FRI-UW-8713
September 1987

VARIABLE SELECTION AND PERFORMANCE OF VARIABLE SUBSETS IN SCALE PATTERN ANALYSIS

by

Nancy D. Davis

Submitted to

International North Pacific Fisheries Commission
by the
United States National Section

September 1987

FISHERIES RESEARCH INSTITUTE
School of Fisheries
College of Ocean and Fishery Sciences
University of Washington
Seattle, Washington 98195

THIS PAPER MAY BE CITED IN THE FOLLOWING MANNER:

Davis, Nancy D. 1987. Variable selection and performance of variable subsets in scale pattern analysis. (Document submitted to annual meeting of the INPFC 1987.) 47 pp. FRI-UW-8713. Fisheries Research Institute, University of Washington, Seattle.

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

Scale features are used in International North Pacific Fisheries Commission studies to determine origins of high seas-caught Pacific salmon (*Oncorhychus* spp.). Results of these studies may vary dependent upon the scale characters used in classification models, but the method for character selection and the number and types of characters used has not been standardized. Scale character data from chinook salmon (*O. tshawytscha*) originating in Asia, western Alaska, central Alaska and southeastern Alaska and British Columbia were used to examine variable selection and performance of variable subsets. Two character selection criteria, Wilks' Lambda and Mahalanobis distance, were compared on the basis of correct reclassification of the scales to their respective groups (i.e., classification accuracy). Subsets selected by these criteria were different. Once highly non-normal characters and collinear characters were removed from character sets, and non-ratio and ratio characters were divided into separate character sets, both selection criteria selected identical subsets. Overall classification accuracy decreased by 1% to 2% with the reduction of the variable sets from 48 to 11 characters. Characters that used life history information resulted in overall classification accuracies ranging 5.9% to 9.0% higher than characters that did not. Overall classification accuracy decreased slightly by 0.3% to 1.3% when non-discriminating characters were forced into classification models. The use of a reduced character set is recommended because of consistency in subset selection and retention of adequate classification accuracy. When reduction in the number of variables is necessary, the stepwise procedure using the Wilks' Lambda selection criterion provided an efficient and repeatable selection methodology, but variable selection may not be required with small character sets that adequately characterize scale patterns.

VARIABLE SELECTION AND PERFORMANCE OF VARIABLE SUBSETS IN SCALE PATTERN ANALYSIS

INTRODUCTION

Scale characters are used in International North Pacific Fisheries Commission (INPFC) studies to determine origins of high seas-caught Pacific salmon (*Oncorhynchus* spp.). Often more characters are measured and calculated than are necessary for estimation of stock proportions, so character subsets that have nearly the same information as the full character set are selected. Methodology for character subset selection and the number and types of characters have not been standardized among researchers using scale pattern data for stock separation. Myers (1986) found that classification results vary dependent upon the scale characters used in the analysis. The INPFC mandate to achieve greater consistency, standardization, and validity in scale pattern studies provided the impetus for this study on character selection methodology.

The objectives of my research were to find character sets that provide consistent subsets, to compare the relative efficiency of several types of characters in describing scale patterns, and to investigate whether inclusion of poor discriminators results in lower classification accuracies. In this paper I also review character selection procedures used in previous INPFC-related salmon scale pattern analyses, and recommend variable selection procedures and character sets for use in scale pattern analysis.

Variable Selection Methods Used and Variable Sets Investigated in Previous INPFC-Related Salmon Scale Pattern Analyses

Three primary methods, *a priori* selection of characters, the Kruskal-Wallis 'H' statistic (Kruskal and Wallis 1952) and Wilks' Lambda (Wilks 1932) have been used by INPFC researchers for variable selection. Studies that have investigated the performance of variable sets include a comparison of scale characters and their factor scores (Bilton and Messinger 1975), characters least sensitive to the body area of scale removal (Knudsen 1985; Knudsen and Davis 1985), and the effect on classification accuracies of character sets selected by the *a priori* and Wilks' Lambda methods (Myers 1986).

Measuring scale features with rulers is tedious and time consuming and severely limits the number of characters that can be measured. In this case, investigators usually opt for *a priori* selection of characters. That is, characters are selected before the scales are measured, based on experience of which variables are likely to be good discriminators. The characters measured are often few in number, and there is little option for creation of character subsets.

Scale pattern analyses with *a priori* selection of scale characters have been conducted by INPFC researchers for many years. Anas and Murai (1969) used three sockeye (*O. nerka*) scale characters to calculate linear discriminant functions (LDF, Fisher 1936) and quadratic discriminant functions. Two types of scale characters, circuli counts (CC) and marine sextuplets (distances between six circuli increments) were measured (see Table 1 for codes and explanations of scale character types). Kato and Ishida (1985, 1986) measured and used nine coho (*O. kisutch*) scale characters (three character types, CC, large distances

[LD], and a marine circuli triplet [MT]) to calculate a LDF. Ishida et al. (1985) studied chum distribution with five characters (two character types, CC and LD) and two characters (CC), respectively, in two polynomial discriminant functions. Major et al. (1975, 1977a, 1977b) selected three chinook (*O. tshawytscha*) scale characters that included three character types (CC, LD and the distance between the 12th and last circuli in the marine zone) for a LDF. Ito et al. (1985, 1986) used nine chinook scale characters (four character types, CC, LD, freshwater triplets [FWT] and MT) in a LDF.

Two early studies did not conform exactly to the *a priori* selection method. Amos et al. (1963) measured ten meristic characters.including three scale characters from pink salmon (*O. gorbuscha*) Scale characters included two character types, CC and LD. A LDF calculated with the ten characters showed that three variables provided 98% separation (between groups). The other seven characters were dropped from the analysis and the LDF was recalculated from the three characters and used to classify high seas samples. By not using all the characters in the LDF, this study differed from the *a priori* method. The study of Tanaka et al. (1969) also did not conform exactly to the a priori method. They compared the distributional ranges of 13 chum (*O. keta*) scale characters, including three character types (CC, LD, average circulus spacing [ACS]). The 13 characters were reduced to eight when it was found that characters in the third and fourth year of growth varied more with sex, age at maturity, and brood year than area of origin. Instead, only characters from the first and second years of growth were used to classify high seas samples.

Computer based digitizing equipment makes data collection of scale measurements less time consuming and enables researchers to measure more scales and more scale features. Brood year or stock specific subset models can be selected from a large suite of scale characters. The problem becomes one of objectively selecting a character subset that provides accurate separation between groups.

The Kruskal-Wallis 'H' statistic is a nonparametric test for intergroup differences (Kruskal and Wallis 1952), and was described by Cook and Lord (1978) as a method for character selection. In this technique, each character is ranked over (classification) groups, the difference between the average sum of ranks for each pairwise group combination is calculated, and scale characters providing the 'best' univariate separation of groups are selected. Highly correlated variables are not used. Cook and Lord selected a subset of six characters from the 11 measured (three character types: CC, LD and the width of the widest circulus) for calculation of a polynomial function. The character subset was limited to six to keep the number of terms in the polynomial function to a 'reasonable' number (210 terms). Marshall et al. (1978, 1979) measured 10 sockeye scale characters that included three character types, LD, CC and MT, and selected six characters for each brood year specific model. Cook et al. (1980) used 16 sockeye scale characters and Cook et al. (1981) and Cook (1982) used 19 scale characters from which six characters were selected for each brood year analysis. Myers et al. (1981) calculated 40 coho scale characters that included five character types, CC, LD, MT, marine triplet ratios (MTR) and large distance ratios (LDR). All 40 characters were screened with the Kruskal-Wallis 'H' statistic to determine the best subset model. Walker and Harris (1982) calculated 48 coho scale characters that included six character types, CC, LD, LDR, MTR, ACS, and marine nonuplets (MN). Walker and Davis (1983) calculated 60 coho scale characters that included seven character types, CC, LD, LDR, MT, MTR, ACS and MN, and screened the total for selection of a subset. Ishida et al. (1984) measured nine chum scale characters (LD, CC and circuli quintuplets), and selected five characters for use in a polynomial discriminant function. Knudsen et al. (1983) used polynomial discriminant analysis for a four group stock separation study of chinook. They calculated 60 scale characters, screened these characters (10 character types, CC, LD, LDR, MT, MTR, ACS, MN, FWT, freshwater triplet ratios

[FWTR], and radius of the focus) with the Kruskal-Wallis 'H' statistic, and selected 3-6 variables for each analysis.

In recent INPFC-scale pattern analyses, there has been a return to the parametric approach to character selection and use of LDF analysis. Conrad (1984), who measured chum scales screened 96 variables by examining group F-statistics and the correlation coefficients between all pairs of characters. Characters considered for inclusion in the LDF either had a large F- value or were negatively correlated with characters having large F-values. The resulting variable subset had five scale characters.

Wilks' Lambda is a multivariate statistic which is an overall measure of group separation based on the differences between all the group centroids (Wilks 1932). Wilks' Lambda is inversely related to the overall multivariate F-ratio (Rao 1952), and for the purposes of variable selection, the Wilks' Lambda and F statistic are equivalent (Habbema and Hermans 1977). Stepwise selection with Wilks' Lambda as the character selection criterion was used in investigations of chinook salmon stock origins (Myers et al. 1984, in press; Myers and Rogers 1985a; Myers 1985, 1986). In these studies, 48 scale characters were screened by the stepwise algorithm to develop brood year and stock specific models. Ten character types were used in these analyses, CC, LD, LDR, MT, FWT, MN, ACS, FWT, MTR, and circuli count ratios (CCR). The stepwise selection procedure, BMDP7M (Dixon et al.1983), selected 2-20 variables for each analysis.

Performance of variable subsets was first studied by Bilton and Messinger (1975). They measured eight sockeye scale characters that included two character types, CC and LD. They compared the classification accuracies from a LDF calculated with scale characters to accuracies from a LDF calculated with factor scores. The factor scores provided less accurate discrimination than the original characters.

Knudsen (1985) and Knudsen and Davis (1985) examined chinook scale character variability attributable to body area of scale removal. Knudsen compared scale character types across five body zones and found that LDR and CCR were the least sensitive to body zone effects. The overall classification accuracy of standards in a LDF with only LDR and CCR characters dropped relative to non-ratio models and ratio models by approximately 5%. Classification accuracies of the unknown scales, however, were generally equal to or greater than those obtained with either non-ratio or ratio characters. Comparisons of non-ratio and ratio characters showed that both did equally well at separating groups, with ratio data yielding a 2% improvement in overall classification accuracy. Accuracies of individual body zone groups varied.

Myers (1986) also examined the performance of character subsets by comparing the classification accuracies of the same group of chinook scales measured and classified with two different character sets: a brood year specific character set selected by Wilks' Lambda (Myers et al. 1984) and an *a priori* character set (Ito et al. 1985). Myers found overall accuracies were somewhat higher with the characters of Myers et al. than with those of Ito et al., but these differences were probably not significant. The classification accuracies of the Asian standards were often higher for the Ito et al. character set than for the Myers et al. character set, but the classification accuracies of the western and central Alaska standards were sometimes dramatically higher with the character set of Myers et al. This study clearly indicated that classification results vary dependent upon the scale characters, and pointed to the need for a character selection methodology that provides consistent character subsets and, ultimately, the most accurate classifications.

METHODS

Scale Pattern Data

I used chinook salmon scale pattern data collected by Myers et al. (1984). Within this large database are composite standards of major inshore stocks from seven brood years (1971-1977) and four regions: Asia (ASIA), western Alaska (WEST), central Alaska (CENT), and southeast Alaska and British Columbia (SEBC). The weighting of brood year standards and scale feature measurements are described by Myers et al. (1984). I selected a subsample from these data, brood year 1973A (BY73) and 1977 (BY77) standards, for my study. These brood years were selected because they represent the extreme range in classification accuracies of the regional analyses of Myers et al. (1984). BY73 had low overall accuracy (71.7%), low accuracy for ASIA (58.5%), and unequal sample sizes per standard: ASIA (n=118), WEST (n=198), CENT (n=134) and SEBC (n=194). BY77 had high overall accuracy (80.0%), high accuracy for ASIA (87.0%), and nearly equal sample sizes per standard: ASIA (n=200), WEST (n=199) CENT (n=199) and SEBC (n=198).

A detailed description of chinook characters and character types are listed in Table 1. A summary list of the categories of character types is also shown in Table 1. The categories of character types are grouped into three types of measures (direct measures, ratios and rates). Direct measures are those characters quantified by simple observation of the scale. Direct measure characters include the width across a life history zone or combination of zones (large distance, LD), the number of circuli in a zone or combination of zones (circuli counts, CC), and small distance measures like circuli doublets (distance across two circuli increments). Similarly, other small distance measures include: circuli triplets (distance across three increments), circuli quadruplets (across four circuli increments), and circuli quintuplets (across five circuli increments). Ratios and rates are another group of scale characters. Ratio characters are a proportion that can be expressed as a percentage, so that the units in the numerator and the denominator cancel each other leaving a dimensionless quantity. Ratios can be formed from either circuli counts or distance measures. Rates are also quotients, but with different units in the numerator and denominator so units do not cancel, e.g., average circulus spacing (ACS, Table 1). A list of character set names and characters that are included in these character sets are listed in Table 2.

Selection Criteria, Stopping Rules, and Classification Accuracies

A stepwise procedure and two character selection criteria were used to investigate consistency in selection of character subsets. Subprogram Discriminant in Statistical Package for the Social Sciences (SPSS, Nie et al. 1975) and BMDP7M program (Dixon 1983) were used for computation of all LDF analyses. Both programs allow stepwise entry of variables into the LDF, but options in these programs do not completely overlap: SPSS provides several methods of variable selection and BMDP provides one.

I used two variable selection methods in SPSS, Wilks' Lambda and the Mahalanobis distance (Mahalanobis 1936). Mahalanobis distance is a generalized measure of distance between two groups. When used as a variable selection procedure, the Mahalanobis distance between all pairs of groups is calculated for each character. Then, for each character, the distance between the two closest groups is listed. From the list of distances between closest groups, the character having the largest distance is selected for entry into the model. The Wilks' Lambda selection criterion is a multivariate statistic which measures separation between all centroids, and is inversely related to the distribution of the overall F-ratio (Rao 1952). The variable selection algorithm selects the variable with the largest F-ratio (i.e., the smallest Wilks' Lambda).

Either selection criteria, Wilks' Lambda or Mahalanobis distance, determines which character is the next candidate for entry in the model. Other constraints called stopping rules must be imposed, otherwise all variables would eventually be entered. The next character selected by the selection criteria must also pass two requirements before entry into the model. The first requirement is a minimum F-to-enter value. The F-to-enter value used in this study was 4.000. The second requirement for entry into the model is a minimum tolerance level. A character with a small tolerance level indicates large interdependency between variables that causes difficulty inverting the covariance matrix and can result in rounding errors in calculation of discriminant coefficients. Minimum tolerance levels of .001 and .010 were used in this study. The final stopping rule is the F-to-remove criterion. With this rule, if the F-value of a character already in the model drops below a minimum value (3.996 in my analyses), then it is removed from the model. If no character meets either the entry or removal criteria, then character selection is terminated. The advantage of using a specific stepwise selection criteria and stopping rule is the objectivity of the method: two researchers starting with the same data would develop identical subset models.

The performance of character subsets was evaluated on the basis of overall correct classification. Classification accuracy compares the group to which a case is classified to the actual group to which the case belongs. Each case is classified to the most likely group based on its discriminant score (posterior probability). The overall correct classification was calculated by taking the proportion of correct classifications for each group, summing the proportions and dividing by the number of groups (four) to obtain the unweighted average.

All cases are used to calculate the classification accuracies in the SPSS analyses, so these accuracies are biased. The data were not divided into learning and training samples, because the amount of information available would be diminished and accuracies might vary depending on how the data were divided.

BMDP7M produces jackknifed classification accuracies that provide a nearly unbiased estimate of misclassification. The program uses a leaving-one-out approach, whereby one observation is removed at a time and the LDF is calculated from the remaining cases. The LDF is then used to classify that observation. Results of classifying each case in this manner are summarized in a jackknife classification matrix. This technique makes efficient use of data by eliminating the need to split data into learning and training data sets.

A total of four selection methods were used. The purpose of the first two methods, the Wilks' Lambda (W) and the Mahalanobis method (M), were to compare the difference in selected subsets between two different selection criteria. Classification accuracies were biased, but since these two methods were used only to compare selection of subsets, it was unimportant that accuracies were biased. However, the third and fourth methods, Wilks' Lambda jackknife (W/J), where the selection criterion was to minimize Wilks' Lambda, and the forced jackknife (F/J), where all characters entered the model after passing the minimum specified tolerance level, were used to compare the ability of several types of character sets to discriminate. The classification accuracies were used for performance evaluation of character sets, and for this reason, jackknife classification accuracies were used.

Determination of Consistency in Subset Selection

Character Set Size and Selection Criteria

The stepwise procedure recalculates the value of the selection criteria after entry of each character into the model. As a result, the likelihood of selection of a particular character may vary according to the characters already in the model. For this reason, two selection criteria might diverge at one step in the selection process and, therefore, ultimately select two different models. With a large number of characters, divergence in character set selection would be more likely to occur. To address this problem, and also to determine which characters would be selected by the two criteria, the W/J, W, and M methods were used to select subset models from the large suite of characters used by Myers et al. (1984) and an *a priori* selected character set used by Ito et al. (1985). Character subsets were compared to determine consistency in subset selection between these three methods.

Character Sets with Transformed, Non-ratio, and Ratio Characters

New character sets were formed to investigate whether the characters had an influence on the consistency of selected subsets. One of the assumptions of the LDF is multivariate normal distribution of the data. Normal probability plots (BMDP5D, Dixon et al. 1983) and Kolmogorov-Smirnov goodness of fit tests (subprogram NPAR, SPSSX, Anon. 1986) were used to examine how well the individual characters fit a normal distribution. A character set, Char1 (n=27, Table 2), was formed as a subset of the character set of Myers et al. (1984) by dropping characters that were highly non-normal. Plots of the distributions of these highly non-normal characters resembled an 'L' shape because some scales had zero values while other scales had positive values for the variable. Highly non-normal characters were found among the FWT, FWTR, MT, and MTR character types because occasionally triplets did not exist on the scale. For example, if there are seven freshwater circuli on a particular scale, then the FWT characters, C41, C42, C43 and the FWTR characters, C46, C47, and C48 do not exist (Table 1). Other scales in the standard may have more freshwater circuli and, therefore, have values for these characters. A plot of the distribution of this character for all scales from a particular standard will be 'L' shaped. Characters with 'L' shaped distributions are referred to as highly non-normal characters. Other characters which failed the goodness of fit test ($p<0.10$) were candidates for log, square root, or arcsin transformations. If the transformation normalized the distribution of a character in all four standards or normalized one standard while not adversely affecting the others, then the transformed character was used in an alternative character set, Char2 (n=27, Table 2). More character sets were created by removing collinear characters from Char1 and Char2 (e.g., use C1 and C5, but not also C6) and by dividing the remaining characters into non-ratio (Char1A and Char2A, n=11, Table 2) and ratio (Char1B and Char2B, n=12, Table 2) character sets. Char 2, Char2A and Char2B contained the transformed homologous characters of Char1, Char1A and Char1B, respectively.

A character set with some collinear characters (Char1) and character sets composed of direct measures (Char1A) and ratio and rate characters (Char1B) were screened by two selection criteria, W and M, to determine consistency in subset selection. Char2, Char2A and Char2B were also screened by these selection criteria to determine consistency in subset selection when some characters have been transformed. Two tolerance levels, .010 and .001 were used to examine whether the tolerance level influenced the subsets selected. All of these character sets were also screened by the W/J method to evaluate their relative ability to separate groups. The F/J method was used to force as many characters as possible from Char1 (tolerance level equal to .001), Char1A (tolerance level equal to .010), and Char1B (tolerance level equal to .001) into the model. The F/J classification accuracies were compared to the W/J accuracies to determine if inclusion of non-selected characters had an effect on overall classification accuracies.

Determination of Effective Discriminators

Discriminant Information from Direct Measures and Rates

Discriminant information contributed by several character types was determined by subdividing Char1A into freshwater and marine triplets (Char3A, Table 2), CC and triplets (Char3B, Table 2) and LD and triplets (Char3C, Table 2). The W/J selection criterion was used to select character subsets from Char3A, Char3B and Char3C and the resulting classification accuracies were compared to determine the contribution of direct measure characters, triplets, CC and LD, towards separation of groups. The W and M criteria were used to check on consistency in subset selection.

The rate character, ACS, was calculated from two direct measures, LD divided by CC. To determine if there was a loss of classification accuracy in subsets selected from character sets with ACS characters instead of direct measures, rate character sets were created. Rate character sets included freshwater and marine triplets in addition to freshwater ACS (Char4A, Table 2), marine ACS (Char4B, Table 2), overall ACS (Char4C, Table 2) and the combined characters, freshwater ACS and marine ACS (Char4D, Table 2) and freshwater ACS, marine ACS and overall ACS (Char4E, Table 2). The W/J selection criterion was used to select character subsets from Char4A through Char4E, and the resulting classification accuracies were compared to determine the contribution of the freshwater, marine, and overall ACS characters towards separation of groups. The W and M criteria were used to check on consistency in subset selection. The stopping rules were the same as used for Char3A through Char3C. The classification accuracies of Char1A and Char4D were compared to determine the contribution of direct measure and rate characters discrimination between groups.

Characters Without Life History Information

To see if identification of scale life history zones affected classification accuracies, character sets with circuli triplets (Char5A, Table 2), triplets and overall CC (Char5B, Table 2), triplets and overall LD (Char5C, Table 2), and triplets and overall CC and LD combined (Char5D, Table 2) were devised. These triplets (T, Table 1) were taken consecutively from the focus out to the last circulus in the first ocean annulus without differentiating between freshwater and marine growth. However, overall circuli count and size are not completely independent of life history interpretation, as scale pattern data were collected out to the edge of the ocean annulus. The W/J criterion was used to select character subsets from Char5A through Char5D, and the resulting classification accuracies were compared to determine the contribution of triplets, CC and LD characters toward the separation of groups. The W and M criteria were used to check on consistency in subset selection. The classification accuracy of Char1A, a character set that includes triplets (FWT and MT) dependent upon consistent life history interpretation, was compared with Char5D to evaluate the value of using characters which require life history information.

Small Distance Measures

Triplets provide information on circuli spacing in small areas of the scale and are intended to 'smooth' the increments of individual circuli on a particular fish scale in order to better characterize the stock. To investigate whether the triplet is a good choice for a smoothing distance, character sets with circuli doublets (Char6A, Table 2), quadruplets (Char6B, Table 2) and quintuplets (Char6C, Table 2) were developed. The W/J criterion was used to select character sets from Char6A through Char6C, and the resulting classification accuracies were compared to determine the relative contribution of the three smoothing distances towards separation of groups. The W and M criteria were used to select subsets from Char6A through Char6C as a check on consistency in subset selection. Char1A includes circuli triplets and was compared with the character set (Char6A, Char6B or Char6C) that resulted in the highest classification accuracy.

Inclusion of Poor Discriminators in Models

A priori selection of scale characters as variables for LDF presumes the variables are good discriminators. Non-discriminating dummy characters were created to examine the effect of forcing them into models. The dummy characters were uncorrelated and randomly drawn from a normal population (mean equal to 370 and standard deviation equal to 70). These values for the mean and standard deviation were chosen because they are similar to actual data for circuli triplets. The dummy variables were forced into discriminant models with Char1A, a character set known to be adequate at discriminating groups. Char7A (Table 2) included Char1A and one dummy character, Char7B (Table 2) included Char1A and two dummy characters, Char7C (Table 2) included Char1A and three dummy characters and Char7D (Table 2) included Char1A and four dummy characters. To examine the effects on classification accuracies of dummy characters when they composed a large portion of the model, Char1A was subdivided into smaller components, CC and LD (Char7E, Table 2) and triplets (Char7G, Table 2). The four dummy characters were combined with CC and LD (Char7F, Table 2) and with triplets (Char7H, Table 2). All the characters were forced into models subject to the minimum tolerance level equal to .010. Resulting classification accuracies were compared to determine the effect of inclusion of known non-discriminating characters in the model.

RESULTS

Determination of Consistency in Subset Selection

Effects of Character Set Size and Variable Selection Criteria

Results of the W and M selection criteria on the character sets of Myers et al. (1984) and Ito et al. (1985) are shown in Table 3, Parts A (BY73) and B (BY77). The results for W/J (A1) and W (A2) are different because for the subset selected by W/J, as reported by Myers et al. (1984), the stepwise process was terminated at step 13, before the stopping rules were invoked. The subset selected by W resulted from letting the selection process continue until no more variables were entered or removed from the model within the constraints of the stopping rule (step 16). Character C35 was dropped from the model after C58 entered, and subsequently C1 and C39 entered the model before the stepwise procedure was completed (Table 3, A2). Since characters can be entered or removed from the model at each step, stopping the stepwise procedure at different points can result in selection of divergent subsets.

For BY73, the W criterion selected a different subset than the M criterion (Table 3, A2 and A3). The M method selected two more characters. Both methods selected the full model from Ito et al.'s (1985) *a priori* character set (Table 3 A4 and A5). Although Ito et al.'s variables are a subset of Myers et al.'s (Table 2), few of the same characters were selected (Table 3, A2 and A5). Ito et al.'s set yielded an overall accuracy of 69.2%, a 2.5 % decrease from the 71.7% overall accuracy obtained with Myers et al.'s set (Table 3, A4 and A1).

For BY77, Myers et al. (1984) allowed the stepwise procedure to continue until no more characters were entered or removed, and, therefore, the same subset was selected by both the W/J and the W methods (Table 3, B1 and B2). Although both methods selected the same character set, the overall classification accuracy (average percentage of scales correctly classified) for W/J was 1.2% lower than for W, because the jackknifing procedure reduces bias in the model. The M method selected five more characters, but classification accuracies are only 0.8% higher than the W selected subset model (Table 3, B2 and B3). Both methods selected the full model from Ito et al.'s (1985) set. The overall W/J classification accuracy of Myers et al.'s set was 80.0%, and the overall W/J classification of Ito et al.'s set was 78.1%, a decrease of 1.9% (Table 3, B1 and B4).

Classification Accuracies with Collinear, Non-Ratio, Ratio, and Transformed Characters

Once the highly non-normal characters were dropped from the Myers et al (1984) data set, the remaining characters were generally normally distributed. Of the 27 characters for each of the four standards (total=108 distributions for each brood year) in Char1 (Table 2), 5.6% (BY73) and 10.2% (BY77) failed the Kolmogorov-Smirnov test at $p<.05$. The distributions of characters which failed the test were slightly right or left skewed or leptokurtic or both.

Results of character selection and classification accuracies using a large character set (Char1, Myers et al.'s [1984] set less highly non-normal characters), and two character sets with collinear characters removed, non-ratio (Char1A), and ratio (Char1B) are shown in Table 4. In addition to the W/J, W, and M methods, an attempt was made to force the full model at either tolerance equal to .001 or tolerance equal to .010. For BY73, the subset selected from the large character set by W/J was different from that selected by W since the tolerance level was lowered from .010 to .001 (Table 4, A1 and A2). The tolerance level was lowered to examine its effects on the order characters were entered, and the final model selected. The W method selected a different subset model from the large character set than the M method (Table 4, A2 and A3). An unsuccessful attempt was made to force all of the characters into the model at tolerance equal to .001 (Table 4, A4). The F/J model contains 10 more characters than the W/J model, yet the overall accuracy of the F/J model was only 1.0% higher than the accuracy of the W/J model (Table 4, A1).

For BY73, W/J, W, and M criteria all selected the same subset model from the non-ratio character set (Table 4, A5 and A6). The full model was forced at tolerance level equal to .010. The overall accuracies of the W/J and F/J models were similar (Table 4, A5 and A7). Ratio characters were more affected by tolerance level than non-ratio characters. The W/J subset at tolerance equal to .010 was different than the W and M methods at tolerance equal to .001 (Table 4, A8 and A9). The full model of ratio characters was successfully forced at the lower tolerance level (.001). The F/J model had an overall accuracy of 1% less than the W/J model (Table 4, A8 and A10). For models with characters selected by W/J, the models with non-ratio characters yielded 3% higher accuracies than those with ratio characters (Table 4, A5 and A8).

For BY77, W/J, W and M criteria selected different subset models from the large character set (Table 4, B1-B3). The F/J was not able to force all characters at a tolerance level of .001 (Table 4, B4). The forced model included 13 more characters than the W/J subset, but a slightly lower overall classification accuracy (Table 4, B1). The W/J, W, and M criteria selected the full model from the non-ratio character set (Table 4, B5 and B6).

The W/J selected different subsets from the ratio character set than the W and M criteria at the lower tolerance level (Table 4, B7 and B8). The lower tolerance level was required to force all the characters into the F/J model (Table 4, B9). For models with characters selected by W/J, the models with non-ratio characters yielded 1.2% higher accuracies than those with ratio characters (Table 4 B5 and B7).

Classification accuracies obtained with character sets that included several transformed characters are shown in Table 5 for both brood year analyses. Results were similar to those obtained with character sets that did not include transformed characters (Table 4).

Determination of Effective Discriminators

Discriminant Information from Direct Measures and Rates

Char1A is composed of one FWT, six MT, and freshwater and marine CC and LD (Table 2). Classification accuracies of models developed from these component character types are in Table 6, Parts A (BY73) and B (BY77). Tolerance level was constant (.010), and the W and M criteria consistently selected the same subset models from all character sets. For BY73, triplets (FWT and MT) yielded an overall accuracy of 56.8% (Table 6, A1). When CC were combined with triplets overall accuracy improved to 60.5% (Table 6, A3). When LD characters were combined with triplets, overall accuracy decreased to 56.3% and was approximately equal to the accuracy obtained with triplets alone (Table 6, A5). When CC and LD were combined with triplets, overall accuracy increased to 69.7% (Table 6, A7). Effects of these changes on individual standards varied.

For BY77, W and M criteria selected identical subsets. The triplet model produced an overall accuracy of 55.5% (Table 6, B1). When CC were combined with triplets, overall accuracy increased to 67.1% (Table 6, B3). When LD characters were combined with triplets, overall accuracy was 68.2% (Table 6, B5). When CC and LD were combined with triplets accuracy increased to 79.0% (Table 6, B7).

Accuracies for WEST, CENT and SEBC were higher in both brood years when CC and LD were used together in the character set than when either character was used separately. The accuracies obtained when CC and LD were used separately were similar, and data from triplets were important for classification of SEBC (Table 6, A1 and B1).

Classification accuracies from character sets with average circuli spacing (ACS) are shown in Table 7, Parts A (BY73) and B (BY77). W and M criteria selected identical character subsets. For BY73, overall classification accuracy obtained with freshwater ACS was 60.4%, and accuracy was slightly higher with ocean ACS (61.0%), but decreased to 59.7% with overall ACS (Table 7, A1, A3 and A5). Accuracy increased to 66.0% when freshwater and ocean ACS were combined (Table 7, A7).

For BY77, overall classification accuracy was approximately equal for freshwater ACS, ocean ACS, or overall ACS (63.4%, 64.1%, 63.7%, respectively) (Table 7, B1, B3, and B5). When freshwater ACS and ocean ACS were combined accuracies increased to 70.7% (Table 7, B7).

Characters Without Life History Information

The results from character sets with triplets (T, Char5A), T and overall CC (Char5B), T and overall LD (Char5C), and T and overall CC and LD combined (Char5D) are shown in Table 8, Parts A (BY73) and B (BY77). The BY73 character sets included eight T characters and the BY77 character sets included nine T characters. For all character sets, the W and M criteria selected identical subsets within the constraints of the stopping rule. For BY73, overall classification accuracies obtained when T characters were used were 42.5%, and increased to 48.9% when CC were added to T characters (Table 8, A1 and A3). Recall that overall CC and LD do contain some life history information. Although not dependent upon differentiation between freshwater and marine growth, these characters do require consistent identification of the end of the measurement zone, that is, the edge of the second winter annulus (Table 1). Accuracy was slightly lower (45.4%) when overall LD was combined with T (Table 8, A5). Accuracy increased to 60.7% when overall CC and LD were combined with T (Table 8, A7).

For BY77, overall classification accuracy obtained when T characters were used was 54.1%, and accuracies increased approximately the same amount (62.8% and 63.4%) when CC and T and LD and T were used (Table 8, B1, B3 and B5). Accuracy was higher (73.1%) when CC and LD were combined (Table 8, B7).

Classification accuracies obtained from triplets with (FWT and MT) and without (T) life history information was determined by comparing the classification accuracies of Char5D and Char1A (Tables 4 and 8). The use of Char5D did not result in accuracies as high as those obtained with Char1A. Overall accuracy of BY73 decreased 9.0% and overall accuracy of BY77 decreased 5.9%. In BY73, accuracies for ASIA, WEST, CENT and SEBC decreased by 0.8%, 10.6%, 16.4% and 8.3%, respectively. In BY77, accuracies for CENT decreased by 16.0%, and accuracies for SEBC decreased by 7.5%. Accuracies for ASIA and WEST (BY77) remained the same with both character sets.

Small Distance Measures

Results of combining circuli increments into small distances such as circuli doublets (Char6A), circuli quadruplets (Char6B), and circuli quintuplets (Char6C) are shown in Table 9. The W/J criterion selected a subset model from Char6A (Table 9, A1 and B1) and selected full models from Char 6B and Char6C (Table 9, A3, A4, B3, and B4). Overall accuracy for BY73 was highest (72.1%) for circuli quadruplets (Table 9, A3). This result was slightly higher than overall accuracy of circuli triplets (69.7%, Char1A, Table 4, A5). The highest accuracy for BY77 was achieved with circuli triplets (79.0%, Char1A, Table 4, B5), only slightly higher than accuracies achieved with circuli doublets (78.8%, Char6A, Table 9, B1).

Inclusion of Poor Discriminators

Forcing unselected characters into models caused small changes in overall classification accuracies as well as shifts of the accuracies in the standards (Char1, Table 4, A4 and B4, Char5D, Table 8, A9 and B9, and Char6A, Table 9, A2 and B2). Classification accuracies resulting from character sets with dummy characters are presented in Table 10. There were small changes in accuracies when non-discriminating characters were forced into the models. For BY73, overall accuracy for Char7D (Char1A with four dummy characters, Table 10, A4) decreased 1% from the overall accuracy of Char1A (Table 4, A5). For BY77, there were smaller differences between the overall accuracies of Char7D (Table 10, B4) and Char1A. Overall accuracy in BY77 for Char7D was 0.1% higher than the overall accuracy for Char1A (Table 4, B5). Char7E (CC and LD, n=4) and Char7G (FWT and MT, n=7) contained few characters and discriminated poorly (Table 10, A5, A7, B5, and B7). Four dummy characters were forced into Char7E and Char7G (Char7F and Char7H, Table 2). For BY73, the CC and LD model (Char7E) decreased in overall classification accuracy from 59.2% to 58.7% when four dummy characters were forced into the model, and accuracy for ASIA decreased by 2.6% (Table 10, A5 and A6). By forcing dummy characters, the triplet model (Char7G) decreased in overall accuracy from 55.9% to 54.6%

and accuracy for ASIA decreased by 1.7% (Table 10, A7 and A8). For BY77, the CC and LD model (Char7E) decreased in overall accuracy 73.0% to 72.0% when dummy characters were forced into the model, and accuracy for CENT decreased by 4.0% (Table 10, B5 and B6). Triplet models (Char7G) decreased in overall accuracy from 54.8% to 53.5% by forcing dummy characters, and accuracy for CENT decreased by 2.5% (Table 10, B7 and B8).

DISCUSSION

Consistency in Subset Selection

Character Set Size and Variable Selection Criteria

A stepwise procedure that uses either the W or M selection criteria is a repeatable method of variable subset selection. These two criteria however, will not always select the same subset models from the large character sets: Myers et al. (1984), Char1, and Char2 (Tables 3, 4, and 5). The W and M criteria differ in that the W is an overall measure of group separation, while the M method uses the distance between the two closest groups. The selection of distinct subsets may be an outcome of the two contrasting approaches and the large number of characters from which to select. The M method often took more steps to derive the same variable subset as the W method. The M criterion often entered a character, removed it and, possibly, entered it again at a later step. For this reason, its behavior was less stable than the W criterion.

The W and M criteria did select identical subsets when characters were selected from character sets with fewer variables (Char1A, Char1B, Char2A and Char2B; Tables 4 and 5). The nature of these character sets are discussed in the following section.

Character Sets with Collinear, Non-Ratio, Ratio, and Transformed Characters

The presence of collinear characters in the same character set can make the outcome of a stepwise procedure more variable. Classification models selected from character sets with many variables, like Char1 and Char2 (Table 2) were sensitive to both selection criteria (W or M), and to tolerance level (Tables 4 and 5). Once highly non-normal characters and collinear characters were removed, and the non-ratio and ratio characters were separated into different character sets (Char1A and Char1B, Table 2), the W and M criteria selected identical subsets (Table 4, A6, A9, B6, and B8; Table 5, A5, A7, B5, and B7). However, characters were not necessarily selected in the same order or in the same number of steps. Results from this study show there was no increase in classification accuracies when transformed characters were used to calculate the LDF (Char2, Char2A, Char2B, Table 5). The LDF appears robust to the non-normality of scale characters for the sample sizes used in this study. The large effort required to find transformations to normalize the distribution of a character in all four standards, or normalize one standard while not adversely effecting the others, does not appear to be necessary.

Reduction in character set size did not seriously decrease classification accuracies. High overall classification accuracies for non-ratio models (Char1A), 69.7% (BY73), and 79.0% (BY77), and ratio models (Char1B), 66.7% (BY73) and 77.8% (BY77), indicated that removal of collinear and highly non-normal characters and separation of non-ratio and ratio characters did not decrease the ability to discriminate among groups (Table 4, A5, A8, B5, and B7). Results showed that overall classification accuracy decreased by 2% (BY73) and 1% (BY77) with the reduction of the variable sets from 48 (Myers et al. 1984) to 11 (Char1A) characters (Tables 3 and 4).

Use of non-ratio characters (Char1A) resulted in accuracies 3.0% (BY73) and 1.2% (BY77) higher than were obtained with ratio characters (Char1B) (Table 4). In contrast, Knudsen and Davis (1985) found that classification models with ratio data had approximately 2% higher overall accuracy than models with non-ratio characters. However, ACS characters in Knudsen and Davis (1985) were included in the non-ratio character set, whereas ACS characters in my study were included in the ratio character set. There is further difficulty comparing these studies since structure of the standards was different. Knudsen and Davis (1985) studied specific hatchery stocks, and I used composite standards of major inshore stocks. In general a 2% difference in overall classification accuracies will probably not have a significant effect on classification results, so that the researcher can take other factors into consideration when deciding whether to use ratio or non-ratio characters.

Several authors have expressed reservations regarding ratios in statistical analyses that assume the multivariate normal distribution of variables (Reyment et al. 1984, Atchley et al. 1976, Humphries et al. 1981). Ratio data are often not normally distributed. The distributions of ratio characters in Myers' et al. (1984) data were less normal than the distributions of non-ratio characters. Seventy-five percent of the characters in the BY77 data set that were transformed were ratio characters. The intent of my study is not to take a definitive position on the use of ratio or non-ratio data in scale pattern analyses, but the use of non-ratio and their ratio homologous characters (e.g., C5 and C11) in the same character set is redundant.

Determination of Effective Discriminators

Discriminant Information from Direct Measures and Rates

Char1A consisted of several types of direct measures, FWT, MT, CC, and LD (Table 2), that my results indicated were useful in discriminating among groups (Table 6). However, classification accuracy gained with each particular character type varied by group and brood year. For example, triplets were valuable in separating SEBC and ASIA from WEST and CENT, CC combined with triplets provided higher accuracy for CENT (BY73 only) and SEBC, and LD and triplets provided higher accuracies for WEST and CENT (BY77 only). The types and number of characters needed to achieve adequate separation, depend not only on the scale patterns of the standards, but also the number of standards.

Char1A can be thought of as an extension of Ito et al.'s (1985) character set. The CC and LD characters in both character sets are the same. Ito et al.'s set is composed of one FWT and three MT characters. Char1A has these same four triplets, plus three more MT characters that Ito et al.'s character set does not contain (i.e., C52, C53, C54, Table 2). The number of triplets are limited by the number of circuli in a zone. C54 was the furthest complete MT from the focus towards the edge of the measurement zone that was present on all scales in the standards. It is likely however, that the number of complete freshwater and marine triplets could vary from one brood year to another. All the complete FWT and MT available in a particular data set should be included in a character set like Char1A from which subset models will be selected.

The combination of LD and CC characters into rate characters, ACS, decreased the ability of the LDF to discriminate between groups. Overall classification accuracies obtained with character sets that included ACS characters (Char4D, Table 2) were 3.7% (BY73) and 8.3% (BY77) lower than the overall accuracy obtained with Char1A, which did not include any rate characters (Tables 4 and 7). The LD and CC characters appear to be more effective discriminators by themselves than when combined into rate characters.

Characters Without Life History Information

Character sets that included triplets with life history information (Char3A, FWT, MT) provided 14.3% (BY73) and 1.4% (BY77) higher overall classification accuracies than character sets with triplets that did not include this information (Char5A, T, Tables 6 and 8). Character sets with freshwater characters, FWT, CC, and LD and marine characters, MT, CC and LD (Char1A) provided 9.0% (BY73) and 5.9% (BY77) higher overall accuracies than triplets (T), and overall CC and LD that did not distinguish between freshwater and marine growth (Char5D, Tables 4 and 8). Consistent interpretation of life history patterns is important because the pattern should be measured the same way each time it is encountered. Otherwise, the scale feature is assigned to the incorrect zone, and many inconsistent interpretations may result in poor characterization of the standard. Ensuring consistency in recognition of life history patterns increases the time required for measuring scales, and this process can be difficult when several scale readers are required to collect data (Myers and Rogers 1985b). Characters that do not contain life history information would not require this effort. However, the increased classification accuracies from character sets with life history information show that these characters are important for discrimination between groups, and are worth the effort involved in their collection.

Small Distance Measures

There were only small differences in overall accuracies among the small distance measures that I tested (circuli doublets, triplets, quadruplets and quintuplets) (Tables 4 and 9). The type of small distance measure that yielded the highest accuracy was brood year dependent. The highest accuracy in BY73 was achieved with quadruplet circuli distances (Table 9). Brood year 1977 showed triplet distances had slightly higher overall accuracy than the other distances. My results suggest that the circuli triplet is an adequate small distance measure. It is unlikely that one could find a single small distance measure that provides maximum accuracy for all standards, especially with a multiple brood year analysis.

Inclusion of Poor Discriminators in Classification Models

A priori models may include characters that would not have otherwise been included by a selection criterion. Historically, forcing non-selected characters into *a priori* models was likely not a problem because the models contained so few characters that they probably all provided some discrimination. My results showed that when characters are forced into models there can be a slight decrease in accuracy with adjustment up and down among the accuracies of the individual standards (Tables 4, 8, and 9). In some cases, forcing all characters in the model was computationally difficult, especially when character sets contained large number of variables (Table 4, A4, B4). If the number of scale characters is small, for example, 11 rather than 27 (Char1A and Char1, respectively), then the whole character set may be used to compute the LDF. Forcing the entire Char1A character set (BY73) decreased the overall accuracy only slightly (0.2%) (Table 4, A7).

Introduction of non-discriminating (dummy) characters into discriminant models had a very slight depressing effect on accuracies (Table 10). This small effect was least when the four dummy variables were forced into a model that already had a substantial number of characters and good discrimination (e.g. Char1A) (Table 10, A1-A4, B1-B4). The decrease in classification accuracy that occurred when non-selected characters were forced into a model was so slight that subset selection may not be required with small character sets that adequately characterize scale patterns. There was a larger effect when the four dummy characters comprised half of the variables and the real characters did not provide enough information for adequate separation (Table 10, A5-A8, B5-B8).

Summary

1. Character sets that provided consistent subsets with the Wilks' Lambda and Mahalanobis selection criteria were Ito et al.'s (1985) character set, the non-ratio character set, Char1A, and the ratio character set, Char1B (Tables 3 and 4). The character sets containing some tranformed characters, Char2A and Char2B, also provided identical subsets with the Wilks' Lambda and Mahalanobis selection methods (Table 5).

2. The relative efficiency of several types of characters were investigated. I found that direct measure characters (circuli counts and large distances) were more effective at describing scale features than rate characters (average circulus spacing, Tables 6 and 7). Characters that depend on consistent identification of life history information were more effective than characters that do not contain this information (Tables 6 and 8). The triplet small distance measure is an adequate interval to describe circuli spacing in small areas of the scale (Tables 4 and 8).

3. Forced inclusion of characters not selected by a character selection criterion or non-discriminating dummy variables, did not seriously reduce the ability of LDF models to separate groups (Tables 4, 8, and 10).

Recommendations

1. If reduction in the number of variables is desired, a stepwise procedure using either the Wilks' Lambda or Mahalanobis selection criteria should be used. A stepwise procedure with a specified selection criterion and stopping rules is an objective methodology for character selection. This procedure also provides a method for selection of characters that can be reproduced. Both the Wilks' Lambda and Mahalanobis criteria produced identical subsets when selecting from character sets with few variables (Tables 6-8), but if it is necessary to choose between these two methods, I recommend the Wilks criterion. The drawback to the Mahalanobis method is that it often takes more steps to derive the same variable subset as the Wilks method. The Mahalanobis criterion often enters a character, removes it, and possibly enters it again at a later step. For this reason, its behavior is less stable than the Wilks criterion. The merit of the Wilks criterion (F-ratio) is its widespread use in previous scale pattern studies, and that the F-to-enter and F-to-remove criteria are widely used in other statistical analyses requiring variable selection.

2. A variable selection procedure may not be necessary at all, if the character set includes a small number of characters that adequately describe the patterns of circuli on scales. An *a priori* selection method is likely to produce similar accuracies in models developed by a stepwise procedure if a character set without collinear and highly non-normal characters (for example, Char1A) is used (Tables 2 and 4). *A priori* selection may result in the inclusion of some non-discriminating variables in the classification models, but my results suggest that the LDF is fairly robust when non-discriminating characters are included. However, classification accuracies were slightly lower when all characters were forced into

models than when the Wilks' Lambda criterion was employed (Tables 4, 8, and 9). This slight depressing effect may be greater in models with low accuracies (Table 10).

3. I recommend Char1A as an appropriate beginning character set for use with either stepwise or *a priori* selection methods. Among the character sets used in this study, Char1A provided consistent subsets with two different selection criteria, and supplied enough discriminant information for adequate separation of groups (Table 4, A5-A7, and B5 and B6).

a. The reduction from 48 to 11 characters by removal of collinear and highly non-normal characters and separation of non-ratio and ratio characters did not decrease the ability to discriminate between groups (Char1A, Table 4).

b. The direct measure characters included in Char1A are easier to interpret biologically, and provided higher classification accuracies than ratio character sets (Tables 4 and 5). Ratio characters may be less sensitive to body zone effects than non-ratio characters (Knudsen 1985). Depending upon scale sample quality, it might be necessary to use ratio characters as a precaution when sample quality is doubtful. However, it is preferable to eliminate from an analysis scales that are nonpreferred.

c. Characters that depend upon consistent identification of life history patterns were found to provide higher classification accuracies than characters that did not (Table 9). Char1A contains life history information with freshwater characters: CC, LD and FWT and marine characters: CC, LD and MT. Consistent interpretation of life history zones is difficult, especially in stocks as diverse as ASIA and SEBC. If scale readers document their interpretations of life history patterns, and consult the documentation throughout the period of data collection, then the ability of finding characters with consistent life history information may be increased. If more scale character information is required for discrimination, it may be useful to isolate more life history zones. Life history patterns that can be consistently recognized by scale readers, perhaps freshwater plus growth, marine annuli or summer growth, may be good candidates for additional characters, but caution should be exercised not to inadvertently reacquire redundant characters.

d. The triplet based on life history information was an adequate choice for describing small distances on scales (Table 9). For both brood years, Char1A included seven triplets. Only complete triplets were used, that is, triplets that included the distance across three circuli, and that were present on every scale in the standards. The number of circuli may be brood year specific. All the complete triplets present in the database should be used in a discriminant analysis.

Acknowledgments

I wish to thank Kate W. Myers for critical review of the manuscript, and thank her and Robert V. Walker for their suggestions made throughout the study. Funding was provided by the National Marine Fisheries Service, National Oceanic and Atmospheric Administration (NOAA), Research Contract Nos. 85-ABC-00006 and 50-ABNF-7-00002.

REFERENCES CITED

Amos, M.H., R.E. Anas, and R.E. Pearson. 1963. Use of discriminant function in the morphological separation of Asian and North American races of pink salmon, *Oncorhynchus gorbuscha* (Walbaum). Int. North Pac. Fish. Comm., Bull. 11:73-100.

Anas, R.E., and S. Murai. 1969. Use of scale characters and a discriminant function for classifying sockeye salmon (*Oncorhynchus nerka*) by continent of origin. Int. North Pac. Fish. Comm., Bull. 26:157-192.

Anonymous. 1986. SPSSX User's Guide, Second Edition. SPSS, Inc., Chicago 988 pp.

Atchley, W.R., C.T. Gaskins, and D. Anderson. 1976. Statistical properties of ratios. 1. Empirical results. Syst. Zool. 25: 137-148.

Bilton, H.T., and H.B. Messinger. 1975. Identification of major British Columbia and Alaska runs of age 1.2 and 1.3 sockeye from their scale characters. Int. North Pac. Fish. Comm., Bull. 32:109-129.

Conrad, R.H. 1984. Origins of chum salmon in the Unimak Island and Shumagin Islands fisheries during June 1983 determined by scale patterns analysis. (Doc. submitted to annual meeting of the Int. North Pac. Fish. Comm., Vancouver, Canada, November 1984.) 21 pp. Alaska Dept. of Fish and Game, Anchorage.

Cook, R.C. 1982. Estimating the mixing proportion of salmonids with scale pattern recognition applied to sockeye salmon (*Oncorhynchus nerka*) in and around the Japanese landbased driftnet fishery area. 264 pp. Ph. D. Dissertation, Univ. of Washington, Seattle.

Cook, R.C., and G.E. Lord. 1978. Identification of stocks of Bristol Bay sockeye salmon *Oncorhynchus nerka,* by evaluating scale patterns with a polynomial discriminant method. U.S. Fish Wildl. Serv., Fish. Bull. 76(2):415-423.

Cook, R.C., R.H. Conrad, K.W. Myers, R.V. Walker and C.K. Harris. 1980. The mixing proportion of Asian and Alaskan sockeye salmon in and around the landbased driftnet fishery area as determined by scale pattern recognition. (Doc. submitted to annual meeting of the Int. North Pac. Fish. Comm., Anchorage, USA, November 1980.) 58 pp. Fish. Res. Inst., Univ. of Washington, Seattle.

Cook, R.C., K.W. Myers, R.V. Walker, and C.K. Harris. 1981. The mixing proportions of Asian and Alaskan sockeye salmon in and around the landbased driftnet fishery area 1972-1976. (Doc. submitted to annual meeting of the Int. North Pac. Fish. Comm., Vancouver, Canada, November 1981.) 81 pp. Fish. Res. Inst., Univ. of Washington, Seattle.

Dixon, W.J., M.B. Brown, L. Engelman, J.W. Frane, M.A. Hill, R.I. Jennrich, and J.D. Toporek. 1983. BMDP Statistical Software 1983 Printing with Additions. 734 pp., Univ. of Calif. Press, Berkeley.

Fisher, R.A. 1936. The use of multiple measurements in taxonomic problems. Ann. Eugen. 7:179-188.

Habbema, J.D.F. and J. Hermans. 1977. Selection of variables in discriminant analysis by F-statistic and error rate. Technometrics 19:487-493.

Humphries, J.M., F.L. Bookstein, B. Chernoff, G.R. Smith, R.L. Elder and S.G. Poss. 1981. Multivariate discrimination by shape in relation to size. Syst. Zool. 30:291-308.

Ishida, Y., S. Ito, and K. Takagi. 1984. Further analysis of scale patterns of Japanese hatchery-reared chum salmon in the North Pacific Ocean. (Doc. submitted to the Int. North Pac. Fish. Comm.) 7 pp. Fishery Agency of Japan, Tokyo, Japan 100.

Ishida, Y., S. Ito, and K. Takagi. 1985. Stock identification of chum salmon based on scale patterns by discriminant function. (Doc. submitted to the Int. North Pac. Fish. Comm.) 7 pp. Fishery Agency of Japan, Tokyo, Japan 100.

Ito, J., Y. Ishida, and S. Ito. 1985. Stock identification of chinook salmon in offshore waters in 1974 based on scale pattern analysis. (Doc. submitted to the Int. North Pac. Fish. Comm.) 14 pp. Fishery Agency of Japan, Tokyo, Japan 100.

Ito, J., Y. Ishida, and S. Ito. 1986. Further analysis of stock identification of chinook salmon in offshore waters in 1974. (Doc. submitted to annual meeting of the Int. North Pac. Fish. Comm., Anchorage, Ak. 1986 October.) 9 pp. Fishery Agency of Japan, Tokyo, Japan 100.

Kato, M. and Y. Ishida. 1985. Scale pattern analysis of coho salmon in the northwest North Pacific Ocean by materials obtained in 1975. (Doc. submitted to Int. North Pac. Fish. Comm.) 8 pp. Fishery Agency of Japan, Tokyo, Japan 100.

Kato, M., and Y. Ishida. 1986. Scale pattern analysis of coho salmon in the northwest North Pacific Ocean using materials obtained by salmon research vessels in 1976. (Doc. submitted to the annual meeting of the Int. North. Pac. Fish. Comm., Anchorage, Ak. 1986 October.) 8 pp. Fishery Agency of Japan, Tokyo, Japan 100.

Knudsen, C.M. 1985. Chinook salmon scale character variability due to body area sampled and possible effects on stock separation studies. 141 pp. MS Thesis, Univ. of Washington, Seattle.

Knudsen, C.M., and N.D. Davis. 1985. Variation in salmon scale characters due to body area sampled. (Doc. submitted to annual meeting of the Int. North Pac. Fish. Comm., Tokyo, Japan, November 1985.) 59 pp. Univ. of Washington, Fish. Res. Inst., FRI-UW-8504. Seattle.

Knudsen, C.M., C.K. Harris, and N.D. Davis. 1983. Origins of chinook salmon in the area of the Japanese mothership and landbased driftnet salmon fisheries in 1980. (Doc. submitted to annual meeting of the Int. North Pac. Fish. Comm., Anchorage, Ak. USA, November 1983.) 71 pp. Univ. of Washington, Fish. Res. Inst., FRI-UW-8315. Seattle.

Kruskal, W.H., and W. A. Wallis. 1952. Use of ranks in one-criterion variance analysis. J. Amer. Stat. Assoc. 47:583-621.

Mahalanobis, P.C. 1936. On the generalized distance in statistics. Proceedings of the National Inst. of Sci., India 12:49-55.

Major, R.L., S. Murai and J. Lyons. 1975. Scale studies to identify Asian and western Alaskan chinook salmon. Int. North Pac. Fish. Comm., Annual Report 1973:80-97.

Major, R.L., S. Murai, and J. Lyons. 1977a. Scale studies to identify Asian and western Alaskan chinook salmon: the 1969 and 1970 Japanese mothership samples. Int. North Pac. Fish. Comm., Annual Report 1974:78-81.

Major, R.L., S. Murai, and J. Lyons. 1977b. Scale studies to identify Asian and western Alaskan chinook salmon. Int. North Pac. Fish. Comm., Annual Report 1975:68-71.

Marshall, S.L., C.K. Harris, D.E. Rogers and R.C. Cook. 1978. Investigations on the continent of origin of sockeye and coho salmon in the area of the Japanese land-based driftnet fishery. Final Rep. FRI-UW-7816. North Pac. Fish. Management Council. 152 pp.

Marshall, S.L., R.C. Cook, C.K. Harris, and R.H. Conrad. 1979. Origins of immature sockeye in and around the area of the Japanese landbased driftnet fishery in 1974 and 1975 as determined by evaluation of scale pattern with a discriminant function. Submitted to the Int. North Pac. Fish. Comm. by the US National Section. Fish. Res. Inst., Univ. of Washington, Seattle.

Myers, K.W., C.K. Harris, C.M. Knudsen, R.V. Walker, N.D. Davis, and D.E. Rogers. (in press). Stock origins of chinook salmon in the area of the Japanese mothership salmon fishery. North American Journal of Fisheries Management.

Myers, K.W. 1985. Racial trends in chinook salmon (*Oncorhynchus tshawytscha*) scale patterns. (Doc. submitted to the Int. North Pac. Fish. Comm.) 56 pp. Fish. Res. Inst., Univ. of Washington, FRI-UW-8503. Seattle.

Myers, K.W. 1986. The effect of altering proportions of Asian chinook stocks on regional scale pattern analysis. (Doc. submitted to the annual meeting of the Int. North Pac. Fish. Comm., Anchorage, USA, November 1986.) 44 pp. Univ. of Washington, Fish. Res. Inst. FRI-UW-8605. Seattle.

Myers, K.W., R.C. Cook, R.V. Walker, and C.K. Harris. 1981. The continent of origin of coho salmon in the Japanese landbased driftnet fishery area in 1979. (Doc. submitted to annual meeting of the Int. North Pac. Fish. Comm, Vancouver, B.C., Canada, November 1981.) 34 pp. Fish. Res. Inst., Univ. of Washington, Seattle.

Myers, K.W., D.E. Rogers, C.K. Harris, C.M. Knudsen, R.V. Walker, and N.D. Davis. 1984. Origins of chinook salmon in the area of the Japanese mothership and landbased driftnet salmon fisheries in 1975-1981. (Doc. submitted to annual meeting of the Int. North Pac. Fish. Comm., Vancouver, Canada, November 1984.) 208 pp. Univ. of Washington, Fish. Res. Inst., Seattle.

Myers, K.W. and D.E. Rogers. 1985a. Determination of stock origins of chinook salmon incidentally caught in foreign trawls in Alaskan FCZ: Part II. 76 pp. Council Doc, 32, North Pacific Fishery Management Council, Anchorage.

Myers, K.W. and D.E. Rogers. 1985b. Precision and variability of scale pattern data from two stocks of southeastern Alaskan coho salmon. Final Report. 133 pp. FRI-UW-8509. Fisheries Research Institute, Univ. of Washington, Seattle.

Nie, N.H., C.H. Hull, J.G. Jenkins, K. Steinbrenner, and D.H. Bent. 1975. Statistical Package for the Social Sciences. Second Edition. McGraw-Hill Book Co. 675 pp.

Rao, C.R. 1952. Advanced Statistical Methods in Biometric Research. Wiley, New York.

Reyment, R.A., R.E. Blackith, and N.A. Campbell. 1984. Multivariate morphometrics. Second Edition. Academic Press, London. 233 pp.

Tanaka, S., M.P. Shepard, and H.T. Bilton. 1969. Origin of chum salmon (*Oncorhynchus keta*) in offshore waters of the North Pacific in 1956-1958 as determined from scale studies. Int. North Pac. Fish. Comm., Bull. 26:57-155.

Walker, R.V., and C.K. Harris. 1982. The continent of origin of coho salmon in the Japanese landbased driftnet fishery area in 1980. (Doc. submitted to annual meeting of the Int. North Pac. Fish. Comm., Tokyo, Japan, November 1982.) 26 pp. Fish. Res. Inst., Univ. of Washington, Seattle.

Walker, R.V., and N.D. Davis. 1983. The continent of origin of coho salmon in the Japanese driftnet fishery areas in 1981. (Doc. submitted to annual meeting of the Int. North Pac. Fish. Comm., Anchorage, USA, November 1983.) 48 pp. Univ. of Washington, Fish. Res. Inst., FRI-UW-8314, Seattle.

Wilks, S.S. 1932. Certain generalizations in the analysis of variance. Biometrika 24:471-474.

Table 1. Description of chinook scale characters.

Character name	Character[a] category	Description[b]
C1[c]	LD	Size Zone 1
C5[c]	LD	Size Zone 2 + Size Zone 3
C6	LD	Size Zone 1 + Size Zone 2 + Size Zone 3
C7	CC	No. circuli Zone 1 + no. circuli Zone 2 + no. circuli Zone 3
C9	ACS	(Size Zone 1 + Size Zone 2 + Size Zone 3)/(no. circuli Zone 1 + no. circuli Zone 2 + no. circuli Zone 3)
C11	LDR	(Size Zone 2 + Size Zone 3)/(Size Zone 1 + Size Zone 2 + Size Zone 3)
C12[c]	CC	No. circuli Zone 1
C16[c]	CC	No. circuli Zone 2 + no. circuli Zone 3
C17	ACS	Size Zone 1/no. circuli Zone 1
C21	ACS	(Size Zone 2 + Size Zone 3)/(no. circuli Zone 2 + no. circuli Zone 3)
C22	MTR	Distance C1 to C3 in Zones 2 + 3/(Size Zone 1 + Size Zone 2 + Size Zone 3)
C23	MTR	Distance C4 to C6 in Zones 2 + 3/(Size Zone 1 + Size Zone 2 + Size Zone 3)
C24	MTR	Distance C7 to C9 in Zones 2 + 3/(Size Zone 1 + Size Zone 2 + Size Zone 3)
C25	MTR	Distance C10 to C12 in Zones 2 + 3/(Size Zone 1 + Size Zone 2 + Size Zone 3)
C26	MTR	Distance C13 to C15 in Zones 2 + 3/(Size Zone 1 + Size Zone 2 + Size Zone 3)
C27	MTR	Distance C16 to C18 in Zones 2 + 3/(Size Zone 1 + Size Zone 2 + Size Zone 3)
C28	MTR	Distance C19 to C21 in Zones 2 + 3/(Size Zone 1 + Size Zone 2 + Size Zone 3)
C29	MTR	Distance C22 to C24 in Zones 2 + 3/(Size Zone 1 + Size Zone 2 + Size Zone 3)
C30	MTR	Distance C25 to C27 in Zones 2 + 3/(Size Zone 1 + Size Zone 2 + Size Zone 3)
C31	MTR	Distance C28 to C30 in Zones 2 + 3/(Size Zone 1 + Size Zone 2 + Size Zone 3)
C32	MTR	Distance C31 to C33 in Zones 2 + 3/(Size Zone 1 + Size Zone 2 + Size Zone 3)
C33	MTR	Distance C34 to C36 in Zones 2 + 3/(Size Zone 1 + Size Zone 2 + Size Zone 3)
C34	MN	Distance C1 to C9 in Zones 2 + 3 (= characters C49 + C50 + C51)
C35	MN	Distance C10 to C18 in Zones 2 + 3 (= characters C52 + C53 + C54)
C36	MN	Distance C19 to C27 in Zones 2 + 3 (= characters C55 + C56 + C57)
C37	MN	Distance C28 to C36 in Zones 2 + 3 (= characters C58 + C59 + C60)
C39[c]	FWT	Distance C2 to C4 in Zone 1 (distance from the outer edge of the first circulus at the focus to the outer edge of C4)

Table 1. Description of chinook scale characters - cont'd.

Character name	Character[a] category	Description[b]
C40	FWT	Distance C5 to C7 in Zone 1
C41	FWT	Distance C8 to C10 in Zone 1
C42	FWT	Distance C11 to C13 in Zone 1
C43	FWT	Distance C14 to C16 in Zone 1
C44	FWTR	Distance C2 to C4 in Zone 1/(Size Zone 1+ Size Zone 2 + Size Zone 3)
C45	FWTR	Distance C5 to C7 in Zone 1/(Size Zone 1 + Size Zone 2 + Size Zone 3)
C46	FWTR	Distance C8 to C10 in Zone 1/(Size Zone 1 + Size Zone 2 + Size Zone 3)
C47	FWTR	Distance C11 to C13 in Zone 1/(Size Zone 1 + Size Zone 2 + Size Zone 3)
C48	FWTR	Distance C14 to C16 in Zone 1/(Size Zone 1 + Size Zone 2 + Size Zone 3)
C49[c]	MT	Distance C1 to C3 in Zones 2 + 3
C50[c]	MT	Distance C4 to C6 in Zones 2 + 3
C51[c]	MT	Distance C7 to C9 in Zones 2 + 3
C52	MT	Distance C10 to C12 in Zones 2 + 3
C53	MT	Distance C13 to C15 in Zones 2 + 3
C54	MT	Distance C16 to C18 in Zones 2 + 3
C55	MT	Distance C19 to C21 in Zones 2 + 3
C56	MT	Distance C22 to C24 in Zones 2 + 3
C57	MT	Distance C25 to C27 in Zones 2 + 3
C58	MT	Distance C28 to C30 in Zones 2 + 3
C59	MT	Distance C31 to C33 in Zones 2 + 3
C60	MT	Distance C34 to C36 in Zones 2 + 3
C63	CCR	No. circuli Zones 2 + 3/(no. circuli Zones 1+2+3)
TR1	T	Distance to C2 to C4
TR2	T	Distance C5 to C7
TR3	T	Distance C8 to C10
TR4	T	Distance C11 to C13
TR5	T	Distance C14 to C16
TR6	T	Distance C17 to C19
TR7	T	Distance C20 to C22
TR8	T	Distance C23 to C25
TR9	T	Distance C26 to C28
FWD1	FWD	Distance C2 to C3 in Zone 1
FWD2	FWD	Distance C4 to C5 in Zone 1
MD1	MD	Distance C1 to C2 in Zones 2 + 3
MD2	MD	Distance C3 to C4 in Zones 2 + 3
MD3	MD	Distance C5 to C6 in Zones 2 + 3
MD4	MD	Distance C7 to C8 in Zones 2 + 3
MD5	MD	Distance C9 to C10 in Zones 2 + 3
MD6	MD	Distance C11 to C12 in Zones 2 + 3
MD7	MD	Distance C13 to C14 in Zones 2 + 3
MD8	MD	Distance C15 to C16 in Zones 2 + 3
MD9	MD	Distance C17 to C18 in Zones 2 + 3

Table 1. Description of chinook scale characters - cont'd.

Character name	Character[a] category	Description[b]
FWQD1	FWQUAD	Distance C2 to C5 in Zone 1
MQD1	MQUAD	Distance C1 to C4 in Zones 2 + 3
MQD2	MQUAD	Distance C5 to C8 in Zones 2 + 3
MQD3	MQUAD	Distance C9 to C12 in Zones 2 + 3
MQD4	MQUAD	Distance C13 to C16 in Zones 2 + 3
MQN1	MQUIN	Distance C1 to C5 in Zones 2 + 3
MQN2	MQUIN	Distance C6 to C10 in Zones 2 + 3
MQN3	MQUIN	Distance C11 to C15 in Zones 2 + 3
D1[d]	RDUM	random dummy character
D2[d]	RDUM	random dummy character
D3[d]	RDUM	random dummy character
D4[d]	RDUM	random dummy character

---

[a] Summary of categories (LD through CCR are after Knudsen [1985] and Knudsen and Davis [1985]):

LD: Large distances (zone sizes)
CC: Circuli count (number of circuli in a zone or combination of zones)
ACS: Average circulus spacing (zone size/number of circuli in a zone)
LDR: Large distance ratios (zone size/total size)
MTR: Marine triplet ratios (distance between circuli triplets in Zones 2 + 3/total size)
MT: Marine triplets (distance between circuli triplets in Zones 2 + 3)
FWT: Freshwater triplets (distance between circuli triplets in Zone 1)
FWTR: Freshwater triplets ratios (distance between circuli triplets in Zone 1/total size)
CCR: Circuli count ratios (circuli count in a zone/total number of circuli)
MN: Marine nonuplets (distance between nine circuli in Zones 2 + 3)
T: Triplets (distance between circuli triplets counted continuously from the focus to the last circulus in the ocean annulus)
FWD: Freshwater doublets (distance between circuli pairs in Zone 1)
MD: Marine doublets (distance between circuli pairs on Zones 2 + 3)
FWQUAD: Freshwater quadruplets (distance between four circuli in Zone 1)
MQUAD: Marine quadruplets (distance between four circuli in Zones 2 + 3)
MQUIN: Marine quintuplets (distance between five circuli in Zones 2 + 3)
RDUM: Random dummy character

[b] Zone 1: The area of the scale from the center of the focus to the outer edge of the last circulus in the freshwater annulus
Zone 2: The area of the scale from the outer edge of the last circulus in the freshwater annulus to the outer edge of the last freshwater circulus
Zone 3: The area of the scale from the outer edge of the last freshwater circulus to the outer edge of the last circulus in the first ocean annulus.
Cn: The nth circulus from the focus of the scale.

[c] Character used by Ito et al. (1985): character name this study = character name Ito et al. C1 = FCL, C5 = OCL, C12 = FCN, C16 = OCN, C39 = F04, C49 = O03, C50 = O06, C51 = O09

[d] Random variables generated from a population with mean = 370.0 and standard deviation = 70.0.

Table 2. Scale character set names and constituent characters.

Character set name	Number of characters	Scale character names
Myers et al. (1984)	48	C1,C5,C6,C7,C9,C11,C12,C16,C17, C21,C22,C23,C24,C25,C26,C27,C28, C29,C30,C31,C32,C33,C34,C35,C36, C37,C39,C40,C41,C42,C43,C44,C45, C46,C47,C48,C49,C50,C51,C52,C53, C54,C55,C56,C57,C58,C59,C60
Ito et al. (1985)	8	C1,C5,C12,C16,C39,C49,C50,C51
Char1	27	C1,C5,C6,C7,C9,C11,C12,C16,C17, C21,C22,C23,C24,C25,C26,C27,C34 C35,C39,C44,C49,C50,C51,C52,C53, C54,C63
Char1A: non-ratio LD, CC, FWT and MT	11	C1,C5,C12,C16,C39,C49,C50,C51, C52,C53,C54
Char1B: ratio LD, CC, FWT and MT	12	C9,C11,C17,C21,C22,C23,C24,C25, C26,C27,C44,C63
Char2 (BY73): Char1 with three transformed characters	27	C1,C5,C6,C7,C9,C11,C12S[a],C16, C17,C21,C22ASIN[b],C23,C24,C25, C26,C27,C34,C35,C39,C44,C49L[c], C50,C51,C52,C53,C54,C63
Char2 (BY77): Char1 with eight transformed characters	27	C1,C5,C6,C7,C9,C11ASIN[d],C12L[e], C16,C17,C21L[f],C22ASIN[b], C23ASIN[g],C24ASIN[h],C25, C26,C27,C34,C35,C39,C44,C49L[c], C50,C51,C52,C53,C54,C63ASIN[i]
Char2A (BY73): non-ratio LD, CC, FWT and MT	11	C1,C5,C12S[a],C16,C39,C49L[c],C50, C51,C52,C53,C54
Char2A (BY77): non-ratio LD, CC, FWT and MT	11	C1,C5,C12L[e],C16,C39,C49L[c],C50, C51,C52,C53,C54
Char2B (BY73): ratio LD, CC, FWT and MT	12	C9,C11,C17,C21,C22ASIN[b], C23,C24,C25,C26,C27,C44,C63
Char2B (BY77): ratio LD, CC, FWT and MT	12	C9,C11ASIN[d],C17,C21L[f],C22ASIN[b], C23ASIN[g],C24ASIN[h],C25,C26,C27, C44,C63ASIN[i]

Table 2. Scale character set names and constituent characters - cont'd.

Character set name	Number of characters	Scale character names
Char3A: FWT and MT	7	C39,C49,C50,C51,C52,C53,C54
Char3B: FWT, MT and CC	9	C39,C49,C50,C51,C52,C53,C54,C12, C16
Char3C: FWT, MT and LD	9	C39,C49,C50,C51,C52,C53,C54,C1, C5
Char4A: FWT, MT and Freshwater ACS	8	C39,C49,C50,C51,C52,C53,C54,C17
Char4B: FWT, MT and Marine ACS	8	C39,C49,C50,C51,C52,C53,C54,C21
Char4C: FWT, MT and Overall-ACS	8	C39,C49,C50,C51,C52,C53,C54,C9
Char4D: FWT, MT, Freshwater ACS and Marine ACS	9	C39,C49,C50,C51,C52,C53,C54,C17, C21
Char4E: FWT, MT, Freshwater ACS, ocean ACS and overall ACS	10	C39,C49,C50,C51,C52,C53,C54,C17, C21,C9
Char5A (BY73): T	8	TR1,TR2,TR3,TR4,TR5,TR6,TR7,TR8
Char5A (BY77): T	9	TR1,TR2,TR3,TR4,TR5,TR6,TR7,TR8, TR9
Char5B (BY73): T and CC	9	TR1,TR2,TR3,TR4,TR5,TR6,TR7, TR8,C7
Char5B (BY77): T and CC	10	TR1,TR2,TR3,TR4,TR5,TR6,TR7,TR8, TR9,C7
Char5C (BY73): T and LD	9	TR1,TR2,TR3,TR4,TR5,TR6,TR7,TR8, C6
Char5C (BY77): T and LD	10	TR1,TR2,TR3,TR4,TR5,TR6,TR7,TR8, TR9,C6
Char5D (BY73): T, CC and LD	10	TR1,TR2,TR3,TR4,TR5,TR6,TR7,TR8, C7,C6
Char5D (BY77): T, CC and LD	11	TR1,TR2,TR3,TR4,TR5,TR6,TR7,TR8, TR9,C7,C6
Char6A: FWD, MD, CC and LD	15	C12,C1,C16,C5,FWD1,FWD2,MD1,MD2, MD3,MD4,MD5,MD6,MD7,MD8,MD9

Table 2. Scale character set names and constituent characters - cont'd.

Character set name	Number of characters	Scale character names
Char6B: FWQUAD, MQUAD, CC and LD	9	C12,C1,C16,C5,FWQD1,MQD1,MQD2, MQD3,MQD4
Char6C: FWQUAD, MQUIN, CC and LD	8	C12,C1,C16,C5,FWQD1,MQN1,MQN2, MQN3
Char7A: Char 1A and 1 RDUM	12	C12,C1,C16,C5,C39,C49,C50,C51, C52,C53,C54,D1
Char7B: Char 1A and 2 RDUM	13	C12,C1,C16,C5,C39,C49,C50,C51, C52,C53,C54,D1,D2
Char7C: Char 1A and 3 RDUM	14	C12,C1,C16,C5,C39,C49,C50,C51, C52,C53,C54,D1,D2
Char7D: Char 1A and 4 RDUM	15	C12,C1,C16,C5,C39,C49,C50,C51, C52,C53,C54,D1,D2,D3,D4
Char7E: CC and LD	4	C12,C16,C1,C5
Char7F: CC, LD and 4 RDUM	8	C12,C16,C1,C5,D1,D2,D3,D4
Char7G: FWT and MT	7	C39,C49,C50,C51,C52,C53,C54
Char7H: FWT, MT and 4 RDUM	11	C39,C49,C50,C51,C52,C53,C54,D1, D2,D3,D4

[a] C12S = square root (C12)
[b] C22ASIN = arcsin (square root C22)
[c] C49L = log (C49 + 1)
[d] C11ASIN = arcsin (square root C11)
[e] C12L = log (C12 + 1)
[f] C21L = log (C21 + 1)
[g] C23ASIN = arcsin (square root C23)
[h] C24ASIN = arcsin (square root C24)
[i] C63ASIN = arcsin (square root C63)

Table 3. Classification accuracies from linear discriminant analysis of chinook scales using characters selected by Wilks' Lambda and the Mahalanobis selection criteria from a large character set used by Myers et al. (1984) and an *apriori* selected character set used by Ito et al. (1985). Character selection subject to the constraints: F-to-enter ≥ 4.000, F-to-remove <3.996 and minimum tolerance = .010. A. Brood year 1973. B. Brood year 1977.

Character set name (no. of characters	Selection method[a]			Number of characters selected	Subset selected	Percent correctly classified				
	W/J	W	M			Unweighted overall average	ASIA	WEST	CENT	SEBC
A. Brood year 1973										
1. Myers et al.[bc] (48)	x			11	C34,C7,C21,C35,C44,C36, C11,C5,C23,C52,C58	71.7	58.5	81.3	64.9	82.0
2. Myers et al. (48)		x		12	C34,C7,C21,C44,C36,C11, C5,C23,C52,C58,C1,C39	72.1	61.9	81.8	63.4	81.4
3. Myers et al. (48)			x	14	C9,C44,C52,C51,C12,C11, C22,C57,C26,C53,C23, C16,C30,C49	72.9	63.6	79.8	64.2	84.0
4. Ito et al. (8)	x			8	Full model	69.2	65.3	71.2	59.0	81.4
5. Ito et al. (8)		x[d]	x[d]	8	Full model	71.6	67.8	72.7	62.7	83.0
B. Brood year 1977										
1. Myers et al.[b] (48)	x			16	C27,C9,C34,C17,C58,C16, C31,C35,C28,C44,C42,C21, C36,C47,C25,C26	80.0	87.0	73.4	73.9	85.9
2. Myers et al. (48)		x		16	Same subset as B.1.	81.2	87.5	74.4	75.4	87.4

Table 3. Classification accuracies from linear discriminant analysis of chinook scales using characters selected by Wilks' Lambda and the Mahalanobis selection criteria from a large character set used by Myers et al. (1984) and an *a priori* selected character set used by Ito et al. (1985). Character selection subject to the constraints: F-to-enter ≥ 4.000, F-to-remove <3.996 and minimum tolerance = .010. - cont'd. A. Brood year 1973. B. Brood year 1977.

Character set name (no. of characters	Selection method[a]			Number of characters selected	Subset selected	Percent correctly classified				
	W/J	W	M			Unweighted overall average	ASIA	WEST	CENT	SEBC
B. Brood year 1977										
3. Myers et al. (48)			x	21	C7,C17,C44,C50,C53,C26, C21,C30,C57,C1,C28,C52, C47,C42,C49,C31,C58,C24, C27,C54,C36	82.0	88.0	76.4	77.4	86.4
4. Ito et al. (8)	x			8	Full model	78.1	91.0	71.9	66.3	83.3
5. Ito et al. (8)		x[d]	x[d]	8	Full model	78.9	92.0	72.9	66.8	83.8

[a] Selection method:

W/J minimize Wilks' Lambda and use the jackknife (leaving-one-out approach) to calculate LDF and nearly unbiased classification accuracies.

W minimize Wilks' Lambda and use all cases to calculate the LDF and biased classification accuracies.

M maximize Mahalanobis distance between two closest groups and use all cases to calculate the LDF and biased classification accuracies.

[b] Results are reported by Myers et al. (1984).

[c] Myers et al. used a subset selected at an intermediate step, i.e., before the stopping constraints were invoked.

[d] Selected subset and resulting classification accuracies are identical using both methods of character selection.

Table 4. Classification accuracies from linear discriminant analysis of chinook scales using characters selected by Wilks' Lambda and the Mahalanobis selection criteria from a large character set (Char1) and non-ratio (Char1A) and ratio (Char1B) character sets. Character selection subject to the constraints: F-to-enter ≥4.000 and F-to-remove <3.996 except where otherwise noted. A. Brood year 1973. B. Brood year 1977.

Character set name (no. of characters)	Selection method^a				Tolerance		Number of characters	Subset selected	Percent correctly classified				
	W/J	W	M	F/J	.010	.001			Unweighted overall average	ASIA	WEST	CENT	SEBC
A. Brood year 1973													
1. Char1 (27)	x				x		13	C9,C34,C7,C21,C35,C44, C54,C27,C17,C23,C50, C26,C25	69.9	57.6	79.8	59.0	83.0
2. Char1 (27)		x				x	13	C9,C34,C7,C21,C35,C6,C44, C54,C27,C17,C23,C50,C26	72.3	60.2	81.8	66.4	80.9
3. Char1 (27)			x			x	12	C26,C21,C6,C25,C34,C12, C39,C23,C53,C50,C16,C63	72.7	62.7	81.3	65.7	80.9
4. Char1 (27)				x		x	23	C9,C34,C7,C21,C35,C6,C44, C54,C27,C17,C23,C50,C26, C25,C22,C51,C16,C11,C53, C24,C39,C1,C63	70.9	58.5	79.8	64.2	80.9
5. Char1A (11)	x				x		10	C12,C16,C5,C50,C49,C52, C51,C53,C1,C39	69.7	57.6	76.8	61.9	82.5
6. Char1A (11)		x^b	x^b		x	x	10	Same subset as A.5.	71.1	60.2	77.3	63.4	83.5
7. Char1A (11)				x	x		11	Full model	69.5	57.6	76.8	61.2	82.5

Table 4. Classification accuracies from linear discriminant analysis of chinook scales using characters selected by Wilks' Lambda and the Mahalanobis selection criteria from a large character set (Char1) and non-ratio (Char1A) and ratio (Char1B) character sets. Character selection subject to the constraints: F-to-enter ≥4.000 and F-to-remove <3.996 except where otherwise noted - cont'd. A. Brood year 1973. B. Brood year 1977.

Character set name (no. of characters)	Selection method^a				Tolerance		Number of characters	Subset selected	Percent correctly classified				
	W/J	W	M	F/J	.010	.001			Unweighted overall average	ASIA	WEST	CENT	SEBC
A. Brood year 1973													
8. Char1B (12)	x				x		9	C9,C27,C22,C21,C26,C23,C25 C44,C11	66.7	54.2	74.2	57.5	80.9
9. Char1B (12)		x^b	x^b			x	9	C27,C22,C21,C26,C23,C25, C44,C11,C63	68.4	57.6	74.2	59.7	82.0
10. Char1B (12)				x		x	12	Full model	65.7	55.1	74.7	53.0	79.9
B. Brood year 1977													
1. Char1 (27)	x				x		9	C27,C34,C17,C54,C35,C16, C21,C44,C5	78.8	86.5	73.9	69.3	85.4
2. Char1 (27)		x^c				x	18	C27,C9,C34,C17,C54,C35,C16, C21,C44,C5,C26,C25,C11,C12, C63,C1,C53,C49	79.5	84.5	73.4	74.9	85.4
3. Char1 (27)			x^c			x	17	C7,C5,C17,C44,C26,C12,C53, C34,C52,C9,C27,C63,C21, C11,C6,C35,C49	80.3	85.0	75.4	75.9	84.8

Table 4. Classification accuracies from linear discriminant analysis of chinook scales using characters selected by Wilks' Lambda and the Mahalanobis selection criteria from a large character set (Char1) and non-ratio (Char1A) and ratio (Char1B) character sets. Character selection subject to the constraints: F-to-enter ≥4.000 and F-to-remove <3.996 except where otherwise noted - cont'd. A. Brood year 1973. B. Brood year 1977.

Character set name (no. of characters)	Selection method[a]				Tolerance		Number of characters	Subset selected	Percent correctly classified				
	W/J	W	M	F/J	.010	.001			Unweighted overall average	ASIA	WEST	CENT	SEBC
B. Brood year 1977													
4. Char1 (27)				x		x	22	C27,C9,C34,C17,C54,C35, C16,C21,C44,C5,C26,C25, C11,C12,C63,C6,C53,C49, C24,C23,C22,C50	78.0	84.0	73.4	71.9	82.8
5. Char1A (11)	x				x		11	Full model	79.0	91.0	72.9	67.8	84.3
6. Char1A (11)		x^{bc}	x^{bc}		x	x^{c}	11	Full model	80.0	92.0	74.4	69.3	84.3
7. Char1B (12)	x				x		9	C27,C9,C17,C23,C26,C11, C44,C25,C63	77.8	83.5	75.4	69.8	82.3
8. Char1B (12)		x^{bc}	x^{bc}			x^{c}	11	C27,C9,C17,C23,C26,C11, C44,C25,C63,C21,C24	77.6	83.5	74.4	70.4	82.3
9. Char1B (12)				x		x	12	Full model	76.3	82.5	72.9	67.8	81.8

Table 4. Classification accuracies from linear discriminant analysis of chinook scales using characters selected by Wilks' Lambda and the Mahalanobis selection criteria from a large character set (Char1) and non-ratio (Char1A) and ratio (Char1B) character sets. Character selection subject to the constraints: F-to-enter ≥4.000 and F-to-remove <3.996 except where otherwise noted - cont'd. A. Brood year 1973. B. Brood year 1977.

[a]Selection method:

W/J	minimize Wilks' Lambda and use the jackknife (leaving-one-out approach) to calculate LDF and nearly unbiased classification accuracies.
W	minimize Wilks' Lambda and use all cases to calculate the LDF and biased classification accuracies.
M	maximize Mahalanobis distance between two closest groups and use all cases to calculate the LDF and biased classification accuracies.
F/J	force all characters into the model subject to specified tolerance level. Jackknife approach to calculate LDF and nearly unbiased classification accuracies.

[b] Selected subset and resulting classification accuracies are identical using both methods of character selection.

[c] Estimated probabilities were used as a stopping rule by maintaining a fixed significance level: probability of F-to-enter ≤.05, probability of F-to-remove >.10 (instead of a fixed F-value).

Table 5. Classification accuracies from linear discriminant analysis of chinook scales using characters selected by Wilks' Lambda and the Mahalanobis selection criteria from a large character set (Char2) a non-ratio character set (Char2A) and a ratio character set (Char2B). All character sets include transformed characters. Characters described on Tables 1 and 2. Character selection subject to the constraints: F-to-enter ≥4.000, F-to-remove <3.996 except where otherwise noted. A. Brood year 1973. B. Brood year 1977.

Character set name (no. of characters)	Selection method[a]			Tolerance		Number of characters selected	Subset selected	Percent correctly classified				
	W/J	W	M	.010	.001			Unweighted overall average	ASIA	WEST	CENT	SEBC
A. Brood year 1973												
1. Char2 (27)	x			x		13	C9,C34,C7,C21,C35,C44, C54,C27,C17,C23,C50, C26,C25	69.9	57.6	79.8	59.0	83.0
2. Char2 (27)		x			x	13	C9,C34,C7,C21,C35,C6,C44, C54,C27,C17,C23,C50,C26	72.3	60.2	81.8	66.4	80.9
3. Char2 (27)			x		x	11	C26,C21,C6,C25,C34,C7, C39,C11,C23,C53,C50	71.9	58.5	80.3	66.4	82.5
4. Char2A (11)	x			x		10	C12S,C16,C5,C49L,C50, C52,C51,C53,C1,C39	68.7	55.9	76.8	59.0	83.0
5. Char2A (11)		x[b]	x[b]		x	10	Same subset as A.4.	70.5	59.3	77.3	61.2	84.0
6. Char2B (12)	x			x		9	C9,C27,C22ASIN,C21,C26, C23,C25,C44,C11	66.5	54.2	74.2	56.7	80.9
7. Char2B (12)		x[b]	x[b]		x	9	C27,C22ASIN,C21,C26,C23, C25,C44,C11,C63	68.2	57.6	74.2	59.0	82.0

Table 5. Classification accuracies from linear discriminant analysis of chinook scales using characters selected by Wilks' Lambda and the Mahalanobis selection criteria from a large character set (Char2), a non-ratio character set (Char2A) and a ratio character set (Char2B). All character sets include transformed characters. Characters are described on Tables 1 and 2. Character selection subject to the constraints: F-to-enter ≥4.000, F-to-remove <3.996 except where otherwise noted - cont'd. A. Brood year 1973. B. Brood year 1977.

Character set name (no. of characters)	Selection method[a]			Tolerance		Number of characters selected	Subset selected	Percent correctly classified				
	W/J	W	M	.010	.001			Unweighted overall average	ASIA	WEST	CENT	SEBC
B. Brood year 1977												
1. Char2 (27)	x			x		12	C27,C9,C34,C17,C54,C35, C21L,C7,C6,C44,C26,C25	78.5	85.5	73.9	70.9	83.8
2. Char2 (27)		x[c]			x	17	C27,C9,C34,C17,C54,C35, C21L,C7,C6,C44,C16,C12L, C26,C25,C49L,C1,C52	80.3	85.5	74.9	76.4	84.3
3. Char2 (27)			x[c]		x	17	C5,C17,C44,C26,C16,C53, C34,C52,C9,C27,C7,C12L, C21L,C35,C1,C50,C23ASIN	81.9	88.0	77.9	76.9	84.8
4. Char2A (11)	x			x		11	Full model	78.4	91.0	72.9	64.8	84.8
5. Char2A (11)		x[bc]	x[bc]		x	11	Full model	80.0	92.5	75.4	66.8	85.4
6. Char2B (12)	x			x		10	C27,C9,C17,C23ASIN,C26, C11ASIN,C44,C25,C63ASIN, C24ASIN	76.9	82.5	74.9	68.3	81.8

Table 5. Classification accuracies from linear discriminant analysis of chinook scales using characters selected by Wilks' Lambda and the Mahalanobis selection criteria from a large character set (Char2), a non-ratio character set (Char2A) and a ratio character set (Char2B). All character sets include transformed characters. Characters are described on Tables 1 and 2. Character selection subject to the constraints: F-to-enter ≥4.000, F-to-remove <3.996 except where otherwise noted - cont'd. A. Brood year 1973. B. Brood year 1977.

Character set name (no. of characters)	Selection method[a]			Tolerance		Number of characters selected	Subset selected	Percent correctly classified				
	W/J	W	M	.010	.001			Unweighted overall average	ASIA	WEST	CENT	SEBC
B.Brood year 1977												
7. Char2B (12)		x[bc]	x[bc]		x	11	C27,C9,C17,C23ASIN,C26, C11ASIN,C44,C25,C63ASIN, C21L,C24ASIN	78.0	84.5	75.9	72.4	79.3

[a] Selection method:

W/J minimize Wilks' Lambda and use the jackknife (leaving-one-out approach) to calculate LDF and nearly unbiased classification accuracies.

W minimize Wilks' Lambda and use all cases to calculate the LDF and biased classification accuracies.

M maximize Mahalanobis distance between two closest groups and use all cases to calculate the LDF and biased classification accuracies.

[b] Selected subset and resulting classification accuracies are identical using both methods of character selection.

[c] Estimated probabilities were used as a stopping rule by maintaining a fixed significance level: probability of F-to-enter ≤.05, probability of F-to-remove >.10 (instead of a fixed F-value).

Table 6. Classification accuracies from linear discriminant analysis of chinook scales using characters selected by Wilks' Lambda and the Mahalanobis selection criteria from character sets with freshwater and marine triplets (Char3A), freshwater and marine triplets, and circuli counts (Char3B), freshwater and marine triplets, and large distance characters (Char3C), and freshwater and marine triplets, circuli counts, and large distance characters (Char1A). Character selection subject to the constraints: F-to-enter ≥4.000, F-to-remove <3.996 and minimum tolerance = .010. A. Brood year 1973. B. Brood year 1977.

Character set name (no. of characters)	Selection method[a]			Number of characters selected	Subset selected	Percent correctly classified				
	W/J	W	M			Unweighted overall average	ASIA	WEST	CENT	SEBC
A. Brood year 1973										
1. Char3A (7)	x			5	C54,C50,C49,C53,C52	56.8	59.3	57.6	38.8	71.6
2. Char3A (7)		x[b]	x[b]	5	Same subset as A.1.	57.0	61.0	58.1	39.6	72.2
3. Char3B (9)	x			7	C12,C54,C16,C50,C49,C53, C52	60.5	61.9	59.6	44.0	76.3
4. Char3B (9)		x[b]	x[b]	7	Same subset as A.3.	61.7	64.4	60.6	45.5	76.3
5. Char3C (9)	x			8	C54,C50,C5,C49,C53,C1, C39,C52	56.3	61.9	62.1	35.8	65.5
6. Char3C (9)		x[b]	x[b]	8	Same subset as A.5.	57.6	64.4	62.6	37.3	66.0
7. Char1A (11)	x			10	C12,C16,C5,C50,C49,C52, C51,C53,C1,C39	69.7	57.6	76.8	61.9	82.5

Table 6. Classification accuracies from linear discriminant analysis of chinook scales using characters selected by Wilks' Lambda and the Mahalanobis selection criteria from character sets with freshwater and marine triplets (Char3A), freshwater and marine triplets, and circuli counts (Char3B), freshwater and marine triplets and large distance characters (Char3C), and freshwater and marine triplets, circuli counts and large distance characters (Char1A). Character selection subject to the constraints: F-to-enter ≥4.000, F-to-remove <3.996 and minimum tolerance = .010 - cont'd. A. Brood year 1973. B. Brood year 1977.

Character set name (no. of characters)	Selection method[a]			Number of characters selected	Subset selected	Percent correctly classified				
	W/J	W	M			Unweighted overall average	ASIA	WEST	CENT	SEBC
B. Brood year 1977										
1. Char3A (7)	x			6	C54,C50,C39,C53,C51, C52	55.5	71.0	44.2	34.2	72.7
2. Char3A (7)		x[b]	x[b]	6	Same subset as B.1.	56.4	72.0	44.7	36.2	72.7
3. Char3B (9)	x			7	C16,C12,C54,C50,C39, C51,C53	67.1	92.0	56.3	43.7	76.3
4. Char3B (9)		x[b]	x[b]	7	Same subset as B.3.	68.8	93.0	58.3	46.2	77.8
5. Char3C (9)	x			8	C5,C54,C1,C50,C52,C39, C53,C51	68.2	91.0	68.8	45.7	67.2
6. Char3C (9)		x[b]	x[b]	8	Same subset as B.5.	69.4	92.5	69.3	47.2	68.2
7. Char1A (11)	x			11	Full model	79.0	91.0	72.9	67.8	84.3

Table 6. Classification accuracies from linear discriminant analysis of chinook scales using characters selected by Wilks' Lambda and the Mahalanobis selection criteria from character sets with freshwater and marine triplets (Char3A), freshwater and marine triplets, and circuli counts (Char3B), freshwater and marine triplets, and large distance characters (Char3C), and freshwater and marine triplets, circuli counts and large distance characters (Char1A). Character selection subject to the constraints: F-to-enter ≥4.000, F-to-remove <3.996 and minimum tolerance = .010 - cont'd. A. Brood year 1973. B. Brood year 1977.

[a] Selection method:

W/J minimize Wilks' Lambda and use the jackknife (leaving-one-out approach) to calculate LDF and nearly unbiased classification accuracies.

W minimize Wilks' Lambda and use all cases to calculate the LDF and biased classification accuracies.

M maximize Mahalanobis distance between two closest groups and use all cases to calculate the LDF and biased classification accuracies.

[b] Selected subset and resulting classification accuracies are identical using both methods of character selection.

Table 7. Classification accuracies from linear discriminant analysis of chinook scales using characters selected by Wilks' Lambda and the Mahalanobis selection criteria from character sets with freshwater average circulus spacing (ACS, Char4A), marine ACS (Char4B), overall ACS (Char4C), freshwater and marine ACS (Char4D), and freshwater, marine and overall ACS (Char4E). All these character sets include freshwater and marine triplets. Character selection subject to the constraints: F-to-enter ≥4.000, F-to-remove <3.996 and minimum tolerance = .010. A. Brood year 1973. B. Brood year 1977.

Character set name (no. of characters)	Selection method[a]			Number of characters selected	Subset selected	Percent correctly classified				
	W/J	W	M			Unweighted overall average	ASIA	WEST	CENT	SEBC
A. Brood year 1973										
1. Char4A (8)	x			7	C54,C50,C17,C39,C49, C53,C52	60.4	58.5	57.1	45.5	80.4
2. Char4A (8)		x[b]	x[b]	7	Same subset as A.1.	61.2	60.2	58.1	46.3	80.4
3. Char4B (8)	x			6	C21,C50,C49,C52,C53,C51	61.0	46.6	78.3	46.3	72.7
4. Char4B (8)		x[b]	x[b]	6	Same subset as A.3.	61.4	46.6	78.3	47.0	73.7
5. Char4C (8)	x			7	C9,C50,C49,C52,C39,C53, C51	59.7	55.1	69.2	36.6	77.8
6. Char4C (8)		x[b]	x[b]	7	Same subset as A.5.	61.2	58.5	70.7	37.3	78.4
7. Char4D (9)	x			8	C21,C50,C49,C52,C17,C39, C53,C51	66.0	47.5	76.8	56.7	83.0
8. Char4D (9)		x[b]	x[b]	8	Same subset as A.7.	66.5	48.3	77.3	56.7	83.5

Table 7. Classification accuracies from linear discriminant analysis of chinook scales using characters selected by Wilks' Lambda and the Mahalanobis selection criteria from character sets with freshwater ACS (Char 4A) marine ACS (Char4B), overall ACS (Char4C), freshwater and marine ACS (Char4D), and freshwater, marine, and overall ACS (Char4E) All these character sets include freshwater and marine triplets. Character selection subject to the constraints: F-to-enter ≥4.000, F-to-remove <3.996 and minimum tolerance = .010 - cont'd. A. Brood year 1973. B. Brood year 1977.

Character set name (no. of characters)	Selection method[a]			Number of characters selected	Subset selected	Percent correctly classified				
	W/J	W	M			Unweighted overall average	ASIA	WEST	CENT	SEBC
A. Brood year 1973										
9. Char4E (10)	x			8	C50,C49,C17,C52,C39,C21, C53,C51	66.0	47.5	76.8	56.7	83.0
10.Char4E (10)		x[b]	x[b]	8	Same subset as A.9.	66.5	48.3	77.3	56.7	83.5
B. Brood year 1977										
1. Char4A (8)	x			7	C54,C17,C39,C50,C53,C51, C52	63.4	67.5	49.7	54.8	81.8
2. Char4A (8)		x[b]	x[b]	7	Same subset as B.1.	65.1	67.5	53.8	55.8	83.3
3. Char4B (8)	x			8	Full model	64.1	66.0	66.3	45.7	78.3
4. Char4B (8)		x[b]	x[b]	8	Full model	65.1	68.0	66.3	47.7	78.3
5. Char4C (8)	x			7	C9,C50,C54,C39,C52,C51, C53	63.7	68.5	58.8	44.7	82.8

Table 7. Classification accuracies from linear discriminant analysis of chinook scales using characters selected by Wilks' Lambda and the Mahalanobis selection criteria from character sets with freshwater ACS, (Char4A), marine ACS (Char4B), overall ACS (Char4C), freshwater and marine ACS (Char4D), and freshwater, marine and overall ACS (Char4E). All these character sets include freshwater and marine triplets. Character selection subject to the constraints: F-to-enter ≥4.000, F-to-remove <3.996 and minimum tolerance = .010 - cont'd. A. Brood year 1973. B. Brood year 1977.

Character set name (no. of characters)	Selection method[a]			Number of characters selected	Subset selected	Percent correctly classified				
	W/J	W	M			Unweighted overall average	ASIA	WEST	CENT	SEBC
B. Brood year 1977										
6. Char4C (8)		x[b]	x[b]	7	Same subset as B.6.	65.0	70.5	58.8	47.7	82.8
7. Char4D (9)	x			9	Full model	70.7	66.5	72.9	59.3	84.3
8. Char4D (9)		x[b]	x[b]	9	Full model	71.6	68.0	73.4	60.3	84.8
9. Char4E (10)	x			10	Full model	71.6	71.0	72.4	58.8	84.3
10. Char4E (10)		x[b]	x[b]	10	Full model	72.9	71.5	72.4	62.3	85.4

[a] Selection method:

W/J minimize Wilks' Lambda and use the jackknife (leaving-one-out approach) to calculate LDF and nearly unbiased classification accuracies.

W minimize Wilks' Lambda and use all cases to calculate the LDF and biased classification accuracies.

M maximize Mahalanobis distance between two closest groups and use all cases to calculate the LDF and biased classification accuracies.

[b] Selected subset and resulting classification accuracies are identical using both methods of character selection.

Table 8. Classification accuracies from linear discriminant analysis of chinook scales using characters selected by Wilks' Lambda and the Mahalanobis selection criteria from character sets that include triplets without life history information (T, Char5A), T and overall circuli counts (Char5B), T and overall large distances (Char5C), and T, overall circuli counts and large distances (Char5D). Character selection subject to the constraints: F-to-enter ≥4.000, F-to-remove <3.996 and minimum tolerance = .010. A. Brood year 1973. B. Brood year 1977.

Character set name (no. of characters)	Selection method^a				Number of characters selected	Subset selected	Percent correctly classified				
	W/J	W	M	F/J			Unweighted overall average	ASIA	WEST	CENT	SEBC
A. Brood year 1973											
1. Char5A (8)	x				5	TR8,TR3,TR5,TR7,TR1	42.5	35.6	30.8	44.8	58.8
2. Char5A (8)		x^b	x^b		5	Same subset as A.1.	44.1	38.1	32.8	45.5	59.8
3. Char5B (9)	x				4	C7,TR5,TR8,TR3	48.9	64.4	34.3	26.1	70.6
4. Char5B (9)		x^b	x^b		4	Same subset as A.3.	49.2	64.4	34.8	26.9	70.6
5. Char5C (9)	x				5	C6,TR8,TR3,TR5,TR7	45.4	44.9	37.4	38.1	61.3
6. Char5C (9)		x^b	x^b		5	Same subset as A.5.	47.0	45.8	38.9	40.3	62.9
7. Char5D (10)	x				7	C7,C6,TR5,TR3,TR6,TR4,TR7	60.7	56.8	66.2	45.5	74.2
8. Char5D (10)		x^b	x^b		7	Same subset as A.7.	61.6	59.3	66.2	46.3	74.7

Table 8. Classification accuracies from linear discriminant analysis of chinook scales using characters selected by Wilks' Lambda and the Mahalanobis selection criteria from character sets that include triplets (T, Char5A), T and overall circuli counts (Char5B), T and overall large distances (Char45C), and T, overall circuli counts and large distances (Char5D). Character selection subject to the constraints: F-to-enter ≥4.000, F-to-remove <3.996 and minimum tolerance = .010 - cont'd. A. Brood year 1973. B. Brood year 1977.

Character set name (no. of characters)	Selection method[a]				Number of characters selected	Subset selected	Percent correctly classified				
	W/J	W	M	F/J			Unweighted overall average	ASIA	WEST	CENT	SEBC
A. Brood year 1973											
9. Char5D (10)				x	10	Full model	61.0	54.2	67.7	46.3	75.8
B. Brood year 1977											
1. Char5A (9)	x				8	TR9,TR1,TR4,TR8,TR2,TR3, TR7,TR5	54.1	72.0	39.2	39.2	66.2
2. Char5A (9)		x[b]	x[b]		8	Same subset as B.1.	55.3	73.0	39.7	40.2	68.2
3. Char5B (10)	x				8	C7,TR9,TR1,TR6,TR4,TR2, TR8,TR5	62.8	90.0	47.2	42.7	71.2
4. Char5B (10)		x[b]	x[b]		8	Same subset as B.3.	64.2	90.0	49.2	45.2	72.2
5. Char5C (10)	x				8	C6,TR9,TR4,TR2,TR8,TR7, TR5,TR6	63.4	87.0	56.3	48.7	61.6
6. Char5C (10)		x[b]	x[b]		8	Same subset as B.5.	64.7	87.5	57.8	49.2	64.1

Table 8. Classification accuracies from linear discriminant analysis of chinook scales using characters selected by Wilks' Lambda and the Mahalanobis selection criteria from character sets that include triplets (T, Char5A), T and overall circuli counts (Char5B), T and overall large distances (Char5C), and T, overall circuli counts and large distances (Char5D). Character selection subject to the constraints: F-to-enter ≥4.000, F-to-remove <3.996 and minimum tolerance = .010 - cont'd. A. Brood year 1973. B. Brood year 1977.

Character set name (no. of characters)	Selection method[a]				Number of characters selected	Subset selected	Percent correctly classified				
	W/J	W	M	F/J			Unweighted overall average	ASIA	WEST	CENT	SEBC
B. Brood year 1977											
7. Char5D (11)	x				10	C7,C6,TR6,TR8,TR2,TR5, TR7,TR9,TR1,TR3	73.1	91.0	72.9	51.8	76.8
8. Char5D (11)		x[b]	x[b]		10	Same subset as B.7.	74.0	91.5	74.4	53.3	76.8
9. Char5D (11)				x	11	Full model	74.1	92.0	73.4	52.8	78.3

[a] Selection method:

W/J minimize Wilks' Lambda and use the jackknife (leaving-one-out approach) to calculate LDF and nearly unbiased classification accuracies.

W minimize Wilks' Lambda and use all cases to calculate the LDF and biased classification accuracies.

M maximize Mahalanobis distance between two closest groups and use all cases to calculate the LDF and biased classification accuracies.

F/J Force all characters into the model subject to specified tolerance level. Jackknife approach to calculate LDF and nearly unbiased classification accuracies.

[b] Selected subset and resulting classification accuracies are identical using both methods of character selection.

Table 9. Classification accuracies from linear discriminant analysis of chinook scales using small distance characters, circuli doublets (Char6A), circuli quadruplets (Char6B) and circuli quintuplets (Char6C). Models selected by the Wilks' Lambda selection criterion. Character selection subject to the constraints: F-to-enter ≥4.000, F-to-remove <3.996 and minimum tolerance = .010. A. Brood year 1973. B. Brood year 1977.

Character set name (no. of characters)	Selection method[a]		Number of characters selected	Subset model	Percent correctly classified				
	W/J	F/J			Unweighted overall average	ASIA	WEST	CENT	SEBC
A. Brood year 1973									
1. Char6A (15)	x		13	C12,C16,C5,MD1,MD3,MD5, MD6,MD2,MD7,MD4,C1, FWD2,FWD1	71.6	64.4	75.3	64.2	82.5
2. Char6A (15)		x	15	Full model	70.9	63.6	74.2	62.7	83.0
3. Char6B (9)	x		9	Full model	72.1	64.4	78.3	62.7	83.0
4. Char6C (8)	x		8	Full model	71.0	61.9	76.3	62.7	83.0
B. Brood year 1977									
1. Char6A (15)	x		13	C16,C5,MD3,C1,C12,MD5, MD9,FWD1,MD2,MD7,MD6, MD4,MD8	78.8	91.0	73.4	68.3	82.3
2. Char6A (15)		x	15	Full model	78.4	91.0	74.4	66.3	81.8

Table 9. Classification accuracies from linear discriminant analysis of chinook scales using small distance characters, circuli doublets (Char6A), circuli quadruplets (Char6B) and circuli quintuplets (Char6C). Models selected by the Wilks' Lambda selection criterion. Character selection subject to the constraints: F-to-enter ≥4.000, F-to-remove <3.996 and minimum tolerance = .010 - cont'd. A. Brood year 1973. B. Brood year 1977.

Character set name (no. of characters)	Selection method[a]		Number of characters selected	Subset model	Percent correctly classified				
	W/J	F/J			Unweighted overall average	ASIA	WEST	CENT	SEBC
B. Brood year 1977									
3. Char6B (9)	x		9	Full model	78.1	91.0	72.4	66.3	82.8
4. Char6C (8)	x		8	Full model	78.0	90.5	72.9	66.3	82.3

[a]Selection method:

W/J minimize Wilks' Lambda and use the jackknife (leaving-one-out approach) to calculate LDF and nearly unbiased classification accuracies.

F/J Force all characters into the model subject to specified tolerance level. Jackknife approach to calculate LDF and nearly unbiased classification accuracies.

Table 10. Classification accuracies from linear discriminant analysis of chinook scale characters and non-discriminating dummy characters. Character set Char7A-Char7D include Char1A and 1-4 dummy variables, respectively. Char7E includes freshwater and marine CC and LD characters, and Char7F includes these CC and LD characters, and 4 dummy variables. Char7G includes freshwater and marine triplets and Char7H includes these triplets and 4 dummy variables. Characters were forced into models subject to the constraint: tolerance = .010.
A. Brood year 1973. B. Brood year 1977.

		Jackknife classifications: percent correct				
Character set name	Number of characters in model	Unweighted overall accuracy	ASIA	WEST	CENT	SEBC
A. Brood year 1973						
1. Char7A	12	69.0	56.8	76.3	60.4	82.5
2. Char7B	13	69.2	55.9	77.3	61.9	81.4
3. Char7C	14	68.8	55.9	76.3	60.4	82.5
4. Char7D	15	68.7	55.9	77.3	59.7	82.0
5. Char7E	4	59.2	55.1	69.7	37.3	74.7
6. Char7F	8	58.7	52.5	69.7	36.6	75.8
7. Char7G	7	55.9	59.3	55.6	38.1	70.6
8. Char7H	11	54.6	57.6	54.5	37.3	69.1
B. Brood year 1977						
1. Char7A	12	79.4	91.5	72.9	68.8	84.3
2. Char7B	13	79.3	91.5	72.9	68.3	84.3
3. Char7C	14	79.1	91.5	72.4	68.3	84.3
4. Char7D	15	79.1	91.0	72.4	68.8	84.3
5. Char7E	4	73.0	87.5	68.3	63.3	72.7
6. Char7F	8	72.0	87.5	67.8	59.3	73.2
7. Char7G	7	54.8	70.0	42.7	34.7	71.7
8. Char7H	11	53.5	69.0	41.7	32.2	71.2