribution with (S–1) numerator and (n.–S–k) denominator degrees of freedom (Seber, 1984). The p-value of the largest $\Psi_{(k+1)}$ among the variables not yet selected is computed. If the p-value is less than a specified threshold, termed p-to-enter, the variable is selected, otherwise the variable selection process terminates. If a new variable is selected, an elimination step is taken. The test statistic $\Psi_{(k+1)}$ is computed for each of the previously selected variables, except the variable added in the immediately preceeding selection step, as if it were being considered for inclusion. The p-value of the smallest $\Psi_{(k+1)}$ among the variables is computed. If the p-value is greater than a specified threshold, termed p-to-remove, the variable is eliminated, and a new elimination step is initiated with one fewer variable. If a variable is not eliminated in an elimination step, the variable selection procedure continues with the initiation of a new selection step.

The thresholds p-to-enter and p-to-remove are specified in Section 6 of a control file. The user must select values appropriate for a given application. The only strict requirement is that p-to-enter must be less than p-to-remove. Only general guidelines for selecting values can be given. As the value of p-to-enter increases, more variables will tend to be selected. A small number of variables can be advantageous in that computations are less time consuming, and problems with colinearity are more likely to be avoided. However, a larger number of variables will tend to contain a larger portion of the information contained in the entire dataset, and variables that are not individually important can provide a substantial information collectively. If the user is certain that only a very small number of variables are important, a small p-to-enter, say 0.05 to 0.10, might be specified. However, a larger p-to-enter of 0.15 to 0.25 might be preferred in most applications. The p-to-remove should be set somewhat larger than p-to-enter, and can be set to 1.0 if the user wishes to implement a forward selection algorithm with no elimination step.

Estimation of Mean Vectors and Dispersion Matrices

The mean vector and the dispersion matrix of the variables selected for inclusion in the model must be estimated for each stock standard and for each major age. The computer program SPAYK allows the user to select one of three estimation options. One option is to use the traditional unbiased sample mean and sample dispersion estimators (Seber, 1984), which have been used previously in the Yukon River chinook stock identification program. One potential disadvantage of these estimators is that they may be sensitive to extreme observations, i.e., outliers, particularly if some of the stock standard sample sizes are small relative to the number of variables selected for inclusion in the model. Because the use of small stock standard sample sizes has been relatively common in Yukon River chinook stock identification studies, the robust estimators of Campbell (1980) (also see Seber, 1984) are implemented as a second option. A third option is to use the robust estimators unless the algorithm fails, in which case the traditional sample estimators are implemented. Section 5 of a control file contains information pertinent to the user's selection of estimation options.

Because Campbell's robust estimators may be less well known than the traditional sample estimators, inclusion of a brief description and a small example is warranted here. The robust estimators are designed to reduce the influence of extreme observations on estimates through use of a weight function. The estimation process is iterative, using the traditional sample estimates as the initial estimates. A weight for each observation is computed as a function of its Mahalanobis distance (Seber, 1984), which is essentially a multivariate z-score. The iterative process is then initiated, with each iteration consisting of two steps. In the first step, estimates of the mean vector and the dispersion matrix are computed, using weights for each observation. In the second step, the weights are recomputed using the most recent estimates of the mean vector and the dispersion matrix. Iterations continue until the greatest absolute change in the estimates is smaller than a user-specified threshold, or until the user specified maximum number of iterations is exceeded.

To demonstrate the potential performance of the robust estimators, a small example dataset consisting of 30 observations of a 4-dimension vector was constructed by pseudo-random sampling from a multivariate normal density. Table 2 lists the observations, the Mahalanobis distance for each observation at the initial and final iteration, and the weight of each observation at the final iteration. The final weights of all but one observation were 1.0, and the weight of the exception is only slightly less than 1.0. The traditional sample estimates and robust estimates of the mean vector and the dispersion matrix of these data were obtained and are presented in Table 3. Because no outliers were detected in the iterative process, the estimates generated by the two methods are very similar.

A second example dataset was constructed by appending two outliers to the data presented in Table 2. The observations, the Mahalanobis distances for each observation at the initial and final iteration, and the weights at the final iteration are listed in Table 4. Note that the outliers (observations 31 and 32) had relatively large initial Mahalanobis distances, with substantially increased Mahalanobis distances on the final iteration. At the beginning of the iteration process, the outliers influence the estimates, and consequently they do not appear too dissimilar to the bulk of the data. As the iteration process continues, the outlier's effect on the estimates is gradually reduced as they are given increasingly smaller weights. On the final iteration, the influence of the two outliers is nearly eliminated. The traditional sample estimates and robust estimates of the mean vector and the dispersion matrix of these data are presented in Table 5. Note that the estimates are quite dissimilar, but that the robust estimates are very similar to the estimates presented in Table 3.

Escapement Age Composition Estimates

An estimate of the age composition of the escapement must be obtained for each stock group listed in Section 3 of a control file. These estimates are used in estimating the stock composition of minor age classes (see METHODS: Estimating Stock Composition below). The escapement age composition estimates must be listed in Section 8 of a control file.

In the Yukon River chinook salmon stock identification example, age composition estimates for stocks spawning in the mainstem Yukon River drainage in Canada are often obtained from fishwheels near the U. S. – Canada border. In this case, the age composition estimates should first be adjusted to account for unequal capture probabilities prior to using SPAYK.

Estimating Stock Composition

Each harvest sample for which stock composition estimates are desired must be assigned a sample number and listed in Section 9 of a control file. Several types of information must be provided for each harvest sample, including descriptive text strings, the number of fish harvested, the estimation of age composition, and a variable indicating how the stock composition is to be estimated. The stock composition can be estimated in one of three ways, depending on the data available.

If an age composition estimate of a harvest and digitized scale samples of the major age classes are available (Method 1), the names of data files containing the digitized scale data must be provided. In this case, the stock composition proportions are parameters of a probability mixture model, and the proportions are estimated using maximum likelihood techniques. In order to specify the likelihood function, it is necessary to introduce some additional notation. Although the following notation is not age specific, maximum likelihood estimation is performed separately for each major age class. Let

S = the number of stock standards in the model,

n = the number of fish in the mixture sample,

x_i = data vector of the i^{th} fish in the sample,

$f(x|i)$ = a multivariate normal pdf with mean μ_i and dispersion Σ_i evaluated at the data vector x, and

π_i = the proportion of a mixture composed of fish from stock i.

With that notation, the likelihood function, L, can be expressed as

$$L = \prod_{i=1}^{n} \sum_{j=1}^{S} \pi_j f(x_i|j). \tag{5}$$

The maximum likelihood estimate of the vector $\pi = (\pi_i)$, denoted $\hat{\pi}$, is the value that maximizes L. Unfortunately, $\hat{\pi}$ can not be expressed in closed form, and estimates must be obtained using numerical optimization techniques. The computer program SPAYK maximizes the likelihood function using the DLCONG subroutine of the Visual Numeric[1] IMSL FORTRAN library. The variance of the estimates is estimated using the Rao-Cramér information limit (Rao, 1973). If more than one stock standard is available for a stock group, estimates are obtained for each individual stock and are then pooled

[1] Use of product names does not constitute endorsement by the Alaska Department of Fish and Game.

on the basis of stock groups (see METHODS: Stock Standards above). The stock compositions of minor age classes are estimated using methods described by Schneiderhan (1997).

If a sample of digitized scale data is unavailable for one of the major age classes for a particular harvest, maximum likelihood estimation is not possible. Similarly, if a sample is available, but the sample size is extremely small, maximum likelihood estimation may not be advisable. In these cases, the stock composition of the major age class can be estimated as if it was a minor age class. This can be accomplished using the computer program SPAYK by specifying NONE as the name of the file containing digitized scale data in Section 9 of a control file. A second major age class must also be listed as a surrogate for the major age class in question in Section 1 of a control file (see METHODS: Age Class Definitions above).

The estimated stock composition within each age class is multiplied by the estimated age class composition of the entire harvest to estimate composition proportions by age class and stock group. These proportions are multiplied by the size of the harvest to obtain harvest by both age class and stock group. The estimated harvests by age class and stock group are rounded to an integer number of fish in such a way that they sum to the total harvest size specified in a control file.

The stock and age composition of harvests from which no digitized scale samples are available can be estimated in one of two ways (Methods 2 and 3). Maximum likelihood estimates from one or more representative harvests assumed to have similar stock composition can be applied (Method 2). The number of representative harvests and their sample numbers must be specified in a control file. If results from more than one representative harvest are used, the estimated proportions are weighted by harvest size. If age composition estimates for the harvest are available, stock composition estimates are applied separately within each age class. If age composition estimates are not available, both the stock group composition and age class composition estimates are obtained from the representative harvests. The stock compositions of harvests are estimated in the order that harvests are listed in a control file. For that reason, estimation for a particular harvest can only be based on representative harvests listed previously in a control file.

The entire harvest can also be assigned to a single stock group (Method 3). This option is useful if harvests are taken in a location or at a time when the harvest can be assumed to consist of fish from only a single stock group. For example, in the Yukon River chinook salmon stock identification project, harvests within the Tanana River drainage are wholly apportioned to the Middle stock group. In this case, as in Method 2, age class composition estimates can be obtained from a sample or taken from one or more representative harvests.

Evaluating Estimation Accuracy

SPAYK has the capability to simulate the estimation accuracy of the maximum likelihood estimator for each major age class, given the variables selected for inclusion in the model. Either unbiased-sample or robust estimators of the mean vector and the dispersion matrix can be employed. For each stock standard, an artificial mixture sample consisting of fish from only that stock is constructed. Samples can be constructed by randomly sampling from the multivariate normal distribution defined by the stock standard, or by randomly sampling individual fish with replacement from the stock standard data. In either case, samples of a size equal to the sample size of the stock standard sample are drawn. A maximum likelihood estimate of the stock composition of each artificial mixture sample is obtained, and the average estimate over all simulations is computed. Simulations are conducted separately for each stock standard, and the results are pooled into the appropriate stock groups. Information pertaining to accuracy simulations is contained in Section 7 of a control file.

These simulations provide an indication of the computer program's ability, given the stock standard data, to distinguish between the stock groups. Such simulation results should be informative, and provide a good index of the estimator performance given the stock standard data available. However, there is no guarantee that the simulation results will accurately reflect estimator performance in more complex mixture samples consisting of fish from multiple stock groups. Additionally, the estimator may perform better with artificial mixtures drawn from the stock standard data used to estimate the mean and dispersion of each stock than with true harvest samples.

Canonical scores (Seber, 1984) for each fish in the stock standard data are computed and written to an output file. A scatter-plot of the first two canonical scores provide an informative visual summary of the differentiability among the stock groups.

CONCLUSIONS

The development of the new program SPAYK was motivated by a desire to streamline steps in the data analysis process used in the Yukon River chinook salmon stock identification project, and to improve statistical methods used wherever feasible. Although additional improvements or efficiencies can always be made in any automated data analysis procedure, and SPAYK is no exception, we believe that success was achieved with respect to both objectives.

SPAYK has replaced five separate computer programs previously used in the Yukon River chinook salmon stock identification project. The need to manually manipulate output from one program as input for a subsequent program has been eliminated, as all components of the analysis are completed in a single execution of SPAYK. Also, the need for subjective decision-making has been largely eliminated. Perhaps most importantly, the variable selection component of the analysis has been automated. The variable selection algorithm implemented in SPAYK closely mimics procedures previously performed manually. However, the automation of that process should improve consistency in the application of the process, and reduce the time required to complete an analysis. Overall, the analysis procedure has been streamlined, and the opportunities for errors to occur have been reduced. In addition, the results should be more consistent between analyses, which may be an important consideration in an ongoing annual program.

Statistical improvements were also made in some aspects of the estimation methods. Most importantly, a maximum likelihood estimator of stock composition was implemented. The superiority of the maximum likelihood estimator in Yukon River chinook salmon applications was empirically demonstrated by Bromaghin and Bruden (1998). The implementation of robust estimators of a mean vector and a dispersion matrix may also be advantageous in many applications. Use of robust estimators should protect against the undue influence of extreme observations, particularly when sample sizes are small, and they may also help avoid violations of the multivariate normal assumption.

An additional advantage of SPAYK is that all program input is read from a control file. A control file can be fairly complicated, and must be constructed with care. If errors are to be made in the estimation process, they will be made as incorrect entries in a control file. However, this approach allows the project leader to focus on the data available, and how to best estimate the stock composition of a harvest, without having to be concerned with the analytical details of the estimation process. An important property of a control file is that it wholly documents data files and options used in the estimation process. A control file can be examined years after an analysis, and all details of the analysis can be easily discerned. An additional advantage of the control file is that it can easily be modified, and the analysis repeated with the changed input. For example, in the Yukon River chinook salmon stock identification project, the original analysis is often performed with preliminary harvest estimates for some fisheries. In such cases, the preliminary harvest estimates in a control file could be replaced with final harvest numbers some months later, and a final analysis would be obtained by simply executing SPAYK again.

The program SPAYK was developed specifically for the Yukon River chinook salmon stock identification project. However, some aspects of the program design were intentionally generalized to accommodate future changes in that project, as well as potential use in other applications. It is likely that SPAYK could be used in other stock composition applications based upon similar scale measurement data with little or no modification.

LITERATURE CITED

Anderson, T. W. 1984. An Introduction to Multivariate Statistical Analysis, 2nd ed. New York: John Wiley & Sons.

Barton, L. H. 1992. Tanana River, Alaska, fall chum salmon radio telemetry study. Alaska Department of Fish and Game, Division of Commercial Fisheries, Fishery Research Bulletin No. 92-01. Juneau.

Beacham, T. D., Withler, R. E., and Stevens, T. A. 1996. Stock identification of chinook salmon, *Oncorhynchus tshawytscha*. Fishery Bulletin 94: 611-627.

Bromaghin, J. F. and Bruden, D. L. 1998. A simulation study of classification and maximum likelihood estimators of the stock composition of Yukon River chinook salmon harvests. Alaska Department of Fish and Game, Division of Commercial Fisheries, Arctic-Yukon-Kuskokwim Region, Regional Information Report No. 3A98-31.

Campbell, N. A. 1980. Robust procedures in multivariate analysis. I. Robust covariance estimation. Applied Statistics 27: 251-258.

Cook, R. and Lord, G. 1978. Identification of stocks of Bristol Bay sockeye salmon, *Oncorhynchus nerka*, by evaluating scale patterns with a polynomial discriminant method. Fishery Bulletin 76: 415-423.

Crane, P. A., Templin, W. D., and Seeb, L. W. 1996. Genetic stock identification of Alaska chinook salmon. Alaska Department of Fish and Game, Division of Commercial Fisheries Management and Development, Regional Information Report No. 5J96-17. Juneau.

Fournier, D. A., Beacham, T. D., Riddell, 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.

Hand, D. J. 1997. Construction and Assessment of Classification Rules. Chichester, England: John Wiley & Sons.

McBride, D. N. and Marshall, S. L. 1983. Feasibility of scale pattern analysis to identify the origins of chinook salmon (*Oncorhymchus tshawytscha* Walbaum) in the lower Yukon River commercial gillnet fishery, 1980-1981. Alaska Department of Fish and Game, Division of Commercial Fisheries, Informational Leaflet No. 208. Juneau.

Metcalf, M. and Reid, J. 1996. FORTRAN 90/95 Explained. New York: Oxford University Press.

Millar, R. B. 1990. Comparison of methods for estimating mixed stock fishery composition. Canadian Journal of Fisheries and Aquatic Sciences 47: 2235-2241.

Moles, A., Rounds, P., and Kondzela, C. 1990. Use of the brain parasite *Myxobolus neurobius* in separating mixed stocks of sockeye salmon. American Fisheries Society Symposium 7: 224-231.

Montgomery, D. C. 1984. Design and Analysis of Experiments, 2nd edition. New York: John Wiley & Sons.

Pella, J. Masuda, M., and Nelson, S. 1996. Search algorithms for computing stock composition of a mixture from traits of individuals by maximum likelihood. United States Department of Commerce, NOAA Technical Memorandum NMFS-AFSC-61.

Rao, C. R. 1973. Linear Statistical Inference and Its Applications, 2nd edition. New York: John Wiley & Sons.

Seber, G. A. F. 1984. Multivariate Observations. New York: John Wiley & Sons.

Schneiderhan, D. J. 1997. A history of scale pattern analysis as applied to stock identification of chinook and chum salmon in the Yukon River. Alaska Department of Fish and Game, Division of Commercial Fisheries Management and Development, Arctic-Yukon-Kuskokwim Region, Regional Information Report No. 3A97-33, Anchorage.

Schaul, L. D. and Clark, J. E. 1990. Use of coded wire tag data to estimate aggregate stock composition of salmon catches in multiple mixed-stock fisheries. American Fisheries Society Symposium 7: 588-603.

Stuart, A. and Ord, J. K. 1991. Kendall's Advanced Theory of Statistics, Volume 2: Classical Inference and Relationship, 5th ed. New York: Oxford University Press.

Wilmot, R. L., Everett, R. J., Spearman, W. J., Baccus, R., Varnavskaya, N. V. and Putivkin, S. V. 1994. Genetic stock structure of Western Alaska chum salmon and a comparison with Russian Far East stocks. Canadian Journal of Fisheries and Aquatic Sciences 51 (Suppl. 1): 84-94.

Table 1. Variables computed from scale measurement data and considered for inclusion in the maximum likelihood mixture model in the computer program SPAYK. The column entitled Variable Set indicates groupings of closely related variables.

Variable Number	Variable Set	Scale Zone	Variable Description
1	1	Freshwater 1	The number of circuli within the zone.
2	1	"	The total distance within the zone.
3	2	"	The distance from the scale focus to circulus 2.
4	2	"	The distance from the scale focus to circulus 4.
5	2	"	The distance from the scale focus to circulus 6.
6	2	"	The distance from the scale focus to circulus 8.
7	3	"	Variable 4 - Variable 3
8	3	"	Variable 5 - Variable 3
9	3	"	Variable 6 - Variable 3
10	4	"	Variable 5 - Variable 4
11	4	"	Variable 6 - Variable 4
12	5	"	The distance from the 4th circulus preceding the end of the zone to the end of the zone.
13	5	"	The distance from the 2nd circulus preceding the end of the zone to the end of the zone.
14	6	"	Variable 2 - Variable 3
15	6	"	Variable 2 - Variable 4
16	7	"	Variable 3/Variable 2
17	7	"	Variable 4/Variable 2
18	7	"	Variable 5/Variable 2
19	7	"	Variable 6/Variable 2
20	8	"	Variable 7/Variable 2
21	8	"	Variable 8/Variable 2
22	8	"	Variable 9/Variable 2
23	9	"	Variable 10/Variable 2
24	9	"	Variable 11/Variable 2
25	10	"	Variable 12/Variable 2
26	10	"	Variable 13/Variable 2
27	11	"	Variable 2/Variable 1
28	12	"	The number of circuli within the first 75% of the zone.
29	13	"	The maximum distance between consecutive circuli.
30	13	"	Variable 29/Variable 2
31	14	Freshwater 2	The number of circuli within the zone.
32	14	"	The total distance within the zone.
33	15	"	The distance from the scale focus to circulus 2.
34	15	"	The distance from the scale focus to circulus 4.
35	15	"	The distance from the scale focus to circulus 6.
36	15	"	The distance from the scale focus to circulus 8.
37	16	"	Variable 34 - Variable 33
38	16	"	Variable 35 - Variable 33
39	16	"	Variable 36 - Variable 33
40	17	"	Variable 35 - Variable 34
41	17	"	Variable 36 - Variable 34
42	18	"	The distance from the 4th circulus preceding the end of the zone to the end of the zone.
43	18	"	The distance from the 2nd circulus preceding the end of the zone to the end of the zone.
44	19	"	Variable 32 - Variable 33
45	19	"	Variable 32 - Variable 34
46	20	"	Variable 33/Variable 32
47	20	"	Variable 34/Variable 32
48	20	"	Variable 35/Variable 32
49	20	"	Variable 36/Variable 32
50	21	"	Variable 37/Variable 32
51	21	"	Variable 38/Variable 32
52	21	"	Variable 39/Variable 32
53	22	"	Variable 40/Variable 32
54	22	"	Variable 41/Variable 32
55	23	"	Variable 42/Variable 32
56	23	"	Variable 43/Variable 32
57	24	"	Variable 32/Variable 31
58	25	"	The number of circuli within the first 75% of the zone.
59	26	"	The maximum distance between consecutive circuli.
60	26	"	Variable 59/Variable 32

- continued -

Table 1. Page 2 of 2.

Variable Number	Variable Set	Scale Zone	Variable Description
61	27	Freshwater Plus Growth	The number of circuli within the zone
62	27	"	The total distance within the zone.
63	28	Total Freshwater	Variable 1 + Variable 31
64	28	"	Variable 2 + Variable 32
65	28	"	Variable 1 + Variable 31 + Variable 61
66	28	"	Variable 2 + Variable 32 + Variable 62
67	29	"	Variable 2/Variable 66
68	29	"	Variable 62/Variable66
69	29	"	Variable 32/Variable 66
70	31	Ocean 1	The number of circuli within the zone.
71	31	"	The total distance within the zone.
72	32	"	The distance from the beginning of the zone to circulus 3.
73	32	"	The distance from the beginning of the zone to circulus 6.
74	32	"	The distance from the beginning of the zone to circulus 9.
75	32	"	The distance from the beginning of the zone to circulus 12.
76	32	"	The distance from the beginning of the zone to circulus 15.
77	33	"	Variable 73 - Variable 72
78	33	"	Variable 74 - Variable 72
79	33	"	Variable 75 - Variable 72
80	33	"	Variable 76 - Variable 72
81	34	"	Variable 74 - Variable 73
82	34	"	Variable 75 - Variable 73
83	34	"	Variable 76 - Variable 73
84	34	"	Variable 76 - Variable 74
85	35	"	The distance from the 6th circulus preceding the end of the zone to the end of the zone.
86	35	"	The distance from the 3rd circulus preceding the end of the zone to the end of the zone.
87	35	"	Variable 71 - Variable 72
88	35	"	Variable 71 - Variable 74
89	35	"	Variable 71 - Variable 76
90	36	"	Variable 72/Variable 71
91	36	"	Variable 73/Variable 71
92	36	"	Variable 74/Variable 71
93	36	"	Variable 75/Variable 71
94	36	"	Variable 76/Variable 71
95	37	"	Variable 77/Variable 71
96	37	"	Variable 78/Variable 71
97	37	"	Variable 79/Variable 71
98	37	"	Variable 80/Variable 71
99	38	"	Variable 81/Variable 71
100	38	"	Variable 82/Variable 71
101	38	"	Variable 83/Variable 71
102	38	"	Variable 84/Variable 71
103	39	"	Variable 85/Variable 71
104	39	"	Variable 86/Variable 71
105	40	"	Variable 71/Variable 70
106	41	"	The number of circuli within the first 50% of the zone.
107	42	"	The maximum distance between consecutive circuli.
108	42	"	Variable 107/Variable 71
109	43	Ocean 2	The total distance within the zone.
110	44	Ocean 3	The total distance within the zone.
111	45	Total Ocean	Variable 71 + Variable 109 + Variable 110
112	45	"	Variable 71/Variable 111
113	45	"	Variable 109/Variable 111

Table 2. Example data, with no outliers, used to compare the traditional sample and robust estimators of the mean vector and the dispersion matrix.

	Variable				Mahalanobis Distance		Final
Observation	1	2	3	4	Initial	Final	Weight
1	0.726	41.683	103.959	5.336	1.959	1.959	1.000
2	0.735	48.323	96.790	4.538	1.100	1.101	1.000
3	0.877	45.146	90.163	1.865	1.578	1.578	1.000
4	0.576	43.088	106.025	6.632	2.923	2.922	1.000
5	0.771	42.140	92.497	3.359	1.161	1.161	1.000
6	0.830	26.376	89.580	2.790	3.429	3.440	0.992
7	0.887	51.480	98.620	1.383	1.837	1.837	1.000
8	0.729	39.998	95.305	4.730	1.077	1.078	1.000
9	0.917	45.784	117.034	1.147	2.552	2.552	1.000
10	0.694	45.400	102.170	4.827	1.286	1.285	1.000
11	0.826	48.852	111.310	2.912	1.250	1.250	1.000
12	0.824	44.736	119.303	2.628	2.280	2.280	1.000
13	0.830	46.974	110.699	2.825	1.235	1.234	1.000
14	0.656	46.540	92.937	5.678	1.839	1.839	1.000
15	0.743	37.155	100.058	4.266	1.488	1.492	1.000
16	0.642	48.504	102.197	6.481	2.107	2.108	1.000
17	0.755	44.084	98.413	4.236	0.508	0.508	1.000
18	0.811	38.547	83.826	2.767	1.801	1.803	1.000
19	0.906	52.813	95.491	0.987	2.281	2.281	1.000
20	0.735	41.022	89.133	3.687	1.979	1.979	1.000
21	0.765	34.187	89.869	4.676	2.601	2.604	1.000
22	0.691	43.282	96.282	4.379	2.158	2.157	1.000
23	0.825	43.899	77.755	3.123	2.495	2.494	1.000
24	0.752	51.597	99.412	5.143	2.830	2.830	1.000
25	0.875	52.064	91.584	1.315	2.407	2.407	1.000
26	0.705	48.399	105.816	4.980	1.157	1.158	1.000
27	0.795	52.464	125.279	3.364	2.423	2.423	1.000
28	0.861	49.529	102.624	2.425	1.352	1.352	1.000
29	0.761	50.645	109.766	4.371	1.480	1.481	1.000
30	0.766	46.928	107.325	3.883	0.703	0.703	1.000

Table 3. Estimated mean vector and dispersion matrix obtained for the example data, with no outliers, using the traditional sample and robust estimators.

Mean Vector

Traditional Sample Estimates				Robust Estimates			
Variable				Variable			
1	2	3	4	1	2	3	4
0.7756	45.0546	100.0409	3.6911	0.7755	45.0594	100.0435	3.6913

Dispersion Matrix

Traditional Sample Estimates					Robust Estimates				
	Variable					Variable			
Variable	1	2	3	4	Variable	1	2	3	4
1	0.0066	0.0777	0.0009	-0.1186	1	0.0067	0.0783	0.0012	-0.1186
2	0.0777	34.5054	25.6523	-1.4569	2	0.0783	34.3401	25.5631	-1.4665
3	0.0009	25.6523	111.2481	0.4082	3	0.0012	25.5631	111.2491	0.4034
4	-0.1186	-1.4569	0.4082	2.3069	4	-0.1186	-1.4665	0.4034	2.3077

Table 4. Example data, with two outliers, used to compare the traditional sample and robust estimators of the mean vector and the dispersion matrix.

	Variable				Mahalanobis Distance		Final
Observation	1	2	3	4	Initial	Final	Weight
1	0.726	41.683	103.959	5.336	1.665	1.946	1.000
2	0.735	48.323	96.790	4.538	1.125	1.100	1.000
3	0.877	45.146	90.163	1.865	1.551	1.574	1.000
4	0.576	43.088	106.025	6.632	1.993	2.928	1.000
5	0.771	42.140	92.497	3.359	0.896	1.172	1.000
6	0.830	26.376	89.580	2.790	3.525	3.442	0.992
7	0.887	51.480	98.620	1.383	1.908	1.837	1.000
8	0.729	39.998	95.305	4.730	1.136	1.079	1.000
9	0.917	45.784	117.034	1.147	2.419	2.546	1.000
10	0.694	45.400	102.170	4.827	0.846	1.295	1.000
11	0.826	48.852	111.310	2.912	1.110	1.239	1.000
12	0.824	44.736	119.303	2.628	1.987	2.280	1.000
13	0.830	46.974	110.699	2.825	1.095	1.225	1.000
14	0.656	46.540	92.937	5.678	1.821	1.846	1.000
15	0.743	37.155	100.058	4.266	1.421	1.496	1.000
16	0.642	48.504	102.197	6.481	2.150	2.106	1.000
17	0.755	44.084	98.413	4.236	0.512	0.501	1.000
18	0.811	38.547	83.826	2.767	1.822	1.809	1.000
19	0.906	52.813	95.491	0.987	2.368	2.281	1.000
20	0.735	41.022	89.133	3.687	1.277	1.987	1.000
21	0.765	34.187	89.869	4.676	2.364	2.596	1.000
22	0.691	43.282	96.282	4.379	0.965	2.165	1.000
23	0.825	43.899	77.755	3.123	2.168	2.490	1.000
24	0.752	51.597	99.412	5.143	1.786	2.812	1.000
25	0.875	52.064	91.584	1.315	2.469	2.411	1.000
26	0.705	48.399	105.816	4.980	1.178	1.161	1.000
27	0.795	52.464	125.279	3.364	2.276	2.419	1.000
28	0.861	49.529	102.624	2.425	0.989	1.336	1.000
29	0.761	50.645	109.766	4.371	1.251	1.467	1.000
30	0.766	46.928	107.325	3.883	0.618	0.701	1.000
31	0.979	53.021	116.629	2.331	2.758	5.880	0.083
32	0.368	28.459	118.199	5.325	4.910	14.663	0.000

Table 5. Estimated mean vector and dispersion matrix obtained for the example data, with two outliers, using the traditional sample and robust estimators.

Mean Vector

Traditional Sample Estimates				Robust Estimates			
Variable				Variable			
1	2	3	4	1	2	3	4
0.7692	44.7850	101.1267	3.6996	0.7761	45.0817	100.0895	3.6876

Dispersion Matrix

Traditional Sample Estimates					Robust Estimates				
	Variable					Variable			
Variable	1	2	3	4	Variable	1	2	3	4
1	0.0129	0.3415	-0.1218	-0.1413	1	0.0067	0.0787	0.0020	-0.1187
2	0.3415	43.1356	18.8419	-2.5849	2	0.0787	34.3349	25.5823	-1.4695
3	-0.1218	18.8419	122.3664	0.6014	3	0.0020	25.5823	111.2896	0.3975
4	-0.1413	-2.5849	0.6014	2.3038	4	-0.1187	-1.4695	0.3975	2.3077

Appendix A. An example control file for use with the computer program SPAYK.

```
11 SECTION 1:  AGE CLASS DEFINITIONS ******************************************
The first data record contains the number of possible age classes. One data
record MUST be present for each possible age class. The contents of each data
record must be:
  a. Freshwater age of all possible age classes;
  b. Marine age of all possible age classes;
  c. An indicator of major age classes (1 = major, 0 = minor);
  d. The freshwater age of the surrogate major age to be used to apportion each
     minor age class;
  e. The marine age of the surrogate major age to be used to apportion each
     minor age class.
8
1 1 0 1 3
1 2 0 1 3
1 3 1 1 4
2 2 0 1 3
1 4 1 1 3
2 3 0 1 4
1 5 0 1 4
2 4 0 1 4
15 SECTION 2:  KEY DEFINITIONS FOR DIGITIZED FILES ****************************
This section defines the keys used when digitizing scales. A set of key
definitions MUST be given for each major age, as specified in SECTION 1 above.
The key is a number preceeding a measurement which indicates which growth zone
the measurement was made in. The possible zones are:
  a. First Freshwater zone
  b. Second Freshwater zone
  c. Freshwater plus growth zone
  d. First Ocean zone
  e. Second Ocean zone
  f. Third Ocean zone
  NOTE: If the freshwater age of fish in the file is 1, enter a 0 for the
  second freshwater zone key.
WARNING: ALL FILES CONTAINING DATA FOR A GIVEN AGE CLASS MUST HAVE BEEN
DIGITIZED USING THE SAME ZONE KEYS!
1 3 1 0 2 4 5 6
1 4 1 0 2 4 5 6
5 SECTION 3:  STOCK GROUP INFORMATION *****************************************
The first data record must contain the number of stock groups. Following
records contain a text string naming each stock group. Note: all contribution
estimates will be pooled into the stock groups defined in this section prior to
output. WARNING: STOCK GROUP NAMES ARE LIMITED TO 30 CHARACTERS.
3
Lower
Middle
Upper
26 SECTION 4:  DIGITIZED ESCAPEMENT SAMPLE DATA FILE SECTION *******************
The first data record contains the number of individual samples. One data
record must follow for each individual sample. The contents of each data record
must be:
  a. Sample number
       The sample number MUST be ordered sequentially from 1 to the number of
       samples.
  b. Stock group membership number
       This number MUST be between 1 and the number of stock groups defined in
       SECTION 3. The ordering corresponds to the ordering of the stock group
       names in SECTION 3. The first stock group name listed in SECTION 3 is
       stock number 1, the second stock group name listed in SECTION 3 is stock
       number 2, etc.
  c. Sample location name
       This is a descriptive name associated with the file. Names are limited to
       30 characters in length.
  d. Freshwater age of fish in the file
       All fish in a single file MUST be of the same age.
  e. Ocean age of fish in the file
       All fish in a single file MUST be of the same age.
```

```
  f. File name
       The file name MUST include the path if the data files and the program exe
       file are not in the same directory.
This is the finest level of escapement data. Stock composition estimates are
computed for each file listed in this section. The estimates are pooled into the
stock groups defined in SECTION 3 prior to output.
9
1  1 EFAndreafsky      1 4 Andr1497.dig
2  1 EFAndreafsky      1 3 Andr1397.dig
3  1 Anvik             1 3 Anvk1397.dig
4  1 Anvik             1 4 Anvk1497.dig
5  2 SalchaChena       1 3 ChSa1397.dig
6  2 Chena             1 4 Chna1497.dig
7  2 Salcha            1 4 Salc1497.dig
8  3 Canada            1 3 SrWr1397.dig
9  3 Canada            1 4 SrWr1497.dig
13 SECTION 5:  ESTIMATION OF MEANS AND VARIANCES ******************************
The first record contains an indicator of which estimator to use.
  a.  1 = Use Robust estimators only
  b.  2 = Use Robust estimators unless they fail
  c.  3 = Use traditional estimators.
If the first record is a 3, no other records should be present. If the first
record is a 1 or a 2, two additional records are required. The second record
contains constants used in computing a weight function used by the Robust
estimator; 2.0 and 1.25 are recommended, and these values should be changed with
caution. The third record contains convergence criteria for the Robust
estimator. The first criterion is the maximum number of iterations to take; the
second criterion is the maximum change allowed in successive iterations before
convergence is achieved; recommended values are 50000 and 0.000001.
2
2.0   1.25
50000       0.000001
11 SECTION 6:  VARIABLE SELECTION **********************************************
The first record is the maximum allowable proportion of missing values a
variable may have; any variable with a larger proportion of missing values is
automatically excluded from the model.
The second record is a threshold value for negative correlations. After
variables with the largest F statistics are elected, other variables having a
negative correlation less than the threshold value with a selected variable are
also included in the preliminary variable set. The third record contains the
p-values for F-to-enter and F-to-remove, respectively, for the stepwise variable
selection procedure. WARNING: THE SECOND P-VALUE MUST BE EQUAL TO OR LARGER THAN
THE FIRST P-VALUE.
0.05
-0.80
0.25 0.30
10 SECTION 7:  ESTIMATION ACCURACY SIMULATION **********************************
This section contains control parameters for the 100% simulation to
demonstrate estimation accuracy. The first number is an indicator of how sample
data are to be simulated; a 1 indicates that samples should be drawn from a
multivariate normal distribution with parameters estimated from the observed
sample data, while a 2 indicates bootstrap samples should be drawn from the
observed sample data.  The second number is the number of simulations to
perform.  The third number is the maximum number of function evaluations
allowed while estimating the stock composition of a sample. IF NO SIMULATION IS
DESIRED, CHANGE THE SECOND NUMBER TO 0, AND LEAVE THE OTHER NUMBERS UNCHANGED.
2   0   100000
10 SECTION 8:  STOCK GROUP ESCAPEMENT AGE COMPOSITION **************************
This section contains the estimated age composition for each stock group. A
single record must be present for each stock group defined in the SECTION 2
above, and the rows of age group proportions MUST be given in the same order
as the stock groups are listed in SECTION 2. Within each row, a proportion MUST
be present for each age class defined in SECTION 1, and the ordering of the
proportions MUST correspond to the ordering of the age classes in SECTION 1.
For Yukon River Chinook applications, if the Canadian data are taken from the
fishwheels at the border, the proportions should be adjusted for selectivity
prior to running this program.
0.00170 0.36541 0.25130 0.00063 0.37883 0.00031 0.00182 0.00000
0.00113 0.23618 0.13966 0.00000 0.60886 0.00000 0.01360 0.00057
0.00000 0.00875 0.10000 0.00305 0.73057 0.01944 0.02618 0.11201
```

42 SECTION 9: HARVEST DATA **

This section contains information for all harvests whose stock composition is to be estimated. The first data record is the number of harvests. One 'block' of information is required for each harvest. The first record of a block contains a sample number (which must run from 1 to the number of harvests), strings for Location-District-PeriodNumber(or strata no.)-Date (each of these strings must be less than 30 characters in length), harvest in numbers of fish, and an indicator of how the stock composition is to be estimated. The indicator must be a 1, 2, or 3; 1 indicates maximum likelihood estimation, 2 indicates results from maximum likelihood estimation of other samples are to be used, and 3 indicates the harvest should be assigned to a particular stock group. The second record contains an indicator of whether age composition estimates are available (1=yes, 0=no). If age composition estimates are available, the following line contains the proportions. Note that a proportions must be given for each age listed in SECTION 1. Subsequent input depends on which estimation method was entered on the first record.

Allocation Method 1: Maximum Likelihood Estimation

A record must be present for each major age as defined in SECTION I. Each record must contain the freshwater age, the ocean age, and the name of the file containing the digitized scale data. If no digitized scale data is available for a major age, list the file name as 'NONE'. The age will then be analyzed as a minor age.

Allocation Method 2: Use Other Samples

The next record must be the number of harvest samples to be used to estimate the stock composition of this harvest. The following record contains the sample numbers of those samples, corresponding to the sample number given in the leftmost number in the first record for each harvest. If multiple samples are being used, indicate samples used in the same row with a space in between. The results for the samples listed will be combined, weighted by harvest size, and applied to this sample.

Allocation Method 3: Assign To Stock Group

The next record must be the stock number (from 1 to the number of stocks as listed in SECTION 3, the first name listed is Stock 1, the second name listed is stock 2, etc.) If no age composition estimates are available, the age composition of one or more other samples must be used. In this case, the next record must be the number of harvest samples to be used to estimate the age composition of this harvest. The following record contains the sample numbers of those samples, corresponding to the sample number given in the leftmost number in the first record for each harvest. If multiple samples are being used, indicate samples used in the same row with a space in between. The results for the samples listed will be combined, weighted by harvest size, and applied to this sample.

```
29
1  'Yukon River' 'District 1,' 'Period 1, Unr.' '6/12' Commercial 11369 1
   1
   0.00002 0.02900 0.13100 0.00000 0.81900 0.00000 0.01800 0.00300
   1 3 Y1p11397.dig
   1 4 Y1p11497.dig
2  'Yukon River' 'District 1' 'Period 2, Unr.' '6/17' Commercial 11154 1
   1
   0.00002 0.03200 0.10100 0.00000 0.85100 0.00000 0.01600 0.00000
   1 3 Y1p21397.dig
   1 4 Y1p21497.dig
3  'Yukon River' 'District 1' 'Period 3, Unr.' '6/20' Commercial 20139 1
   1
   0.00000 0.02100 0.11375 0.00000 0.84325 0.00000 0.02200 0.00000
   1 3 Y1p31397.dig
   1 4 Y1p31497.dig
4  'Yukon River' 'District 1' 'Period 5, Unr.' '6/24' Commercial 7394 1
   1
   0.00000 0.06800 0.11000 0.00000 0.80900 0.00000 0.01300 0.00000
   1 3 Y1p41397.dig
   1 4 Y1p41497.dig
5  'Yukon River' 'District 1' 'Period 6, Unr.' '6/27' Commercial 13006 1
   1
   0.00000 0.07100 0.09100 0.00000 0.82710 0.00000 0.01090 0.00000
   1 3 Y1p51397.dig
   1 4 Y1p51497.dig
```

```
6  'Yukon River' 'District 1' 'Period 7, Res.' '6/28' Commercial 770 1
   1
   0.00000 0.32432 0.20270 0.00000 0.44595 0.00000 0.02703 0.00000
   1 3 Y1p61397.dig
   1 4 Y1p61497.dig
7  'Yukon River' 'District 1' 'Period 8, Res.' '6/30' Commercial 477 1
   1
   0.00000 0.32877 0.16438 0.00000 0.50685 0.00000 0.00000 0.00000
   1 3 None
   1 4 Y1p71497.dig
8  'Yukon River' 'District 1' 'Period 4, Res.' '6/22' Commercial 2075 2
   0
   2
   6 7
9  'Yukon River' 'District 1' 'Subsistence' 'Season Total' ' ' 7550 2
   0
   1
   1
10  'Yukon River' 'District 1' 'Test Fish' 'Season Total' ' ' 2811 2
   0
   1
   1
11  'Yukon River' 'District 2' 'Period 1, Unr.' '6/16' Commercial 7266 1
   1
   0.00000 0.02013 0.10403 0.00000 0.86913 0.00000 0.00671 0.00000
   1 3 Y2p11397.dig
   1 4 Y2p11497.dig
12 'Yukon River' 'District 2' 'Period 2, Unr.' '6/19' Commercial 9583 1
   1
   0.00000 0.04422 0.16327 0.00000 0.78231 0.00000 0.00680 0.00340
   1 3 Y2p21397.dig
   1 4 Y2p21497.dig
13 'Yukon River' 'District 2' 'Period 3, Unr.' '6/23' Commercial 15248 1
   1
   0.00000 0.03679 0.09365 0.00000 0.85953 0.00000 0.01003 0.00000
   1 3 Y2p31397.dig
   1 4 Y2p31497.dig
14 'Yukon River' 'District 2' 'Period 4, Res.' '6/25' Commercial 311 2
   0
   2
   6 7
15 'Yukon River' 'District 2' 'Period 5, Unr.' '7/1-7/2' Commercial 6955 1
   1
   0.00000 0.05630 0.09920 0.00000 0.82574 0.00000 0.01877 0.00000
   1 3 Y2p51397.dig
   1 4 Y2p51497.dig
16 'Yukon River' 'District 2' 'Subsistence' 'Season Total' ' ' 9350 2
   0
   1
   11
17 'Yukon River' 'District 3' 'Subsistence' 'Season Total' ' ' 6311 2
   0
   2
   1 11
18 'Yukon River' 'District 4B' 'All Periods' 'Season Total' 'Gillnet Harvests' 50 2
   1
   0.00000 0.03125 0.12500 0.00000 0.84375 0.00000 0.00000 0.00000
   5
   11 12 13 14 15
19 'Yukon River' 'District 4C' 'All Periods' 'Season Total' 'Gillnet Harvests' 495 2
   1
   0.00000 0.00000 0.08046 0.00000 0.90805 0.00000 0.01149 0.00000
   5
   11 12 13 14 15
20 'Yukon River' 'District 4' 'All Periods' 'Season Total' 'Fishwheel Harvests' 912 2
   1
   0.00000 0.14670 0.24956 0.00000 0.59664 0.00000 0.00710 0.00000
   5
   11 12 13 14 15
```

```
21 'Yukon River' 'District 4' 'Subsistence' 'Season Total' ' ' 11415 2
   0
   3
   18 19 20
22 'Yukon River' 'District 5' 'All Periods' 'Season Total' 'Fishwheel Harvests' 2026 3
   1
   0.00000 0.16667 0.19697 0.00000 0.58081 0.00505 0.02525 0.02525
   3
23 'Yukon River' 'District 5' 'All Periods' 'Season Total' 'Gillnet Harvests' 1652 3
   1
   0.00000 0.02069 0.15172 0.00000 0.77241 0.00690 0.02069 0.02759
   3
24 'Yukon River' 'District 5' 'Subsistence' 'Season Total' ' ' 17735 3
   0
   3
   2
   22 23
25 'Yukon River' 'District 6' 'All Periods' 'Season Total' 'Commercial Fishwheel' 2728 3
   1
   0.00000 0.56646 0.18701 0.00000 0.24652 0.00000 0.00000 0.00000
   2
26 'Yukon River' 'District 6' 'Subsistence' 'Season Total' ' ' 3930 3
   0
   2
   1
   25
27 'Yukon River' 'District 6' 'Sport Fish' 'Season Total' ' ' 2017 3
   1
   0.00113 0.23618 0.13966 0.00000 0.60886 0.00000 0.01360 0.00057
   2
28 'Yukon River' 'Canada' 'All Periods' 'Season Total' 'Commercial' 5311 3
   0
   3
   2
   22 23
29 'Yukon River' 'Canada' 'All Periods' 'Season Total' 'Non-Commercial' 11217 3
   0
   3
   2
   22 23
```

Appendix B. An example output file, corresponding to the control file in Appendix A, created by the computer program SPAYK.

```
Program SPAYK

Control File: 97RobBootB.ctl

Date: 19981130
Time: 152137.440


Reformating escapement data files...
      36 fish processed in file Andr1397.dig.
      42 fish processed in file Anvk1397.dig.
      55 fish processed in file ChSa1397.dig.
     150 fish processed in file SrWr1397.dig.
      97 fish processed in file Andr1497.dig.
      82 fish processed in file Anvk1497.dig.
     139 fish processed in file Chna1497.dig.
      48 fish processed in file Salc1497.dig.
     311 fish processed in file SrWr1497.dig.


Performing preliminary variable screening for age 1.3.
    Variable       Between Sample
    Selected        F-Statistic
        1             11.45
        5             10.01
        8              9.83
       10              6.04
       12              9.10
       14              8.12
       18             11.40
       21              9.00
       23              7.29
       25              9.90
       27             16.93
       28             10.28
       30              5.76
       62             67.37
       65             38.58
       67             75.64
       71             10.86
       74             10.32
       78             11.13
       81             10.70
       88             15.01
       92             20.65
       96             22.33
       99             18.41
      104              4.71
      105              5.00
      106             16.09
      107              6.83
      109              3.92
      111             12.39
```

Performing preliminary variable screening for age 1.4.

Variable Selected	Between Sample F-Statistic
2	17.39
5	15.65
8	13.41
10	9.20
13	10.52
14	17.17
16	15.72
20	9.73
23	9.73
26	23.41
27	14.06
28	11.15
29	6.86
61	32.15
65	38.45
68	19.52
70	4.54
72	18.89
79	10.20
82	8.43
85	7.78
90	8.22
97	10.73
100	10.75
103	4.87
105	8.35
106	4.19
107	2.55
109	2.25
111	3.59

Adding variables having large negative correlations with selected variables for age 1.3.

Adding Variable 19. It has a correlation of -0.965 with Variable 1
Adding Variable 94. It has a correlation of -0.920 with Variable 88
Adding Variable 93. It has a correlation of -0.958 with Variable 106

Adding variables having large negative correlations with selected variables for age 1.4.

Adding Variable 18. It has a correlation of -0.925 with Variable 28
Adding Variable 94. It has a correlation of -0.941 with Variable 70

Performing final variable selection from preliminary variable set for age 1.3.

Variable 67 added with an F-to-enter p-value of 0.0000 < 0.2500.
Variable 96 added with an F-to-enter p-value of 0.0130 < 0.2500.
Variable 1 added with an F-to-enter p-value of 0.0012 < 0.2500.

```
Performing final variable selection from preliminary variable set for age 1.4.
    Variable  65 added with an F-to-enter p-value of 0.0000 < 0.2500.
    Variable  72 added with an F-to-enter p-value of 0.0000 < 0.2500.
    Variable   8 added with an F-to-enter p-value of 0.0000 < 0.2500.
    Variable  68 added with an F-to-enter p-value of 0.0000 < 0.2500.
    Variable   5 added with an F-to-enter p-value of 0.0000 < 0.2500.
    Variable  10 added with an F-to-enter p-value of 0.0002 < 0.2500.
    Variable  85 added with an F-to-enter p-value of 0.0011 < 0.2500.
    Variable  82 added with an F-to-enter p-value of 0.0000 < 0.2500.
    Variable  26 added with an F-to-enter p-value of 0.0000 < 0.2500.
    Variable  18 added with an F-to-enter p-value of 0.0003 < 0.2500.
    Variable 103 added with an F-to-enter p-value of 0.0522 < 0.2500.
    Variable 100 added with an F-to-enter p-value of 0.0623 < 0.2500.
    Variable  90 added with an F-to-enter p-value of 0.0001 < 0.2500.
    Variable 107 added with an F-to-enter p-value of 0.0000 < 0.2500.

Stock groups are identified in the output as follows:
     1 = Lower
     2 = Middle
     3 = Upper

Processing data for harvest sample    1:
    Yukon River
    District 1,
    Period 1, Unr.
    6/12
    Commercial

      26 fish processed in file Y1p11397.dig.
      26 fish had complete records and were used in the analysis.
```

Estimated Stock Composition For Major Age 1.3.

Stock Group	Estimate	Standard Error
1	0.2118	0.1701
2	0.0041	0.2217
3	0.7841	0.2244

```
     167 fish processed in file Y1p11497.dig.
     167 fish had complete records and were used in the analysis.
```

Estimated Stock Composition For Major Age 1.4.

Stock Group	Estimate	Standard Error
1	0.3510	0.0601
2	0.1841	0.0467
3	0.4650	0.0657

Stock by age estimates:

Stock Group	Age 1.1	1.2	1.3	2.2	1.4	2.3	1.5	2.4
1	0	265	315	0	3268	0	15	0
2	0	6	6	0	1714	0	37	0
3	0	59	1168	0	4330	0	152	34

Processing data for harvest sample 2:
Yukon River
District 1
Period 2, Unr.
6/17
Commercial

22 fish processed in file Y1p21397.dig.
22 fish had complete records and were used in the analysis.

Estimated Stock Composition For Major Age 1.3.

Stock Group	Estimate	Standard Error
1	0.2078	0.1608
2	0.1958	0.2600
3	0.5964	0.2295

191 fish processed in file Y1p21497.dig.
190 fish had complete records and were used in the analysis.

Estimated Stock Composition For Major Age 1.4.

Stock Group	Estimate	Standard Error
1	0.3143	0.0543
2	0.2553	0.0496
3	0.4303	0.0601

Stock by age estimates:

Stock Group	Age 1.1	1.2	1.3	2.2	1.4	2.3	1.5	2.4
1	0	157	234	0	2984	0	12	0
2	0	172	221	0	2423	0	45	0
3	0	27	672	0	4085	0	122	0

Processing data for harvest sample 3:
Yukon River
District 1
Period 3, Unr.
6/20
Commercial

17 fish processed in file Y1p31397.dig.
17 fish had complete records and were used in the analysis.

Estimated Stock Composition For Major Age 1.3.

Stock Group	Estimate	Standard Error
1	0.2378	0.4891
2	0.0758	0.2062
3	0.6863	0.2736

192 fish processed in file Y1p31497.dig.
190 fish had complete records and were used in the analysis.

Estimated Stock Composition For Major Age 1.4.

Stock Group	Estimate	Standard Error
1	0.3821	0.0571
2	0.2351	0.0477
3	0.3827	0.0555

Stock by age estimates:

Stock Group	Age 1.1	1.2	1.3	2.2	1.4	2.3	1.5	2.4
1	0	274	545	0	6489	0	39	0
2	0	102	174	0	3993	0	112	0
3	0	48	1572	0	6499	0	292	0

Processing data for harvest sample 4:
Yukon River
District 1
Period 5, Unr.
6/24
Commercial

13 fish processed in file Y1p41397.dig.
13 fish had complete records and were used in the analysis.

Estimated Stock Composition For Major Age 1.3.

Stock Group	Estimate	Standard Error
1	0.6157	0.3629
2	0.0000	0.7601
3	0.3843	0.2070

181 fish processed in file Y1p41497.dig.
180 fish had complete records and were used in the analysis.

Estimated Stock Composition For Major Age 1.4.

Stock Group	Estimate	Standard Error
1	0.3463	0.0567
2	0.1721	0.0443
3	0.4816	0.0627

Stock by age estimates:

Stock Group	Age 1.1	1.2	1.3	2.2	1.4	2.3	1.5	2.4
1	0	485	501	0	2071	0	7	0
2	0	0	0	0	1029	0	16	0
3	0	18	313	0	2881	0	73	0

```
Processing data for harvest sample     5:
    Yukon River
    District 1
    Period 6, Unr.
    6/27
    Commercial

      15 fish processed in file Y1p51397.dig.
      15 fish had complete records and were used in the analysis.

        Estimated Stock Composition For Major Age 1.3.
            Stock           Standard
            Group  Estimate   Error
              1     0.6133    0.3611
              2     0.1435    0.3049
              3     0.2432    0.2585

     181 fish processed in file Y1p51497.dig.
     181 fish had complete records and were used in the analysis.

        Estimated Stock Composition For Major Age 1.4.
            Stock           Standard
            Group  Estimate   Error
              1     0.2584    0.0512
              2     0.2247    0.0480
              3     0.5168    0.0618


        Stock by age estimates:
          Stock  Age
          Group     1.1    1.2    1.3    2.2    1.4    2.3    1.5    2.4
            1         0    712    726      0   2780      0      7      0
            2         0    194    170      0   2417      0     29      0
            3         0     17    288      0   5560      0    106      0

Processing data for harvest sample     6:
    Yukon River
    District 1
    Period 7, Res.
    6/28
    Commercial

      11 fish processed in file Y1p61397.dig.
      11 fish had complete records and were used in the analysis.

        Estimated Stock Composition For Major Age 1.3.
            Stock           Standard
            Group  Estimate   Error
              1     0.8751    0.4236
              2     0.0000    0.5509
              3     0.1249    0.1271
```

20 fish processed in file Y1p61497.dig.
20 fish had complete records and were used in the analysis.

Estimated Stock Composition For Major Age 1.4.

Stock Group	Estimate	Standard Error
1	0.4535	0.1919
2	0.0582	0.1283
3	0.4883	0.1864

Stock by age estimates:

Stock Group	Age 1.1	1.2	1.3	2.2	1.4	2.3	1.5	2.4
1	0	248	137	0	156	0	2	0
2	0	0	0	0	20	0	1	0
3	0	2	19	0	168	0	17	0

Processing data for harvest sample 7:
Yukon River
District 1
Period 8, Res.
6/30
Commercial

21 fish processed in file Y1p71497.dig.
21 fish had complete records and were used in the analysis.

Estimated Stock Composition For Major Age 1.4.

Stock Group	Estimate	Standard Error
1	0.2255	0.1948
2	0.1811	2.4826
3	0.5934	0.1925

Stock by age estimates:

Stock Group	Age 1.1	1.2	1.3	2.2	1.4	2.3	1.5	2.4
1	0	116	43	0	55	0	0	0
2	0	37	12	0	44	0	0	0
3	0	4	23	0	143	0	0	0

Processing data for harvest sample 8:
Yukon River
District 1
Period 4, Res.
6/22
Commercial

Stock and age composition estimates taken from harvest samples: 6 7

Stock by age estimates:

Stock Group	Age 1.1	1.2	1.3	2.2	1.4	2.3	1.5	2.4
1	0	606	300	0	351	0	3	0
2	0	62	20	0	106	0	2	0
3	0	10	70	0	517	0	28	0

Processing data for harvest sample 9:
Yukon River
District 1
Subsistence
Season Total

Stock and age composition estimates taken from harvest samples: 1

Stock by age estimates:

Stock Group	Age 1.1	1.2	1.3	2.2	1.4	2.3	1.5	2.4
1	0	176	209	0	2170	0	10	0
2	0	4	4	0	1138	0	25	0
3	0	39	776	0	2875	0	101	23

Processing data for harvest sample 10:
Yukon River
District 1
Test Fish
Season Total

Stock and age composition estimates taken from harvest samples: 1

Stock by age estimates:

Stock Group	Age 1.1	1.2	1.3	2.2	1.4	2.3	1.5	2.4
1	0	65	78	0	808	0	4	0
2	0	1	1	0	424	0	9	0
3	0	15	289	0	1071	0	38	8

Processing data for harvest sample 11:
Yukon River
District 2
Period 1, Unr.
6/16
Commercial

14 fish processed in file Y2p11397.dig.
14 fish had complete records and were used in the analysis.

Estimated Stock Composition For Major Age 1.3.

Stock Group	Estimate	Standard Error
1	0.0000	0.8783
2	0.0000	0.6040
3	1.0000	0.4081

142 fish processed in file Y2p11497.dig.
142 fish had complete records and were used in the analysis.

Estimated Stock Composition For Major Age 1.4.

Stock Group	Estimate	Standard Error
1	0.2924	0.0600
2	0.1199	0.0480
3	0.5876	0.0785

Stock by age estimates:

Stock Group	Age 1.1	1.2	1.3	2.2	1.4	2.3	1.5	2.4
1	0	0	0	0	1847	0	3	0
2	0	0	0	0	757	0	5	0
3	0	146	756	0	3711	0	41	0

Processing data for harvest sample 12:
Yukon River
District 2
Period 2, Unr.
6/19
Commercial

28 fish processed in file Y2p21397.dig.
28 fish had complete records and were used in the analysis.

Estimated Stock Composition For Major Age 1.3.

Stock Group	Estimate	Standard Error
1	0.1284	0.1257
2	0.0000	0.2245
3	0.8716	0.2272

133 fish processed in file Y2p21497.dig.
132 fish had complete records and were used in the analysis.

Estimated Stock Composition For Major Age 1.4.

Stock Group	Estimate	Standard Error
1	0.2668	0.0588
2	0.1512	0.0452
3	0.5821	0.0760

Stock by age estimates:

Stock Group	Age 1.1	1.2	1.3	2.2	1.4	2.3	1.5	2.4
1	0	301	201	0	2000	0	3	0
2	0	0	0	0	1133	0	9	0
3	0	123	1364	0	4364	0	53	32

Processing data for harvest sample 13:
Yukon River
District 2
Period 3, Unr.
6/23
Commercial

17 fish processed in file Y2p31397.dig.
17 fish had complete records and were used in the analysis.

Estimated Stock Composition For Major Age 1.3.

Stock Group	Estimate	Standard Error
1	0.5101	0.3924
2	0.0000	0.6317
3	0.4899	0.2160

141 fish processed in file Y2p31497.dig.
140 fish had complete records and were used in the analysis.

Estimated Stock Composition For Major Age 1.4.

Stock Group	Estimate	Standard Error
1	0.4326	0.0693
2	0.1759	0.0533
3	0.3915	0.0648

Stock by age estimates:

Stock Group	Age 1.1	1.2	1.3	2.2	1.4	2.3	1.5	2.4
1	0	530	728	0	5670	0	16	0
2	0	0	0	0	2306	0	30	0
3	0	31	700	0	5130	0	107	0

```
Processing data for harvest sample   14:
    Yukon River
    District 2
    Period 4, Res.
    6/25
    Commercial

    Stock and age composition estimates taken from harvest samples:    6   7

        Stock by age estimates:
          Stock  Age
          Group    1.1    1.2    1.3    2.2    1.4    2.3    1.5    2.4
            1        0     91     45      0     53      0      1      0
            2        0      9      3      0     16      0      0      0
            3        0      1     10      0     78      0      4      0


Processing data for harvest sample   15:
    Yukon River
    District 2
    Period 5, Unr.
    7/1-7/2
    Commercial

      20 fish processed in file Y2p51397.dig.
      19 fish had complete records and were used in the analysis.

        Estimated Stock Composition For Major Age 1.3.
            Stock          Standard
            Group  Estimate   Error
              1     0.3325   0.1758
              2     0.0000   0.2517
              3     0.6675   0.2296


     153 fish processed in file Y2p51497.dig.
     153 fish had complete records and were used in the analysis.

        Estimated Stock Composition For Major Age 1.4.
            Stock          Standard
            Group  Estimate   Error
              1     0.3180   0.0599
              2     0.1870   0.0718
              3     0.4950   0.0677


        Stock by age estimates:
          Stock  Age
          Group    1.1    1.2    1.3    2.2    1.4    2.3    1.5    2.4
            1        0    349    229      0   1826      0      9      0
            2        0      0      0      0   1074      0     23      0
            3        0     42    461      0   2843      0     99      0
```

```
Processing data for harvest sample   16:
    Yukon River
    District 2
    Subsistence
    Season Total


    Stock and age composition estimates taken from harvest samples:   11

        Stock by age estimates:
          Stock  Age
          Group    1.1    1.2    1.3    2.2    1.4    2.3    1.5    2.4
            1        0      0      0      0   2377      0      4      0
            2        0      0      0      0    974      0      6      0
            3        0    188    973      0   4775      0     53      0

Processing data for harvest sample   17:
    Yukon River
    District 3
    Subsistence
    Season Total


    Stock and age composition estimates taken from harvest samples:    1  11

        Stock by age estimates:
          Stock  Age
          Group    1.1    1.2    1.3    2.2    1.4    2.3    1.5    2.4
            1        0     90    107      0   1732      0      6      0
            2        0      2      2      0    837      0     14      0
            3        0     69    652      0   2723      0     65     12

Processing data for harvest sample   18:
    Yukon River
    District 4B
    All Periods
    Season Total
    Gillnet Harvests

    Stock composition estimates taken from harvest samples:   11  12  13  14  15

        Stock by age estimates:
          Stock  Age
          Group    1.1    1.2    1.3    2.2    1.4    2.3    1.5    2.4
            1        0      1      2      0     15      0      0      0
            2        0      0      0      0      7      0      0      0
            3        0      0      4      0     21      0      0      0
```

```
Processing data for harvest sample   19:
    Yukon River
    District 4C
    All Periods
    Season Total
    Gillnet Harvests

    Stock composition estimates taken from harvest samples:   11  12  13  14  15

        Stock by age estimates:
          Stock  Age
          Group    1.1     1.2     1.3     2.2     1.4     2.3     1.5     2.4
            1        0       0      11       0     156       0       1       0
            2        0       0       0       0      72       0       1       0
            3        0       0      29       0     221       0       4       0

Processing data for harvest sample   20:
    Yukon River
    District 4
    All Periods
    Season Total
    Fishwheel Harvests

    Stock composition estimates taken from harvest samples:   11  12  13  14  15

        Stock by age estimates:
          Stock  Age
          Group    1.1     1.2     1.3     2.2     1.4     2.3     1.5     2.4
            1        0     105      61       0     189       0       0       0
            2        0       1       0       0      88       0       1       0
            3        0      28     167       0     267       0       5       0

Processing data for harvest sample   21:
    Yukon River
    District 4
    Subsistence
    Season Total

    Stock and age composition estimates taken from harvest samples:   18  19  20

        Stock by age estimates:
          Stock  Age
          Group    1.1     1.2     1.3     2.2     1.4     2.3     1.5     2.4
            1        0     830     580       0    2820       0       8       0
            2        0       8       0       0    1308       0      16       0
            3        0     219    1567       0    3988       0      71       0
```

Processing data for harvest sample 22:
Yukon River
District 5
All Periods
Season Total
Fishwheel Harvests

Stock by age estimates:

Stock Group	Age 1.1	1.2	1.3	2.2	1.4	2.3	1.5	2.4
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	338	399	0	1177	10	51	51

Processing data for harvest sample 23:
Yukon River
District 5
All Periods
Season Total
Gillnet Harvests

Stock by age estimates:

Stock Group	Age 1.1	1.2	1.3	2.2	1.4	2.3	1.5	2.4
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	34	251	0	1276	11	34	46

Processing data for harvest sample 24:
Yukon River
District 5
Subsistence
Season Total

Age composition estimates taken from harvest samples: 22 23

Stock by age estimates:

Stock Group	Age 1.1	1.2	1.3	2.2	1.4	2.3	1.5	2.4
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	1794	3134	0	11828	101	410	468

Processing data for harvest sample 25:
Yukon River
District 6
All Periods
Season Total
Commercial Fishwheel

Stock by age estimates:

Stock Group	Age 1.1	1.2	1.3	2.2	1.4	2.3	1.5	2.4
1	0	0	0	0	0	0	0	0
2	0	1545	510	0	673	0	0	0
3	0	0	0	0	0	0	0	0

Processing data for harvest sample 26:
Yukon River
District 6
Subsistence
Season Total

Age composition estimates taken from harvest samples: 25

Stock by age estimates:

Stock Group	Age 1.1	1.2	1.3	2.2	1.4	2.3	1.5	2.4
1	0	0	0	0	0	0	0	0
2	0	2226	735	0	969	0	0	0
3	0	0	0	0	0	0	0	0

Processing data for harvest sample 27:
Yukon River
District 6
Sport Fish
Season Total

Stock by age estimates:

Stock Group	Age 1.1	1.2	1.3	2.2	1.4	2.3	1.5	2.4
1	0	0	0	0	0	0	0	0
2	2	476	282	0	1228	0	28	1
3	0	0	0	0	0	0	0	0

Processing data for harvest sample 28:
Yukon River
Canada
All Periods
Season Total
Commercial

Age composition estimates taken from harvest samples: 22 23

Stock by age estimates:

Stock Group	Age 1.1	1.2	1.3	2.2	1.4	2.3	1.5	2.4
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	537	939	0	3542	30	123	140

```
Processing data for harvest sample   29:
    Yukon River
    Canada
    All Periods
    Season Total
    Non-Commercial

    Age composition estimates taken from harvest samples:   22  23

        Stock by age estimates:
          Stock  Age
          Group    1.1    1.2    1.3    2.2    1.4    2.3    1.5    2.4
            1        0      0      0      0      0      0      0      0
            2        0      0      0      0      0      0      0      0
            3        0   1135   1982      0   7481     64    259    296

Program completed at 153558.180.
Execution time =   14.35 minutes.
Normal program termination...
```

Appendix C. An example of the data file format required by the computer program SPAYK.

```
andr1392.dig   Combined East and West Fork data files (adef1392.dig,adwf1392.dig)
33420  00 71924113  434 8 0 66511 381 121 151 151 101 101  81  71 101  7     0
1  91  51  62 143 173 223 173 113 203 213 193 183 223 183 273 223 203 233 183 14
3 233 183 273 193 253 223 153 133 143 173 143 274420
33420  00 71924113  414 8 0 84021 381 131 151  91 101  91 101  51  91  7     0
2 122  93 153 103 163 153 123 173 183 163 183 193 173 203 133 143 233 233 213 22
3 303 213 193 383 213 173 183 143 233 234435
33420  00 639241130 434 8 0 65011 311  91 161  91  91  61  31  91  61  3     0
1  62 102 112 112  72  93  73 153  83 133 113 123 103 113 143  93 163 193 193 18
3 153 123 153 173 103 213 173 143 113 113  83 154387
33420  00 639241130 424 8 0 67511 411 111 141 141  81  51  91  61  71  8     0
1  52  92 112 122 112 113 133 153 163 113 173 193 163 193 153 153 173 193 173 21
3 173 193 223 233 153 213 203 103 243 163 184391
33420  00 639241130 414 8 0 74511 391 171 141  81 141  41  41  81  71  7     0
1 102  92 152 132  92 143 113 153 193 203 233 173 193 183 203 173 153 113 183 20
3 183 183 243 233 213 183 173 183 173 114382
33420  00 639241130 374 8 0 65011 361 141 121  71  91  71  81  91  61 12     0
2  82 122  92 113 103 153 163 103 193 163 183 153 193 153 123 173 163 193 263 18
3 253 243 153 203 143 124322
33420  00 639241130 434 8 0 69011 331 131  81  91 101  61 101  61  82 14     0
2 102  72 122 113 133  83 153  93 183 143 163 133 153 143 173 113 163 153 163 13
3 173 183 133 173 173 183 173 123 143 173 163 144325
33420  00 639241130 414 8 0 68511 331 111  61 111  51  91  81 101 101  8     0
2  82  82  83 133 153 113 143 173 133 163 163 183 163 173 163 153 133 193 153 10
3 183 163 223 153  93 163 153 103 143 134393
```