Stock Identification of Sockeye Salmon (*Oncorhynchus nerka*) with Scale Pattern Recognition[1]

RODNEY C. COOK

Fisheries Research Institute, University of Washington, Seattle, WA 98195, USA

COOK, R. C. 1982. Stock identification of sockeye salmon (*Oncorhynchus nerka*) with scale pattern recognition. Can. J. Fish. Aquat. Sci. 39: 611–617.

The non-Gaussian density functions underlying polynomial discrimant functions are employed in a classification scheme designed for sockeye salmon (*Oncorhynchus nerka*). A leaving-one-out approach is used to estimate the smoothing parameters in the density functions and to obtain nearly unbiased estimates of expected actual error rates in the classification scheme. The result is that all available observations of known origin may be used to determine the discriminant rule and estimate classification error rates. These are needed to obtain point estimates of the proportions of subpopulations present in areas of intermingling. Several additional improvements over the polynomial discriminant method are noted. The scheme is applied to scale measurement data of sockeye salmon from Bristol Bay, the Gulf of Alaska, and the Kamchatka Peninsula.

Key words: stock identification, discriminant analysis, sockeye salmon

COOK, R. C. 1982. Stock identification of sockeye salmon (*Oncorhynchus nerka*) with scale pattern recognition. Can. J. Fish. Aquat. Sci. 39: 611–617.

Un schéma de classification conçu pour le saumon nerka (*Oncorhynchus nerka*) fait appel à des fonctions de densité non gaussiennes sous-jacentes à des fonctions discriminantes polynomiales. Une approche selon laquelle un terme de l'alternative est mis de côté a servi à estimer les paramètres de lissage dans les fonctions de densité et à obtenir des estimations presque non biaisées des taux d'erreur réels anticipés dans le schéma de classification. Comme résultat, on peut utiliser toutes observations disponibles d'origine connue pour déterminer la règle discriminante et estimer les taux d'erreur de classification. On a besoin de ces données dans des estimations ponctuelles des proportions de sous-populations là où il y a mélange. Nous mentionnons plusieurs autres améliorations sur la méthode discriminante polynomiale. Nous appliquons ce schéma à des mensurations d'écailles de saumons nerka de la baie Bristol, du golfe d'Alaska et de la péninsule du Kamchatka.

Received April 22, 1981
Accepted January 7, 1982

Reçu le 22 avril 1981
Accepté le 7 janvier 1982

THIS paper presents a pattern recognition technique that utilizes scale measurement data to classify sockeye salmon (*Oncorhynchus nerka*) to region of origin. The purpose is to develop a nonparametric approach that utilizes the leaving-one-out or jackknife procedures currently available with parametric approaches. (BMDP7M provides this for linear discriminant functions.) The method described is applicable to Gaussian and non-Gaussian scale pattern distributions. The method is discussed with reference to the sampling strategy required to obtain valid point and confidence interval estimates for the mixing proportion of subpopulations in mixed stock fisheries. These considerations are applicable to any classification procedure.

Discriminant function analyses have been used to identify subpopulations of sockeye salmon by evaluating scale growth patterns. Most of these studies have used discriminant functions based upon the assumption that the underlying data are normally distributed: namely linear or quadratic discriminant functions (Anas 1964; Mason 1966; Anas and Murai 1969; Major et al. 1975; Bilton and Messinger 1975; and others). Recently, nonparametric approaches have been used. The polynomial discriminant method of Specht (1966) has been applied by Cook and Lord (1978) to Bristol Bay sockeye salmon. This method may be implemented with no particular regard to the underlying distributions.

In racial studies of sockeye salmon the goal is not usually to identify the origin of an individual salmon, but to estimate the proportions of subpopulations in areas of intermingling (the mixing proportion). Worlund and Fredin (1962) noted a set of linear relationships from which estimates of these proportions may be obtained. This is referred to as the classification matrix correction procedure by Cook and Lord (1978).

Variance estimators have also been developed (Worlund

[1]Contribution No. 581, College of Fisheries, University of Washington, Seattle, WA 98195, USA.

Printed in Canada (J6481)
Imprimé au Canada (J6481)

and Fredin 1962; Pella and Robertson 1979). These estimators require two sets of data: (1) the results of classifying a number of salmon from the mixed or unknown population, and (2) the results of classifying salmon from each of the classes (or subpopulations) of known origin. Thus, the available observations from the subpopulations of known origin must be subdivided into learning and testing samples. The learning samples are used to estimate the discriminant functions and the testing samples are used to appraise the effectiveness of these functions and to estimate the elements of the classification matrix (required to make point and interval estimates of the relative proportions of the subpopulations of concern). It would be advantageous to pool these samples if a method were available to make unbiased estimates of the elements of the classification matrix. Many methods are available for estimating error rates in classification problems.

Error rates (i.e. the off-diagonal elements of the classification matrix) have been discussed extensively in the literature. Hills (1966) reviewed the problem, but most of the approaches appear unsatisfactory, including the method being used by most investigators involved in salmon stock identification. They divide the data into two subsamples, one being used to construct the discriminant function(s). The function(s) are then used to classify the remaining data, and classification error rates are estimated by the observed proportions of those misclassified. This approach requires fairly large samples (not always available), and is inefficient (i.e. does not utilize the full sample). The leaving-one-out approach originally proposed by Lachenbruch (1965) appears appropriate because it may be used no matter what the distribution of the observations, it uses all of the observations, and it gives nearly unbiased estimates of the probabilities of misclassification. The discriminant rule is estimated by omitting one sample observation and then the observation left out is classified with the rule. This is accomplished for all observations and the results are tallied. This gives an almost unbiased estimate for the expected actual error rate (Lachenbruch 1975). Lachenbruch (1967) has shown that in the two-class case the estimates for the probability of misclassification are unbiased for a discriminant function based on $n_1 - 1$ and n_2 observations (or n_1 and $n_2 - 1$ observations). Thus, pooling the learning and testing samples into, say, training samples deserves further attention.

Methods

The estimation procedure(s) used in a stock identification problem will determine certain requirements for the pattern recognition scheme. Let the results of any given classification scheme be represented by the vector $\mathbf{R}$ where the element R_i is the proportion of the mixed population that would classify as class i. Similarly, let the proportions of the mixed population belonging to classes i be represented by the vector $\mathbf{U}$. Thus, $\mathbf{R}$ is a function of $\mathbf{U}$. If the classification scheme is represented by the matrix $\mathbf{C}$, where the element C_{ij} is the fraction of the population from class j that would classify as class i ($\Sigma C_{ij} = 1$, for all j), then the following relation exists:

$$\mathbf{CU} = \mathbf{R}$$
$$\text{or } \mathbf{U} = \mathbf{C}^{-1}\mathbf{R}$$

The proportions of the various classes present ($\mathbf{U}$) are unknown and can be estimated:

$$\hat{\mathbf{U}} = (\hat{\mathbf{C}})^{-1}\hat{\mathbf{R}}$$

where $\hat{\mathbf{C}}$ and $\hat{\mathbf{R}}$ are computed as follows.

A test sample from each class j is used to estimate a set of multinomial probabilities ($\hat{C}_{ij}$s for fixed j), and an unknown sample from the mixed population is used to obtain $\hat{\mathbf{R}}$. Thus, $\hat{\mathbf{C}} = \mathbf{c}$, where c_{ij} is the proportion of the test sample from class j that classified as class i. Also $\hat{\mathbf{R}} = \mathbf{r}$, where r_i is the proportion of the unknown sample that classified as class i.

The fundamental assumption inherent in all pattern recognition schemes is that the subpopulations are modeled properly and that the mixed population consists only of individuals from the modeled subpopulations. Further, the elements of $\hat{\mathbf{C}}$ must be nearly unbiased, so that $\hat{\mathbf{U}}$ is nearly unbiased also.

The elements of $\mathbf{C}$ may be estimated without significant bias by utilizing the leaving-one-out approach. Therefore $\hat{\mathbf{U}}$ should remain nearly unbiased over the range of practical application. The benefits are obvious. All available observations may be used both to determine the decision rule and to estimate $\mathbf{C}$. Thus the point estimates ($\hat{\mathbf{U}}$) will be more precise than if only a subsample of observations of known origin were used. Sample size requirements will be reduced: for a required precision of estimation (for $\hat{\mathbf{U}}$) fewer observations will be needed to implement the pattern recognition scheme because the observations may be utilized both to establish the discriminant rule and to estimate $\mathbf{C}$. This is not the case in the traditional approach. The above remarks about precision relate directly to the variance of $\hat{\mathbf{U}}$: the variance of $\hat{\mathbf{U}}$ is inversely related to sample size. The formulations of Worlund and Fredin (1962) and Pella and Robertson (1979) demonstrate this. However, it must be cautioned that these formulae require that $\hat{\mathbf{R}}$ and $\hat{\mathbf{C}}$ be statistically independent.

This criterion has never been met in practice. All of the applications cited in the introduction would produce $\hat{\mathbf{R}}$ and $\hat{\mathbf{C}}$ that are dependent. For $\hat{\mathbf{R}}$ and $\hat{\mathbf{C}}$ to be independent they must be obtained from independent classification rules (i.e. estimated with independent sets of "training" samples). Thus one training sample would be used to estimate a classification procedure that in turn would be used to classify the individuals comprising the unknown sample (resulting in $\hat{\mathbf{R}}$). A second independent training sample would be used to estimate another independent classification procedure that would be used to obtain $\hat{\mathbf{C}}$ with the leaving-one-out approach. Thus, we are back to subdividing the available samples of known origin and increasing the complexity of the analysis. Intuitively this does not appear to be necessary, as evidenced by earlier studies. Point estimates are not affected; however, variance estimates may be optimistic. Additional terms in the variance formulae of Pella and Robertson (1979) would be required to account for this dependence. However, this dependence is not "classic" in that the individuals comprising different samples are classified to obtain $\hat{\mathbf{R}}$ and $\hat{\mathbf{C}}$. A modified variance formulation would be too conservative and produce larger than needed confidence intervals. Simulations have shown that Pella and Robertson's variance formula is already conservative when $\hat{\mathbf{R}}$ and $\hat{\mathbf{C}}$ are independent (Cook unpublished). Thus with a fixed number of samples of known origin the choice is between a more precise $\hat{\mathbf{U}}$ with possibly optimistic confidence intervals or a less precise $\hat{\mathbf{U}}$ (because the test samples must be subdivided) and conservative confidence intervals. The conservative nature of Pella and Robertson's variance formula

may help account for the optimism in the former approach. With this in mind and in keeping with historic applications, this paper will present the former approach; however, this or any other stock identification procedure is easily adapted to the latter approach.

Discriminant function analyses depend upon the recognition of underlying patterns differing among classes of objects. In this case, scale patterns characterize a sockeye salmon of a particular geographic 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 training sample of p-tuples for a number of salmon from one origin establishes a region in p-space characteristic of that class of sockeye. Training samples from salmon of different and known origins establish regions in p-space which may be separated by decision surfaces. These regions are described mathematically by multivariate probability density functions. A sockeye salmon of unknown origin may be classified according to which region its p-tuple occupies. Determination of this region requires that the value of the probability density function for each class is calculated at the point defined by the p-tuple of the unknown salmon. The largest result determines that class to which that sockeye is assigned; the probability that the unknown salmon belongs to this class is apparently the largest. The decision boundaries are thus defined as the points at which values of the density functions are equal.

These decision surfaces may be "shifted" by multiplying the values of the density functions by an a priori probability. I prefer to adjust the values of the a priori probabilities so that the percentage of fish belonging to a certain class that were missassigned to all other classes approximately equals the percentage of fish from all other classes missassigned to that certain class. This avoids situations where large numbers of misclassifications in one direction cause a decrease in overall classification accuracy. By shifting the decision surface the number of these misclassifications may be decreased dramatically while the corresponding misclassifications in the other direction are not significantly increased. Accuracies are generally higher than if a priori probabilities are left equal.

The decision rule is formalized below. Choose

$$d(\mathbf{X}) = \theta_r$$

such that

$$h_r f_r(\mathbf{X}) \geq h_s f_s(\mathbf{X}) \text{ for all } s \neq r$$

where $\mathbf{X}$ = vector of scale measurement data from a salmon to be classified, $d(\mathbf{X})$ = the decision on an $\mathbf{X}$, θ_k = the classes (subpopulations), $f_k(\mathbf{X})$ = the value of the estimated density function, for θ_k at point $\mathbf{X}$, and h_k = the a priori probabilities.

By using normal density functions in the decision rule, linear discriminant functions may be obtained if the variance–covariance matrices of the data are considered similar for each class. Welch (1939) noted this for Fisher's (1936) linear discriminant function. The quadratic discriminant function results when the variance–covariance matrices are considered dissimilar (Smith 1947). Specht's (1966) polynomial discriminant functions result from using a density function of the form described by Parzen (1962) and extended by Murthy (1966) to the multivariate case. The density function used by Specht (1966) was chosen for this study because it is non-Gaussian and is appropriate for modeling any foreseeable scale pattern distribution.

The probability density function for each class is approximated by the average of the contribution of the ith training sample observation to the estimated density. This "contribution" is modeled with what closely resembles the multivariate normal density function. The key difference is that the variance–covariance matrix is replaced by $\sigma^2 I$ where σ is a "smoothing parameter." The result is that the overall probability density function is an average of Gaussian-type density functions. As each of these Gaussian-type functions is estimated with one observation, that observation vector replaces the mean vector and the $\sigma^2 I$ replaces the variance–covariance array. By judiciously choosing σ, the estimated overall density function will pass smoothly through the data points representing that class. (A formula for obtaining σ is presented later.) The result of interest is that the estimated density becomes identical with the true parent density under the limiting conditions of a large number of training samples whether the parent density is Gaussian or not.

The density function is

$$(1)\quad f_a(\mathbf{X}_{bj}) = \frac{1}{\sigma_a^p (2\pi)^{p/2}} \frac{1}{n_a} \sum_{i=1}^{n_a} \exp - \frac{(\mathbf{X}_{ai} - \mathbf{X}_{bj})'(\mathbf{X}_{ai} - \mathbf{X}_{bj})}{2\sigma_a^2}$$

where i = observation (fish) number, $\mathbf{X}_{ai}$ = ith observation from class θ_a, n_a = test sample size for class θ_a, $\mathbf{X}_{bj}$ = the observation to be classified, p = the dimension of $\mathbf{X}$, and σ_a = the smoothing parameter for θ_a.

The density function employing the leaving-one-out approach is thus

$$(2)\quad f_a^j(\mathbf{X}_{aj}) = \frac{1}{\sigma_a^p (2\pi)^{p/2}} \frac{1}{n_a - 1} \sum_{\substack{i=1 \\ i \neq j}}^{n_a} \exp - \frac{(\mathbf{X}_{ai} - \mathbf{X}_{aj})'(\mathbf{X}_{ai} - \mathbf{X}_{aj})}{2\sigma_a^2}$$

where $\mathbf{X}_{aj}$ = the observation to be classified.

Equation (1) is used to calculate the value of the estimated density function for θ_a at $\mathbf{X}_{bj}$, and equation (2) is used to calculate the value of the estimated density function for θ_a at $\mathbf{X}_{aj}$. Subsequent classification of the fish comprising the unknown sample(s) utilizes equation (1) and the training sample for each of the established classes.

The direct use of this estimated density function in the decision rule provides certain advantages over the polynomial discriminant method. As mentioned previously, its adaptability to the leaving-one-out rule allows reduced sample size requirements. The full character of the training samples is retained; no information is lost by reducing the scale measurement data to sets of coefficients for the discriminant functions. Truncation of the polynomial discriminant functions is bypassed; the decisions resulting from direct density estimates are the same as those that would be obtained from untruncated polynomials. Also, a different smoothing parameter may be

used in the density function for each class. (When polynomial discriminant functions are used for classification the smoothing parameter value must be held constant across classes.) Smoothing parameter estimates are usually different for different distribution functions.

Estimation of the smoothing parameter is done with a leaving-one-out modification of the maximum likelihood method (Habbema et al. 1974). The leaving-one-out approach is used because the likelihood function goes to infinity as the smoothing parameter goes to zero. The method is thus choose the value $\hat{\sigma}_a$ such that

$$g(\hat{\sigma}_a) = \max \sum_{j=1}^{n_a} f_a^j(\mathbf{X}_{aj}),\ \hat{\sigma} > 0$$

When the data are "standardized"[2] the smoothing parameter appears to take on values in the range of from 0.5 to 1.5.

In his original work Specht (1966) determined experimentally that finding a good value for the smoothing parameter(s) is not difficult. Plotting accuracy versus smoothing parameter value results in the peak of the curve being sufficiently broad. Generally I agree. It appears that for standardized data the value 1.0 may be chosen arbitrarily without incurring much loss of classification accuracy. The estimation procedure of Habbema et al. (1974) should suffice otherwise.

Application

The Japanese conduct a land-based high seas driftnet fishery for salmon in the western North Pacific Ocean. Since 1962 this fishery has been the world's second largest salmon fishery for all species combined. In the 1970s the fleet shifted its effort to the eastern portion of the fishery area and sockeye salmon catches increased. Although it was clear that Asian sockeye were present in the eastern area, no conclusions could be made about the presence or absence of North American stocks. The United States unilaterally assumed management authority for her stocks of Pacific salmon throughout their migratory range by passing Public Law 94-265 and there was then a need to initiate a study of the origin of sockeye salmon intercepted by this fishery. Evidence accumulated from this study could then be used by the International North Pacific Fisheries Commission in their deliberations concerning the regulation of the Japanese land-based driftnet fishery. To appraise the methodology developed in the last section, scale measurement data collected for the above-mentioned study were analyzed. The same sets of scale characters were used.

Three regions and stocks were defined: Kamchatka, Bristol Bay, and the south coast of Alaska from about 160°W to 145°W (Gulf of Alaska). A training sample for each age-class and each region (class) was compiled by subsampling in proportion to run size from a large collection of scale samples representing the major stocks within the region. Scale samples were provided by the Alaska Department of Fish and Game, the Fishery Agency of Japan, and by the Pacific Scientific Institute of Fisheries and Oceanography (TINRO), USSR (see Marshall et al. (1978) for a detailed description of the definition of the training samples; sample collection, preparation, and viewing; and scale character selection).

Age 2.2 Sockeye in 1975

The data from the age 2.2 sockeye (i.e. fish which migrated to sea after two winters in freshwater and have spent two winters at sea) collected in 1975 were reanalyzed using direct density estimation. The scale characters used were (1) the number of circuli in the freshwater growth zone, (2) the size of the freshwater growth zone, (3) the number of circuli in the first half of the 1st yr's ocean growth zone, (4) the number of circuli in the second half of the 1st yr's ocean growth zone, (5) the distance between the outer edge of the last freshwater circulus and the 6th ocean circulus, and (6) the distance between the 15th and 18th ocean circuli. There were 200, 199, and 201 scales selected from the Bristol Bay, Gulf of Alaska, and Kamchatka regions, respectively. Each sample was subdivided into two subsamples, one being the training sample and one being reserved for subsequent testing. The estimated smoothing parameters were $\hat{\sigma}_B = 1.0$, $\sigma_G = 1.3$, and $\hat{\sigma}_K = 1.1$. The observations comprising the training samples were then classified. A priori probabilities were adjusted so that the number of fish from a certain class that were misassigned to all other classes equaled the number of fish misassigned to that certain class from all other classes (Cook and Lord 1978). The a priori probabilities were $h_B = 0.143$, $h_G = 0.603$, and $h_K = 0.254$. The results of the classification scheme were tallied (Table 1) and the classification matrix was estimated as

$$\hat{\mathbf{C}}_1 = \begin{bmatrix} 0.7900 & 0.1700 & 0.0396 \\ 0.1700 & 0.5500 & 0.2772 \\ 0.0400 & 0.2800 & 0.6832 \end{bmatrix}$$

The leaving-one-out approach for estimating the elements of the classification matrix should give nearly unbiased results. This may be checked by comparing classification matrices estimated with the samples used to establish the classification procedure and with an independent set of test samples. Indeed, Cook and Lord (1978) did not detect significant differences between elements of classification matrices estimated with test samples used to determine the a priori probabilities and corresponding elements estimated with a second independent test group. Now, to check the possible bias in $\hat{\mathbf{C}}_1$, the fish in the second group of subsamples were classified like fish from an unknown sample (Table 2). Thus, a second classification matrix, independent of the training samples used to establish the classification scheme, was estimated as

$$\hat{\mathbf{C}}_2 = \begin{bmatrix} 0.8400 & 0.2222 & 0.0700 \\ 0.1500 & 0.5253 & 0.2400 \\ 0.0100 & 0.2525 & 0.6900 \end{bmatrix}$$

Because the columns of the classification matrix are sets of multinomial probabilities, the classification matrices were compared column by column. The chi-square test statistics

[2]All data points were "standardized." That is, the mean and standard deviation for each scale character were calculated from the training samples (all classes 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.

TABLE 1. Classification array obtained from direct density estimation for training samples of age 2.2 mature sockeye salmon from Bristol Bay, the Gulf of Alaska, and the Kamchatka Peninsula in 1975.

Calculated decision	Correct decision		
	Bristol Bay	Gulf of Alaska	Kamchatka
Bristol Bay	79	17	4
Gulf of Alaska	17	55	28
Kamchatka	4	28	69
Training sample sizes	100	100	101

TABLE 2. Classification array obtained from direct density estimation for test samples of age 2.2 mature sockeye salmon from Bristol Bay, the Gulf of Alaska, and the Kamchatka Peninsula in 1975.

Calculated decision	Correct decision		
	Bristol Bay	Gulf of Alaska	Kamchatka
Bristol Bay	84	22	7
Gulf of Alaska	15	52	24
Kamchatka	1	25	69
Test sample sizes	100	99	100

TABLE 3. Classification array obtained from polynomial discriminant functions for test samples of age 2.2 mature sockeye salmon from Bristol Bay, the Gulf of Alaska, and the Kamchatka Peninsula in 1975 (from Marshall et al. 1978).

Calculated decision	Correct decision		
	Bristol Bay	Gulf of Alaska	Kamchatka
Bristol Bay	103	17	5
Gulf of Alaska	14	75	35
Kamchatka	8	32	88
Test sample sizes	125	124	128

were 2.08, 0.89, and 1.12. No bias was detectable ($\alpha = 0.05$) at the sample sizes used. The hypotheses that the samples used to establish $\hat{\mathbf{C}}_1$ and $\hat{\mathbf{C}}_2$ were drawn from the same populations were not rejected.

The classification scheme was sufficiently accurate. Comparisons of classification matrices showed that it performed as well as the polynomial discriminant functions used in Marshall et al. (1978). The column by column chi-square test statistics were 2.03, 0.78, and 0.005 for the age 2.2 sockeye (Table 3). No significant differences were detected ($\alpha = 0.05$).

This data set was also used to check the effects of the smoothing parameter on the accuracy of the classification scheme. Figure 1 gives the percentage correctly classified for a test sample as a function of its smoothing parameter. The smoothing parameters were held constant at their maximum likelihood estimates for the other classes. This method appears to give smoothing parameter estimates larger than those

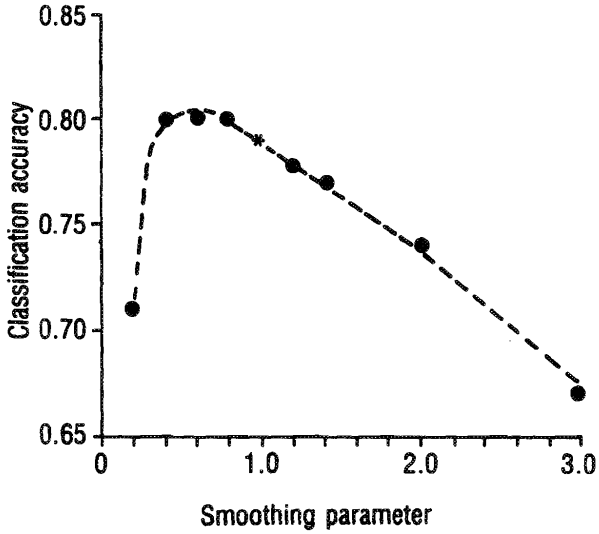

FIG. 1. The percentage of the Bristol Bay training sample (age 2.2 mature sockeye in 1975) correctly classified as a function of the smoothing parameter for that class. The modified maximum likelihood estimate is indicated with an asterisk (*).

providing maximum classification accuracy. Although differences of a few percent are not significant, in critical applications one may wish to try several combinations of smoothing parameters and select that combination providing the best accuracy. The modified maximum likelihood estimation procedure of Habbema et al. (1974) should give good results or, at least, provide a starting point for selecting alternative combinations.

AGE 2.3 SOCKEYE IN 1975

The data from the age 2.3 sockeye salmon collected in 1975 were reanalyzed using direct density estimation. The scale characters used were (1) the number of freshwater circuli, (2) the number of circuli in the first half of the 1st yr's ocean growth, (3) the size of the 1st yr's ocean growth, (4) the distance from the outer edge of the last freshwater circulus to the 3rd ocean circulus, (5) the distance from the 3rd to the 6th ocean circulus, and (6) the distance from the 12th to the 18th ocean circulus. There were 195, 199, and 233 scales available from the Bristol Bay, Gulf of Alaska, and Kamchatka regions, respectively. Again, a training and test sample were established for each class. The smoothing parameters were estimated as $\hat{\sigma}_B = 1.1$, $\sigma_G = 1.0$, $\sigma_K = 1.1$. The fish in the training samples were classified (Table 4) and the classification matrix was estimated as

$$\hat{\mathbf{C}}_1 = \begin{bmatrix} 0.7041 & 0.2200 & 0.0598 \\ 0.1633 & 0.5000 & 0.2906 \\ 0.1327 & 0.2800 & 0.6496 \end{bmatrix}$$

The a priori probabilities were $h_B = 0.363$, $h_G = 0.256$, and $h_K = 0.381$. Again, to check the potential bias in $\hat{\mathbf{C}}_1$, the fish in the test samples were classified (Table 5). The resulting classification matrix was

$$\hat{C}_2 = \begin{bmatrix} 0.7113 & 0.1212 & 0.1034 \\ 0.2062 & 0.5253 & 0.3448 \\ 0.0825 & 0.3535 & 0.5517 \end{bmatrix}$$

Column by column comparison of $\hat{C}_1$ and $\hat{C}_2$ resulted in chi-square test statistics of 1.63, 3.76, and 3.32. No bias was detected ($\alpha = 0.05$) at the sample size used. Also, the classification scheme performed as well as the polynomial discriminant functions in Marshall et al. (1978) (Table 6). The column by column chi-square test statistics were 0.56, 0.64, and 0.30 for the age 2.3 sockeye.

Conclusions and Recommendations for Implementation

Traditional discriminant function analyses are aimed at estimating the proportions of subpopulations in areas of mixing and require that the available samples of known origin be subdivided. One set of subsamples (the learning samples) is used to determine the discriminant rule and the other set of subsamples (the test samples) is used to estimate the elements of the classification matrix. The leaving-one-out approach allows the subsamples to be combined into training samples which may be used both to estimate the discriminant rule and the elements of the classification matrix. Thus, all available subsamples may be utilized to obtain statistically valid point estimates for the proportions of subpopulations in areas of mixing. However, to obtain statistically valid variance estimates it may be necessary to subdivide the available samples of known origin. Otherwise, optimistic confidence intervals may result.

TABLE 4. Classification array obtained from direct density estimation for training samples of age 2.3 mature sockeye salmon from Bristol Bay, the Gulf of Alaska, and the Kamchatka Peninsula in 1975.

Calculated decision	Correct decision		
	Bristol Bay	Gulf of Alaska	Kamchatka
Bristol Bay	69	22	7
Gulf of Alaska	16	50	34
Kamchatka	13	28	76
Training sample sizes	98	100	117

TABLE 5. Classification array obtained from direct density estimation for test samples of age 2.3 mature sockeye salmon for Bristol Bay, the Gulf of Alaska, and the Kamchatka Peninsula in 1975.

Calculated decision	Correct decision		
	Bristol Bay	Gulf of Alaska	Kamchatka
Bristol Bay	69	12	12
Gulf of Alaska	20	52	40
Kamchatka	8	35	64
Test sample sizes	97	99	116

TABLE 6. Classification array obtained from polynomial discriminant functions for test samples of age 2.3 mature sockeye salmon from Bristol Bay, the Gulf of Alaska, and the Kamchatka Peninsula in 1975 (from Marshall et al. 1978).

Calculated decisions	Correct decisions		
	Bristol Bay	Gulf of Alaska	Kamchatka
Bristol Bay	90	22	8
Gulf of Alaska	16	65	43
Kamchatka	14	37	95
Test sample sizes	120	124	146

Acknowledgments

I wish to thank Colin Harris, Robert L. Burgner, Gary E. Lord, Donald A. McCaughran, and Scott L. Marshall for their editorial advice, guidance, and efforts in obtaining scale samples. I am grateful to the personnel of the Alaska Department of Fish and Game, the Fishery Agency of Japan, and TINRO, USSR, for making these samples available. Many others providing services at the Fisheries Research Institute deserve thanks. This research was primarily supported by NOAA, National Marine Fisheries Serice, under Contract No. 03-6-208-35470, and by the North Pacific Fisheries Management Council.

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: 199–129.

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. Fish. Bull. 76: 415–423.

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

HABBEMA, J. D. F., J. HERMANS, AND K. VAN DERR BROEK. 1974. A stepwise discriminant analysis program using density estimation, p. 101–110. Compstat. 1974, Proc. Computational Stat. Physica Verlag, Wein.

HILLS, M. 1966. Allocation rules and their error rates. J. R. Stat. Soc. B28: 1–31.

LACHENBRUCH, P. A. 1965. Estimation of error rates in discriminant analysis. Ph.D. thesis, Univ. California, Los Angeles.

1967. An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. Biometrics 23: 639–645.

1975. Discriminant analysis. Hafner Press, New York. 128 p.

MAJOR, R. L., A. 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.

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 landbased driftnet fishery. Univ. Washington, Fish. Res. Inst. Tech. Rep. FRI-UW-7816: 152 p.

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, p. 43–56. *In* P. R. Krishnaiah [ed.] Multivariate analysis. Academic Press Inc., New York.

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

PELLA, J. J., AND T. L. ROBERTSON. 1979. Assessment of composition of stock mixtures. Fish. Bull. U.S. 77: 387–398.

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, p. 143–153. *In* Symposium on pink salmon. H. R. MacMillan Lectures in Fisheries, Univ. British Columbia, Vancouver, Canada.