IDENTIFICATION OF STOCKS OF BRISTOL BAY SOCKEYE SALMON, *ONCORHYNCHUS NERKA*, BY EVALUATING SCALE PATTERNS WITH A POLYNOMIAL DISCRIMINANT METHOD[1]

RODNEY C. COOK[2] AND GARY E. LORD[3]

ABSTRACT

A polynomial discriminant method is developed for the racial classification of stocks of sockeye salmon. The method is based upon the nonparametric estimation of the multivariate probability densities of the scale characteristics for each stock considered. Errors in classification are examined and a correction procedure is extended to the n-class case. As an example, sockeye salmon of age 2.2 sampled on the high seas are classified to river of origin based on freshwater scale growth patterns. Also, freshwater and marine scale characters are evaluated for stock identification purposes involving certain Bristol Bay runs.

Racial analysis of high-seas salmon has important applications both in life history studies of various stocks and in management considerations of these stocks. As a result, many have examined the characteristics of scale structure to differentiate salmon subpopulations. Konovalov (1971) notes that some investigators were ignoring many characteristics in scale structure which arise under the effects of ecological factors in specific bodies of water. When the ecological conditions affecting scale characters are seriously considered, statistically significant differences between subpopulations can often be found. The ability to recognize salmon subpopulations depends upon the differences between the stocks in terms of examined characteristics and the accuracy of the analytic technique. Various discriminant function analyses have been traditionally used.

Fukuhara et al. (1962), Amos et al. (1963), and Dark and Landrum (1964) used linear discriminant functions based upon morphological characteristics to identify the continent of origin of Pacific salmon. Scale characteristics and linear discriminant functions were used by Anas (1964) and Mason (1966). Anas and Murai (1969) used linear and quadratic discriminant functions. Recent investigations by Major et al. (1975) and Bilton and Messinger (1975) used unspecified discriminant function techniques, probably similar to those of Anas and Murai (1969). These and other studies show the utility of discriminant function methodology for identifying races of Pacific salmon.

Salmon managers need a flexible and easily implemented stock identification technique. This paper applies a generalized discriminant function technique to measurements of sockeye salmon scales to attempt to fulfill this need.

DISCRIMINANT FUNCTION ANALYSES OR PATTERN RECOGNITION[4]

Discriminant function analysis depends on the recognition of underlying patterns differing among classes of objects. In this case, scale patterns characterize a sockeye salmon of a particular origin. A set of p-scale characters (a p-tuple or vector in p-space) measured on an individual salmon provides a description of that salmon. A sample of p-tuples for a number of salmon from one origin (the learning sample) establishes a region in p-space characteristic of that class of sockeye. Samples from salmon of different and known origins establish regions in p-space which may be separated by decision surfaces. A sockeye salmon of unknown origin may be classified according to which region its p-tuple occupies. The accuracy of classification depends upon the precision with

[1]Contribution No. 478, College of Fisheries, University of Washington, Seattle, Wash.

[2]Fisheries Research Institute, University of Washington, Seattle, WA 98195.

[3]Applied Physics Laboratory, University of Washington, Seattle, WA 98195.

[4]A good text on pattern recognition is given by Patrick (1972). A review of the literature is given by Das Gupta (1973).

Manuscript accepted November 1977.

which the regions are described and the inherent separation between them.

These regions are described mathematically by multivariate probability density functions. Fisher's (1936) linear discriminant function defines a linear decision surface (hyperplane) derived by describing these regions as multivariate normal density distributions with common variance-covariance matrices (Welch 1939). Quadratic discriminant functions have been developed (Smith 1947). The resulting decision surfaces are nonlinear. The quadratic discriminant function does not require common variance-covariance matrices. Anas and Murai (1969) compared the classificatory abilities of the linear and quadratic discriminant functions. They found (in agreement with Isaacson 1954) that even if the assumption that the distributions have common variance-covariance matrices is violated, the linear discriminant function would still give good results for large sample sizes. But the quadratic function gave slightly better results.

All investigators utilizing discriminant analyses to separate races of Pacific salmon have assumed that the density distributions of measurements from a particular class of salmon were multivariate normal. The frequency distributions of scale characters in Major et al. (1975) show that multimodal and skewed distributions occur for chinook salmon scale characters even in the univariate case. In many other cases, the underlying distribution functions may be non-Gaussian. Discriminant functions based upon non-Gaussian distributions or obtained by distribution-free methods are preferable to those based upon an unrealized assumption of normality.

Nearly all of the discriminant function analyses used in the investigations of Pacific salmon have been two-class analyses designed to determine the continent of origin of salmon taken on the high seas. For the two-class situation only one discriminant function need be calculated. These two-class problems are a special case of the many-class problems in which a separate discriminant function is calculated for each class. Bilton and Messinger (1975) calculated discriminant functions for each of several runs in a classification study on sockeye salmon. If several stocks of salmon intermingle and are to be classified, analyses of this type are needed.

Specht's (1966) polynomial discriminant method does not require that the underlying density distributions be multivariate normal nor that they have common variance-covariance matrices. Since this method is nonparametric, various scale characteristics may be used for discrimination with no particular regard to the underlying distributions. Thus, the method is flexible and practical.

Specht (1966) uses an estimated probability density function of the form described by Parzen (1962) and extended by Murthy (1966) to the multivariate case. The underlying multivariate density for each class is modeled by a sum of functions that are multivariate Gaussian in form, one such function for each fish in the learning sample for that class. This set of functions is complete. Therefore, for each class the underlying continuous probability density, Gaussian or not, may be approximated arbitrarily closely by such a sum. A power series expansion of this estimated denisty then results in a polynomial term in the density function, the coefficients of which are functions of the observations (fish) in the learning sample. One such set of coefficients is computed for each class to be considered. These polynomials determine the nonlinear decision surfaces and are the basis for discrimination.

The individual multivariate Gaussian functions (which when summed model the underlying multivariate distribution for that class) contain a "smoothing parameter," σ, which appears in the place of a standard error. This parameter is then incorporated in the estimates of the polynomial coefficients. The reader is referred to Specht (1966) for a discussion of the effect of this smoothing parameter and for the algorithm for the calculation of the sets of polynomial coefficients $\{D_{k_i \ldots k_j \ldots k_h}\}$. The polynomial discriminant function is:

$$\begin{aligned} P(X) = {} & D_0 + D_1 X_1 + D_2 X_2 + \ldots + D_p X_p \\ & + D_{11} X_1{}^2 + \ldots + D_{k_1 k_2} X_{k_1} X_{k_2} \\ & + \ldots + D_{pp} X^2{}_p + D_{111} X_1{}^3 \\ & + \ldots + D_{k_1 k_2 k_3} X_{k_1} X_{k_2} X_{k_3} + \ldots \\ & + D_{ppp} X^3{}_p + \ldots \\ & + D_{k_1 \ldots k_j \ldots k_h} X_{k_1} \ldots X_{k_h} \\ & + \ldots, \end{aligned}$$

where p = dimension of the vector X (set of scale characters)

$1 \leq k_j \leq p$
$j = 1, 2, \ldots, h$
h = the degree of the variable portion of the term.

The decision on an unknown X (set of scale measurements from a salmon of unknown origin) is thus:

$$\text{Choose } d(X) = \theta_r \text{ so that } h_r P^r(X) \geq h_s P^s(X) \text{ for all } s \neq r$$

where $d(X)$ = the decision on an unknown X
θ_i's = the classes (origins)
$p^i(X)$ = the polynomial value for X calculated using the discriminant function for class θ_i
h_i = the a priori probability, the uses of which will be described later.

APPLICATION OF THE METHOD

Three scale sample sets are required to implement the polynomial discriminant method: learning samples, test samples, and unknown samples. The learning and testing samples are collected from each subpopulation when they are segregated (i.e., in the rivers of origin). Scale characters to be measured in the unknown sample for the required discrimination are determined by evaluating characters measured in the learning samples. The learning samples and the characters selected are used to calculate the coefficients in the polynomial discriminant functions. To calculate these coefficients, the value for the smoothing parameter and the point at which the discriminant function should be truncated must be determined. Various circumstances will dictate different choices. When a smoothing parameter of 1.5 was chosen, all terms in the discriminant function greater than the fourth order contributed negligibly to polynomial values and so were truncated in our applications. Often, polynomial discriminant functions of lower order yield adequate results.[5]

[5]A polynomial discriminant function with six variables and of the fourth order will contain 210 terms. Since our calculations were performed by computer, we chose not to delete the third or fourth degree terms. However, if more than six variables are used, it would be wise to truncate further in order to keep the number of terms down.

The fish comprising the test samples are classified to test the effectiveness of the polynomial discriminant method and to determine the a priori probabilities. (Each test sample consists of fish from one class.) Finally, fish collected from the zone of intermingling are classified to determine the degree of intermingling in the area of interest.

Appraisal of the method using scale samples of sockeye salmon collected from the 1967 escapement in five Bristol Bay rivers showed large percentages of fish comprising the test samples were correctly classified. However, misclassified fish in the test group (set of test samples from all rivers being considered) were not assigned to the rivers in proportion to the known relative test sample sizes. To balance these misclassifications, wherever a greater number of fish comprising the test group was assigned to a particular river than should have been (according to the relative test sample sizes), the a priori probability for that river was lowered. Corresponding increases were made for those classes with insufficient assignment. By alternatively using the decision procedure of the polynomial discriminant method and adjusting the a priori probabilities, we obtained solutions so that the number of fish belonging to a certain river that were misassigned to all other rivers approximately equaled the number of fish misassigned to that certain river from all other rivers. Thus the a priori probabilities were not used in the manner their name suggests, but a priori knowledge may dictate test sample sizes. The relative test sample sizes in the test group may be in the relative proportions to be expected in the unknown sample (i.e., historical relative run sizes). The adjustment procedure, then, shifts the nonlinear decision surfaces between the probability densities so that the incorrectly identified samples are assigned to the various rivers in the proportions dictated by the test sample sizes in the test group. However, the primary purpose of the adjustment procedure is not to balance the misclassifications but to maximize the number of correct classifications. As the misclassifications are balanced, the number of correct classifications generally increases. At this point the result is a classification method that maximizes the total number of correct classifications and balances misclassification rates for a test group in which the test sample sizes are in particular proportions.

However, it is obvious that the proportions of fish from the various classes in the test group would rarely be identical to those proportions in the unknown sample. Thus, imbalance among the misclassified fish will recur, unless the expected accuracy of classification is very good (near 100%). We have devised a method to correct for this.

Based upon the results of classification of the known test group, the classification matrix, C, is estimated:

$$\hat{C} = \begin{bmatrix} \hat{c}_{11} & \hat{c}_{12} & \cdots & \hat{c}_{1n} \\ \hat{c}_{21} & \hat{c}_{22} & \cdots & \hat{c}_{2n} \\ \vdots & & & \vdots \\ \hat{c}_{n1} & \hat{c}_{n2} & \cdots & \hat{c}_{nn} \end{bmatrix},$$

where $\hat{c}_{ij}$ is an estimate of the fraction of fish allocated to class i belonging to class j, such that $\sum_{i=1}^{n} \hat{c}_{ij} = 1.0, \forall j$. (Note that for each j the $\hat{c}_{ij}$'s are a set of estimated multinomial probabilities and that each test sample size should be adequate.) If the discrimination is error-free, C would be an identity matrix. The adjustment of a priori probabilities causes the initially estimated classification matrix to evolve to the point where

$$\hat{C}T = R_t \text{ such that } T \approx R_t.$$

The ith component of the vector T is the fraction of fish in the test group from test sample i (class i), and the ith component of the vector R_t is the fraction of fish in the test group allocated to class i by the adjusted polynomial discriminant method.

The test samples comprising T are not independent of the classification scheme since they are used to determine the a priori probabilities used in the decision rule. Hence, the estimated probabilities in the classification matrix may not be unbiased. However, we did chi-square tests that show elements of the classification matrix are not significantly different when estimated with either the test samples used to determine the a priori probabilities or a second independent test group. Thus, we prefer to use only one test group to determine the a priori probabilities and to estimate the elements of the classification matrix because the test sample sizes will be larger (and the variance of the $\hat{c}_{ij}$'s smaller) if we do not subdivide the fish available.

Now, let u_i be the fraction of fish in a sampled group that belong to the ith class. The vector U is then unknown except for the obvious side condition $\sum_{i=1}^{n} u_i = 1$. The classification matrix now operates on U to give:

$$CU = R_u \text{ [6]}$$

where the ith component of R_u is the fraction of fish in the unknown sample allocated to the ith class. Since C is estimated, R_u is known and since C is usually nonsingular, we can estimate U by

$$\hat{U} = \hat{C}^{-1} R_u.$$

Each point estimate ($\hat{u}_i$) obtained will have some variance. This variability will depend upon the accuracy with which fish from class i are classified, the accuracy with which the elements of C are estimated, and variance due to sampling error encountered when obtaining the unknown sample. Thus, if any u_i is small, then its estimate ($\hat{u}_i$) may be negative. Such solutions are meaningless. In such cases the classes with negative solutions should be dropped (assume such $u_i \approx 0$) and the analyses repeated.

We did simulation work to evaluate the classification matrix correction procedure for the two- and three-class situations. Five hundred simulated experiments were done for each situation. For the two-class case the average error of the classification results was 0.100 while that of the corrected estimates was 0.055. In 84% of the experiments the corrected estimate was closer to the true value than classification result. For the three-class case the average error of the classification results was 0.127 while that of the corrected estimates was 0.054. In 89% of the experiments the corrected estimate was closest to the true value. The results of these simulations show that the classification correction procedure improves estimates of the true proportion of a class present.

This classification matrix correction procedure will reduce to the correction procedure developed for the two-class case by Worlund[7] in the following manner:

[6]A similar relationship and a least squares solution technique is given by Worlund and Fredin (1962).

[7]Worlund, D. D. 1960. A method for computing the variance of an estimate of the rate of intermingling of two salmon populations. Unpubl. manuscr., 13 p. Bur. Commer. Fish., Biol. Lab., Seattle, Wash.

$$U = C^{-1} R_u$$

or

$$\begin{bmatrix} u_1 \\ \\ u_2 \end{bmatrix} = \begin{bmatrix} \dfrac{c_{22}r_1 - c_{12}r_2}{c_{11}c_{22} - c_{21}c_{12}} \\ \\ \dfrac{c_{11}r_2 - c_{21}r_1}{c_{11}c_{22} - c_{21}c_{12}} \end{bmatrix}$$

Generally $$u_i = \frac{r_i\, c_{jj} - c_{ij}r_j}{c_{ii}c_{jj} - c_{ji}c_{ij}}.$$

Since $$r_j = 1 - r_i,$$
$$c_{ji} = 1 - c_{ii},$$
$$c_{jj} = 1 - c_{ij},$$

substitution yields $$u_i = \frac{r_i - c_{ij}}{c_{ii} - c_{ij}},$$

which is the correction formula of Worlund and Fredin (1962) (except for differences in notation and terminology) that has been used in many two-class Pacific salmon stock identification studies.

Application to Sockeye Salmon Samples Taken in High Seas Sampling

A problem of interest to the nations bordering the North Pacific Ocean is the origin of sockeye salmon taken on the high seas. The rivers of origin of sockeye salmon south of the central Aleutian Islands in summer are of particular interest to the United States since an index of their overall relative abundance is used to forecast the numbers of mature fish returning to Bristol Bay in the following year (Rogers 1975). These fish are primarily of Bristol Bay origin (Hartt 1962, 1966; Hartt et al. 1975). Knowledge of the relative abundance of the various runs of the Bristol Bay stock south of the central Aleutians would be useful for forecast purposes and might provide insight into the high seas life history of the various runs.

In order to recognize age 2.2 immature sockeye salmon on the high seas in 1976, the freshwater growth patterns of scales from three of the major rivers in Bristol Bay were examined.[8] Scales from the smolt outmigrations of 1974 for the Kvichak and Naknek Rivers were used as learning and testing samples. For the Egegik River scales from age 2.2 adult fish returning to spawn in 1976 were used as learning and testing samples because smolt scales were unavailable. The freshwater scale patterns of fish from these runs were used to classify the sockeye salmon captured south of Adak Island during summer 1976 after having spent two winters in the ocean.

The scale patterns were examined under a microprojector of the type described by Dahlberg and Phinney (1968). The widths of the summer, winter, and plus growth zones were measured in terms of circuli counts and distance. The width of the widest circulus was also measured. Each scale character was then ranked over all classes (rivers) and the Kruskal-Wallis statistic (Kruskal and Wallis 1952) calculated. The difference between the average sum of ranks for each pairwise class combination was also calculated. On the basis of these statistics the scale characters providing the best univariate separation were selected for use in the polynomial discriminant method. Highly dependent scale characters were not used.

By examining the learning samples, six scale characteristics were chosen for use in the polynomial discriminant method: 1) The number of the circuli in the first winter growth zone, 2) the number of circuli in the second summer growth zone, 3) the number of circuli in the plus growth zone, 4) the width of the first summer growth zone, 5) the width of the second winter growth zone, and 6) the width of the widest circulus.[9] Learning sample sizes of 25, 25, and 24 for the Egegik, Kvichak, and Naknek River classes, respectively, were used to calculate the coefficients in the polynomial function for each class. The classificatory ability of these functions was then tested.

The relative test sample sizes for each class were determined by examining run size data. According to the average run sizes of age 2.3 salmon for the last 8 yr approximately equal numbers of fish from each class were expected to occur in the unknown sample. However, since the Kvichak River test sample size was twice that of the Egegik or Naknek River sample size, the fish in the latter test samples were given a weight of 2 when the a priori

[8]Age designation indicates fish which migrated to sea after two winters in freshwater and have spent two winters at sea. They are expected to return from the ocean primarily at age 2.3, or after spending three winters at sea.

[9]It should be mentioned that all data points were "normalized." That is, the mean and standard deviation for each scale character were calculated from the learning samples (all categories combined). All data points were then transformed by subtracting off the mean and dividing by the standard deviation for the appropriate scale character. This is done for numerical purposes.

probabilities were adjusted. After adjusting the a priori probabilities, we obtained the results given in Table 1. The classification matrix was then estimated:

$$\hat{C} = \begin{bmatrix} 0.800 & 0.040 & 0.167 \\ 0.080 & 0.740 & 0.208 \\ 0.120 & 0.220 & 0.625 \end{bmatrix}$$

where the subscripts of the matrix elements ($\hat{c}_{ij}$'s) were 1, 2, and 3 for the Egegik, Kvichak, and Naknek River classes, respectively. Seventy-two percent of the fish in the test group were correctly classified. The fish in the high seas sample were then classified with the adjusted polynomial discriminant method.

Of the 101 sockeye salmon, 25 were classified as Egegik River fish, 22 as Kvichak River fish, and 54 as Naknek River fish. The resultant vector was:

$$R_u = \begin{bmatrix} 0.267 \\ 0.222 \\ 0.511 \end{bmatrix}.$$

The estimated unknown vector was thus:

$$\hat{C}^{-1} R_u = \begin{bmatrix} 1.300 & 0.037 & -0.360 \\ -0.078 & 1.498 & -0.478 \\ -0.222 & -0.534 & 1.837 \end{bmatrix} \begin{bmatrix} 0.267 \\ 0.222 \\ 0.511 \end{bmatrix}$$

$$= \begin{bmatrix} 0.171 \\ 0.067 \\ 0.761 \end{bmatrix} = \hat{U}.$$

Based upon preliminary data for the 1977 Bristol Bay sockeye salmon run from the Alaska Department of Fish and Game, the actual unknown vector was:

$$U = \begin{bmatrix} 0.325 \\ 0.061 \\ 0.614 \end{bmatrix}.$$

The classification matrix correction procedure gave a slightly better estimate than the direct results of the polynomial discriminant method. The differences between the u_i's and the $\hat{u}_i$'s were due to bias and variability. (We are presently examining methods to reduce the variability of our $\hat{u}_i$'s.)

A problem with the high seas sample is that some of these sockeye salmon originate in rivers other than those considered. Although the three classes considered will account for nearly all of the age 2.2 sockeye salmon bound for Bristol Bay, some may be non-Bristol Bay fish. When the Bristol Bay runs are at a low point in their cycle, up to 20% of the high seas sockeye salmon at Adak Island may be non-Bristol Bay fish (Hartt et al. 1975). The possible bias from classifying the non-Bristol Bay fish into the classes established should be considered since 1977 is a low year in the sockeye salmon run cycle.

TABLE 1.—Results of the polynomial discriminant method on a known test group of Bristol Bay sockeye salmon. The a priori probabilities were 0.340, 0.332, and 0.328 for the Egegik, Kvichak, and Naknek River classes, respectively.

Calculated decisions	Correct decisions			Total (all calculated decisions)
	Egegik	Kvichak	Naknek	
Egegik	40	2	8	50
Kvichak	4	37	10	51
Naknek	6	11	30	47
Total (all correct decisions)	50	50	48	148

In conclusion, the polynomial discriminant method can be used to identify certain runs of sockeye salmon on the high seas by differences in freshwater scale growth patterns. Possibly the relative proportions of sockeye salmon that will be returning to inshore areas can be predicted. Eventually the method will be used to predict one year in advance the relative run sizes to the major Bristol Bay rivers by sampling these sockeye on the high seas.

Application to Inshore Fishery Stock Separation

A problem of interest to the Alaska Department of Fish and Game is the separation of stocks in commercial catches in inshore areas, particularly the separation of Kvichak, Naknek, and Egegik River sockeye salmon. The Division of Commercial Fisheries is collecting data on scale measurements for growth studies. They are interested in how well these data and the polynomial discriminant method can separate Bristol Bay sockeye salmon stocks.

Scale data from samples of the 1973 spawning escapement were examined. Each of two age-classes was examined separately. Distance and circuli counts to both the freshwater and saltwater annuli were examined for use in the polynomial discriminant method with the Kruskal-Wallis and multiple comparison procedures. The accuracy of classification for age 1.2 and age 2.2 sockeye salmon was examined for each age-group with known test groups.

The degree of separation for age 1.2 sockeye salmon is shown in Table 2. (Egegik River fish are historically insignificant in this age-class.) The scale characters providing this separation were: 1) the circuli count to the first annulus, 2) the distance to the first annulus, 3) the distance from the first to the second annulus, 4) the distance from the second to the third annulus, 5) the circuli count from the third annulus to the edge of the scale, and 6) the distance from the third annulus to the edge of the scale. Ninety-five percent of the fish in the test group were correctly classified.

The degree of separation for age 2.2 sockeye salmon is shown in Table 3. The scale characters providing this separation were: 1) the circuli count to the first annulus, 2) the distance to the first annulus, 3) the circuli count from the first to the second annulus, 4) the distance from the second to the third annulus, 5) the distance from the third to the fourth annulus, and 6) the circuli count from the fourth annulus to the edge of the scale. Seventy-seven percent of the fish in the test group were correctly classified.

Thus, the polynomial discriminant method can provide adequate separation with a given data base. The data collected for growth studies provide good separation in some cases. Sockeye salmon from the Egegik, Kvichak, and Naknek Rivers are distinguishable in terms of these scale measurements and it should be possible to estimate their relative proportions in catch samples.

TABLE 2.—Results of the polynomial discriminant method on 1.2 age Bristol Bay sockeye salmon from 1973. The a priori probabilities were 0.52 and 0.48 for the Kvichak and Naknek River classes, respectively.

Calculated decisions	Correct decisions		Total (all calculated decisions)
	Kvichak	Naknek	
Kvichak	18	0	18
Naknek	2	19	21
Total (all correct decisions)	20	19	39

TABLE 3.—Results of the polynomial discriminant method on 2.2 age Bristol Bay sockeye salmon from 1973. The a priori probabilities were 0.342, 0.330, and 0.328 for the Egegik, Kvichak, and Naknek River classes, respectively.

Calculated decisions	Correct decisions			Total (all calculated decisions)
	Egegik	Kvichak	Naknek	
Egegik	20	3	3	26
Kvichak	1	22	4	27
Naknek	5	1	14	20
Total (all correct decisions)	26	26	21	73

COMMENTS AND CONCLUSIONS

The key to successful implementation of the polynomial discriminant method is the choice of scale characters that reflect differences between the subpopulations of concern. The scale characters that are most likely different are those that are formed when the populations are geographically separated. Genetic and environmental influences on scale formation probably interact to create these differences. Although it is likely that no single characteristic will provide the required separation, a group of characteristics analyzed with multivariate techniques (e.g., the polynomial discriminant method) will often provide this required separation. The polynomial discriminant function technique requires no consideration of the underlying probability density functions for these scale characters because these density functions are estimated nonparametrically. Once the characters that provide the best separation are determined (by rank order comparison procedures in this paper) the discriminant function analysis may be implemented.

A learning sample is needed to calculate the discriminant function for each subpopulation. These fish comprising these samples must be collected before or after the populations intermingle (either as smolts or returning adults in the respective rivers). Learning samples must be taken from the same year class and freshwater age-group as the unknown (mixed) population if the scale characters are known or thought to vary from year to year. Using Specht's (1966) algorithm and the data from these learning samples, the coefficients in the discriminant functions are calculated. The next step is to appraise the effectiveness of these polynomial discriminant functions.

By classifying a group of test samples the proportion of correctly identified fish and the classification error rates can be determined. The proportion of correctly identified fish will likely be low until a good set of a priori probabilities is determined. As the a priori probabilities are adjusted to balance the classification error rates, the proportion of correctly identified fish will generally increase. The proportion of correctly identified fish, when the classification error rates are satisfactorily balanced, gives an indicator of the effectiveness of the polynomial discriminant method. The classification error rates specific to these final a priori probabilities are now estimated so that they may be corrected for when the polynomial discriminant method is applied to the unknown mixed sample. This is done with the classification matrix correction procedure.

First, the fish in the unknown mixed sample are classified with the polynomial discriminant method (using the adjusted a priori probabilities). The proportions resulting for each subpopulation and the decision matrix allow simple algebraic solution for the estimated true proportions of the various subpopulations in the zones of intermingling.

Estimates of this type are often needed in particular management situations involving Pacific salmon. By using scale samples and the polynomial discriminant method, the proportions of the major classes present in areas where the subpopulations mix can be estimated. We have considered only two possible applications in this paper: high seas monitoring for predictive purposes and the analysis of catch samples. Many other possibilities exist for other situations and other salmon species: the timing of inshore runs could be examined in estuarine areas or in river systems, the continent of origin of salmon on the high seas could be examined (for those species or areas not already analyzed), or the intermingling of hatchery and native populations could be analyzed for certain fisheries. Since scale samples are relatively easy to collect and exchange and since computers are readily available to do the necessary calculations, the polynomial discriminant method is a flexible and practical tool for the racial analysis of Pacific salmon, particularly sockeye salmon.

ACKNOWLEDGMENTS

Many thanks are due to Colin Harris, Allan C. Hartt, and Robert L. Burgner for their editorial advice and guidance. We also wish to thank James B. Scott for his meticulous work with the scale data. Many others providing services to the deep sea tagging project at the Fisheries Research Institute deserve thanks. The scale samples and data from the Bristol Bay river systems were collected by the Alaska Department of Fish and Game. We are grateful to Paul Krasnowski and other Alaska Department of Fish and Game personnel for providing these vital samples. This research was primarily supported by NOAA, National Marine Fisheries Service, under Contract No. 03-6-208-35470. The inshore stock separation project was supported by the Alaska Department of Fish and Game.

LITERATURE 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.
1964. Sockeye salmon scale studies. Int. North Pac. Fish. Comm., Annu. Rep. 1963:158-162.

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.

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.

DAHLBERG, M. L., AND D. E. PHINNEY.
1968. A microprojector for use in scale studies. Prog. Fish-Cult. 30:118-120.

DARK, T. A., AND B. J. LANDRUM.
1964. Analysis of 1961 red salmon morphological data. Int. North Pac. Fish. Comm., Annu. Rep. 1962:110-115.

DAS GUPTA, S.
1973. Theories and methods in classification: A review. *In* T. Cacoullos (editor), Discriminant analysis and applications, p. 77-137. Academic Press, N.Y.

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

FUKUHARA, F. M., S. MURAI, J. J. LALANNE, AND A. SRIBHIBHADH.
1962. Continental origin of red salmon as determined from morphological characters. Int. North Pac. Fish. Comm., Bull. 8:15-109.

HARTT, A. C.
1962. Movement of salmon in the North Pacific Ocean and Bering Sea as determined by tagging, 1956-1958. Int. North Pac. Fish. Comm., Bull. 6, 157 p.
1966. Migrations of salmon in the North Pacific Ocean and Bering Sea as determined by seining and tagging, 1959-1960. Int. North Pac. Fish. Comm., Bull. 19, 141 p.

HARTT, A. C., G. E. LORD, AND D. E. ROGERS.
1975. Monitoring migrations and abundance of salmon at sea. Int. North Pac. Fish. Comm., Annu. Rep. 1973:73-79.

ISSACSON, S. L.
1954. Problems in classifying populations. *In* O. Kempthorne, T. A. Bancroft, J. W. Gowen, and J. L. Lush (editors), Statistics and mathematics in biology, p. 107-117. Iowa State Coll. Press, Ames.

KONOVALOV, S. M.
1971. Differentiation of local populations of sockeye salmon *Oncorhynchus nerka* (Walbaum). (Translated by L. V. Sagen, 1975. Univ. Wash. Pub. Fish., New Ser., 6, 290 p.)

KRUSKAL, W. H., AND W. A. WALLIS.
1952. Use of ranks in one-criterion variance analysis. J. Am. Stat. Assoc. 47:583-621.

MAJOR, R. L., S. MURAI, AND J. LYONS.
1975. Scale studies to identify Asian and western Alaskan

chinook salmon. Int. North Pac. Fish. Comm., Annu. Rep. 1973:80-91.

MASON, J. E.
1966. Sockeye salmon scale studies. Int. North Pac. Fish. Comm., Annu. Rep. 1964:117-118.

MURTHY, V. K.
1966. Nonparametric estimation of multivariate densities with applications. *In* P. R. Krishnaiah (editor), Multivariate analysis, p. 43-56. Academic Press, N.Y.

PARZEN, E.
1962. On estimation of a probability density function and mode. Ann. Math. Stat. 33:1065-1076.

PATRICK, E. A.
1972. Fundamentals of pattern recognition. Prentice-Hall, Englewood Cliffs, N.J., 504 p.

ROGERS, D. E.
1975. Forecast of the sockeye salmon run to Bristol Bay in 1976. Univ. Wash., Fish. Res. Inst. Circ. 76-1, 49 p.

SMITH, C. A. B.
1947. Some examples of discrimination. Ann. Eugen. 13:272-282.

SPECHT, D. F.
1966. Generation of polynomial discriminant functions for pattern recognition. Stanford Univ., Tech. Rep. 6764-5, 127 p.

WELCH, B. L.
1939. Note on discriminant functions. Biometrika 31:218-220.

WORLUND, D. D., AND R. A. FREDIN.
1962. Differentiation of stocks. *In* Symposium on pink salmon, p. 143-153. H. R. MacMillan Lectures in Fisheries, Univ. B.C., Vancouver, Can.