Using the Coordinates of some Character Points of Scales for Differentiation of Pacific Salmon Stocks

O.M. Zaporozhets and **G.V. Zaporozhets**
Kamchatka Research Institute of Fisheries & Oceanography (KamchatNIRO),
18, Naberazhnaya Str., Petropavlovsk-Kamchatsky, 683602, Russia

Zaporozhets, O.M., and G.V. Zaporozhets. 2000. Using the coordinates of some character points of scales for differentiation of Pacific salmon stocks. N. Pac. Anadr. Fish Comm. Bull. No. 2: 325–329.

Keywords: Stock identification, population, scales, sockeye salmon

Abstract: Classification accuracy of adult sockeye salmon (*Oncorhynchus nerka*) from three known populations (one fluvial and two limnobiotic) spawning in the Paratunka River basin (East Kamchatka) was assessed by two methods. The first method used scale structure (spacing of circuli triples and so on), the second method used these characteristics, in conjunction with X and Y coordinates of some typical points at the inner boundary between the fresh-water scale zone and the marine first year scale zone. Discriminate analysis with additional use of two-dimensional parameters revealed a significant increase in average accuracy of classification.

INTRODUCTION

Fishery management and monitoring of national resources within international waters require a reliable method of identifying origin of salmon stocks. There are many methods used currently, but the most promising is differentiation of adult fish scale structure (Anas and Murai 1969; Tanaka et al. 1969; Cook and Lord 1978; Cook 1982; Nikolayeva and Semenets 1983; Myers 1985; Davis 1987; Millar 1988; Schwartzberg and Fryer 1993; Bernard and Myers 1994; Kayev 1998; Temnykh 1998; and others). Scale characteristics include distances from the scale center to various structures (some circuli, circuli triplets, annual ring, fresh-water growth zone and marine growth zone boundaries and so on) and distances between the structures, and the number of circuli within various zones on the scale.

Usually the scale characteristics are sufficient for identification of large stocks on the basis of representative standard samples. However, in some cases (for example, to differentiate various populations of a species which spawn in closely spaced spawning grounds, or to estimate adult return to different hatcheries within one river) traditional characteristics might be not enough. Therefore, in addition to these characteristics, we also use two-dimensional scale parameters. One of these parameters can be X and Y coordinates of some typical points in relation to the scale center. We have suggested that several edge points delineating the boundary of the first marine year zone might be useful for identification of stocks inhabiting different ecological niches during early life.

Identification of stock origin from scale shape and otholiths, using the Fourie function, has been done (Jarvis et al. 1978; Bird et al. 1986). However, when using the Fourie function the coordinates of a group of edge points in relation to the center of estimation area must be found. We have proposed using the scale center as the center of coordinates, which allows measurement of the vertical and horizontal asymmetry. Scale characteristics may be related to locomotory function of fish. It is possible that scale structures are determined by the level of adaptation of individuals to intense or prolonged swimming.

Two-dimensional parameters of the first year zone on scales were successfully used as additional characteristics for identification of wild and hatchery chum salmon entering spawning grounds of Paratunka River, West Kamchatka (Zaporozhets and Zaporozhets 1999). We decided, therefore, to apply the method to sockeye salmon (*Oncorhynchus nerka*) from different populations inhabiting one river to see if we could improve the separation of two such closely related populations.

In this study we used scale measurement data from fish whose origin was known, in order to estimate potential benefit from using the two-dimensional method.

MATERIALS AND METHODS

Adult sockeye salmon scale samples from three populations were collected from spawning grounds by scientists working in KamchatNIRO (Eugeny Pogodayev and authors). Spawning grounds were all in the Paratunka River basin: Blidzneye Lake, Dalneye

All correspondence should be addressed to O.M. Zaporozhets.
e-mail: Zaporozhets@newmail.ru

Lake and upper reaches of Paratunka River. Scales were taken according to NPAFC recommendations (Knudsen 1985; Davis et al. 1990). Age was estimated by Eugeny Pogodayev.

The radiuses of all circuli in the first annual zone were measured by one reader, along a long axis of the scale between reference lines connecting the edges of extreme circuli (Fig. 1). These measurements were analyzed using "BioSonics" OPRS (OPRS, BioSonics Inc., Seattle, WA) at the magnification of x82. We also measured X and Y coordinates from the center (focus) of scales to six characteristic points as follows: 1) left lowest termination of the first year growth zone, 2) left extreme of the first year zone, 3) upper extreme of the first year growth zone, 4) maximum radius point at the first year growth zone, 5) right extreme of the first year growth zone, 6) right lowest termination of the first year growth zone.

Fig. 1. Central zone of sockeye salmon scale demonstrating characteristic axes and points: 0, center (focus) of scale; 1, left lowest termination of the first year growth zone; 2, left extreme of the first year zone; 3, upper extreme of the first year growth zone; 4, maximum radius point at the first year growth zone; 5, right extreme of the first year growth zone; 6, right lowest termination of the first year growth zone; 1–6, reference line; 0–4, axis used for making measurements.

Geometrical coordinates are defined as follows:

- X_0, Y_0 - focus of scale;
- $X_1<0$, $|Y_1<0| \rightarrow$ Max,
- $|X_2<0| \rightarrow$ Max, Y_2 – any location,
- X_3 - any location, $Y_3 \rightarrow$ Max,
- $X_4 \rightarrow$ Max, $Y_4 \rightarrow$ Max,
- $X_5 \rightarrow$ Max, $Y_5 >0$,
- $X_6>0$, $|Y_6<0| \rightarrow$ Max;
- $Y_1 \approx Y_6$;

Scale Variables

Initially OPRS measurement data consisted of radial and coordinate parameters.

I. Radial parameters: distances from the center of the focus to the outer edge of each circulus that was marked on the OPRS monitor. These data were reformatted into an initial set of 11 variables:
1) - R1 - central zone size;
2) - NF - total number of fresh-water sclerites;
3) - RF - radius of last fresh-water circulus;
4) - NR - number of sclerites until first marine year circulus;
5-11) - average circulus spacing in seven initial triplets (T1 - T7), assessed from the equation

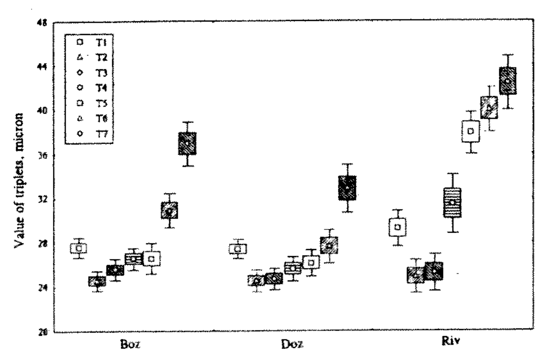
$$T_n = (DR_n+DR_{n+1}+DR_{n+2})/3),$$

Where $DR_n = R_{n+1} - R_n$, and R_{n+1} И R_n - radii of two neighbouring sclerites.

The triplets were assessed as particular sclerites united in groups by three in every group to 'smooth' the individual circulus increments on a particular fish scale and to make characteristics of stocks more certain (Davis 1987; Bernard and Myers 1994). We did not separate the fresh-water triplets from marine triplets because the number of fresh-water sclerites in fluvial sockeye can be very low (Fig. 2). The separation would, therefore, complicate matters.

II. Coordinate parameters: X and Y coordinates of the seven points on the scale mentioned above (including the center) were estimated from the center of the focus to the outer edge of each circulus marked on the OPRS monitor. The system was reformatted by subtracting coordinates: coordinates of the points (X_n, Y_n) minus coordinates of the center (X_0, Y_0).

Fig. 2. Dispersion diagrams of seven triplets in three sockeye salmon populations in the Paratunka River basin: BOZ, "Blizhneye Lake"; DOZ, "Dalneye Lake"; RIV, "Fluvial". Central square, diamond, etc., mean; boxes, standard error; whiskers, 95% confidence interval.



Variable Selection

Transformed data were tested for missing values, inter-group correlation and normal distribution through the STATISTICA-5.1 program (StatSoft, Inc. 1998). Variables which failed the tests were excluded from the analysis. Extreme outliers were sought through distribution diagrams (Scatterplot with Box Plot). Extreme mean values were also excluded from the analysis. All individuals having an incomplete set of variables were excluded from the analysis automatically according to the Casewise deletion of missing data option.

Two-factor dispersion analysis ANOVA (origin and year of birth, a = 0.05) was used to select further variables. The task was to find sub-sets of variables which were significantly different in different sockeye populations, but similar among different aged individuals.

In final studies we used only four circuli triplets T_3-T_6, first radius R_1 and coordinates: X_1, Y_2, Y_3, Y_4, Y_5. We analyzed data from measurements of 180 individual scale samples collected in 1996, including 86 scale samples from Blidzneye Lake sockeye (1), 64 samples from Dalneye Lake sockeye (2), and 28 from Paratunka River sockeye (3).

Tests of Model Performance

Discriminate analysis of data was accomplished using the program STATISTICA 5.1 after preliminary computation. Classification accuracy was tested through the five-fold cross-validation. This kind of validating has been useful when there is no particular test sample.

Classification is repeated five times, each time a sub-sample being excluded in constructing a test-sample for cross-validating. Thus, every sub-sample should act five-1 times as a participant of learning sample and as a test-sample. Cross-validation costs, estimated for five test samples, are averaged to get five time errors for estimation of the cross-validation cost, together with its standard error. The cost is expressed as the percent of mistakenly classified observations. Accordingly, the data were divided into three equal sub-samples, every sub-sample having 48%, 36% and 16% of Blidzneye Lake, Dalneye Lake and Paratunka River sockeye, respectively. Then three-fold classification of learning and test samples was accomplished.

RESULTS

Two-factor ANOVA (origin and year of birth) showed there was no significant difference ($p < 0.05$) between the three sockeye populations in their average size of initial spacing between circuli triplets (T_1, T_2). Otherwise, NF and T_2 variables depended significantly on the age of fish ($p < 0.05$). These variables were excluded from the analysis.

The most common significant differences among populations occurred with the following characteristics: X_1, Y_1, X_2, Y_3, Y_4, X_5, T_5, T_6. However, after analysis of correlation matrix and preliminary simulation the following characteristics were selected: R_1, X_2, Y_2, Y_3, Y_4, Y_5, T_3, T_4, T_5, T_6.

Step-by-step discriminate analysis of three equal data sub-samples using ten selected variables indicated that one of the characteristics (Y_4, T_3, T_4) was surplus and the program did not select that characteristic for the model. Nevertheless, those characteristics were not completely excluded from the analysis because average accuracy of classification would have been decreased.

The model, which included nine variables, obtained from program selection, classified learning samples with an average accuracy of about 84±4%, and test sub-samples with average accuracy of 77±6% (Table 1). Results with test subsamples were best for Blidzneye Lake sockeye (88±5%). Dalneye Lake sockeye classification with the model was poor (60±7%) (Fig. 3). Average cost of classification was 23±6%.

Step-by-step discriminate analysis of three equal data sub-samples using 5 radial variables (R_1, T_3, T_4, T_5, T_6) showed that one of these variables (T_3) was surplus in two of three cases.

The model, which included four to five variables, selected as a result of simulation, classified learning samples with an average accuracy of about 63±4%, and test sub-samples with an accuracy of 59±7% (Table 1, Fig. 3). Average cost of classification was increased by 1.4 times (32±7%) compared to the first method.

Examination of canonical analysis results showed that contribution of coordinate characteristics to differentiation of three sockeye salmon populations in first and second discriminate functions took about 60% and more than 70% respectively (Fig. 4).

Table 1. Classification accuracy (mean % ± standard error) of sockeye salmon origin assessed with "coordinate" characteristics (RT+XY) and without these characteristics (RT). BOZ, Blizhneye Lake; DOZ, Dalneye Lake; RIV, Fluvial Stock of Paratunka River.

	RT+XY				RT			
Samples	BOZ	DOZ	RIV	M±m	BOZ	DOZ	RIV	M±m
Learning	95.9±1.5	69.0±0.8	87.7±4.6	84.2±4.2	62.9±5.9	60.9±5.1	63.9±11.6	62.5±4.1
Test	88.3±4.7	60.3±6.7	81.7±9.7	76.8±5.6	57.6±4.1	42.8±16.3	77.9±6.4	59.4±7.3
Classification effort	11.7±4.7	39.7±6.7	18.3±9.7	23.2±5.6	42.5±4.1	57.0±16.4	22.2±6.4	31.9±7.3

Fig. 3. Classification accuracy among three sockeye salmon populations (Pop. 1, Blizhneye Lake; Pop. 2, Dalneye Lake; Pop. 3, Paratunka River) in learning samples (above) and test samples (below), compared by two methods: 1 using coordinate characteristics; 2 without these characteristics.

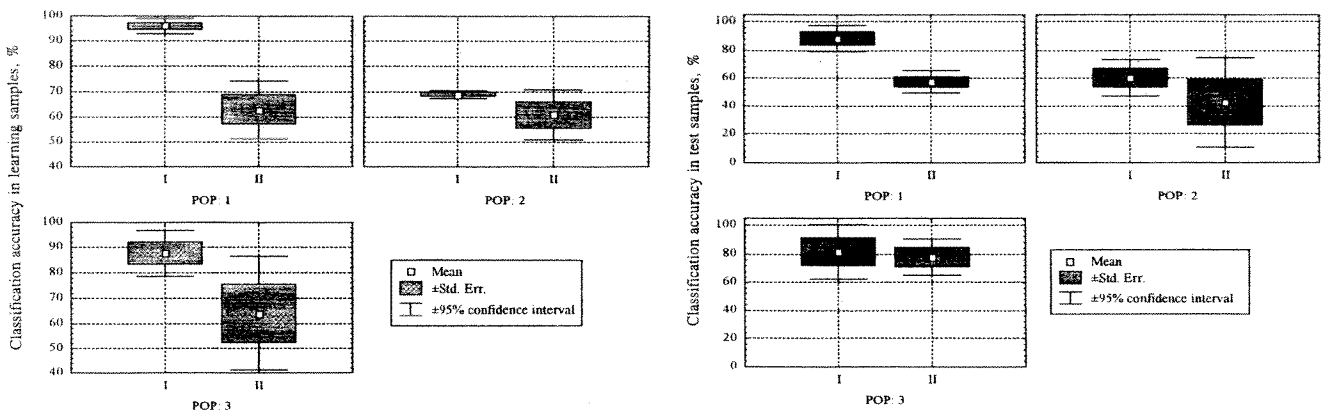
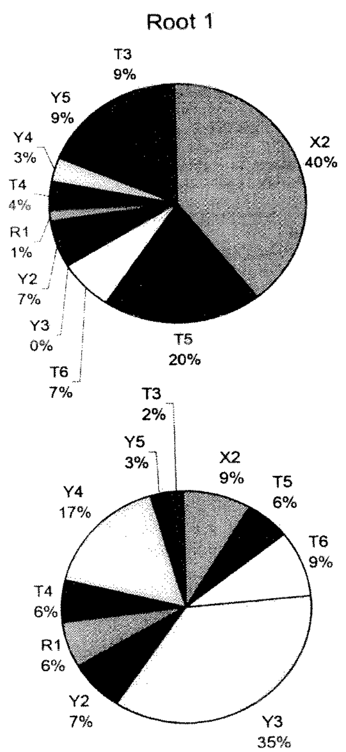


Fig. 4. Reliable contribution of different variables to discrimination of bench-mark sockeye salmon samples from three populations spawned in 1996 in Paratunka River. Above, first canonical function; below, second canonical function.



DISCUSSION

Statistically reliable differences in scale characteristics of sockeye salmon from three populations (two limnetic and one fluvial) indicated that limnetic growth of fish in these populations during embryogenesis and feeding was different.

There was much more similarity between the two limnetic populations than between them and the fluvial population. This indicates similar conditions for spawning and development during the first fresh-water year in lakes. However, the two lakes differed in geomorphology: Blidzneye Lake is a shallow-water lake, warming quickly in summer; Dalneye Lake is deeper and more stable in water temperature regime (Pogodayev 1995). Feeding of Blidzneye Lake juvenile sockeye salmon is poor compared to that of Dalneye Lake juvenile sockeye: in Blidzneye Lake abundance of competitors for food and predators (juvenile stickleback, *Gasteroteus aculeatus* and coho, *O. kisutch*) is several times greater than abundance of juvenile sockeye salmon (Pogodayev 1995).

Growth conditions for the juvenile sockeye salmon in Paratunka River and its tributaries are different to those in lakes. Rapid flow and powerful spring floods cause these juvenile sockeye salmon to migrate to the sea early. Consequently, only a few narrow fresh-water sclerites occur on the scales. Later, rapid growth of these sockeye in the ocean results in many wide marine sclerites (Fig. 2).

Complicated age structure of sockeye populations makes the discriminate analysis of scale characteristics difficult. Therefore, accuracy of classification of multiple-age samples, despite conducting variable selection procedures, is lower than classification accuracy of mono-age samples. Nevertheless, we were unable to collect enough mono-age samples for this study.

Classification accuracy of test sub-samples made with and without coordinate characteristics (Table 1) spoke in favor of using these additional characteristics. The more complex model, including both types of characteristics, was 1.3 times better at classifying sockeye populations than the radial model, and the cost of classification was 1.4 times lower.

It could be suggested that higher classification accuracy of complex models arises because there are more variables. However, simulations, carried out with the same number of variables, suggest otherwise: classification accuracy of learning samples in the model with five radial variables (R_1, T_3, T_4, T_5, T_6) was 62.5±4.1%, in the model with five coordinate variables (X_2, Y_2, Y_3, Y_4, Y_5) - 77.3±0.5%, and in the model with five complex variables (X_2, Y_3, T_3, T_4, T_5) - 82.3±1%.

Significant contribution of coordinate variables to the total discrimination of groups (Fig. 4) speaks in favor of using the variables as additional parameters for population differentiation.

ACKNOWLEDGEMENTS

We are sincerely grateful to our colleagues from KamchatNIRO: Eugeny Pogodayev, for age reading from adult sockeye scale and for consultations on scales structure features; Y. N. Orlova, for creation of scale images in OPRS; and A.A. Nikolayeva for assistance in our work.

REFERENCES

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 Pacific Fish. Comm. Bull. No. 26: 157–192.

Bernard, R.L., and K.W. Myers. 1994. The performance of quantitative scale pattern analysis in the identification of hatchery and wild steelhead. (NPAFC Doc. 62) 18p. FRI-UW-9410. Fisheries Research Institute, School of Fisheries, University of Washington, Seattle.

Bird, J.L., D.T. Eppler, and D.M. Checkly. 1986. Comparisons of herring otoliths using Fourier series shape analysis. Can. J. Fish. Aquat. Sci. 43: 1228–1234.

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

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.

Davis, N.D. 1987. Variable selection and performance of variable subsets in scale pattern analysis. (INPFC Doc. 3235) 47p. FRI-UW-8713. Fisheries Research Institute, University of Washington, Seattle.

Davis, N.D., K.W. Myers, R.V. Walker, and C.K. Harris. 1990. The Fisheries Research Institute's High-Seas Salmonid Tagging Program and Methodology for Scale Pattern Analysis. BioSonics, Inc., Am. Fish. Soc. Symp. 7: 863–879.

Jarvis, R.S., H.F. Klodowski, and S.S. Sheidon. 1978. New method of quantifying scale shape and an application to stock identification in walleye. Trans. Am. Fish. Soc. 107: 528–534.

Kayev, À.M. 1998. Scale identification of origin and life history of chum salmon *Oncorhynchus keta* from the Sea of Okhotsk. Vopr. Ichthiologii 38: 650–658.

Knudsen, C.M. 1985. Chinook salmon scale character variability due to body area sampled and possible effects on stock separation studies. Master's thesis. University of Washington, Seattle. 141p.

Millar, R.B. 1988. Statistical methodology for estimating composition of high seas salmonid mixtures using scale analysis. (INPFC Doc. 3284) 18p. FRI-UW-8801. Fisheries Research Institute, University of Washington, Seattle.

Myers, K.W. 1985. Racial trends in chinook salmon (*Oncorhynchus tshawytscha*) scale patterns. (INPFC Doc. 2913). 56p., University of Washington, Fisheries Research Institute, Seattle.

Nikolayeva, Ye.T., and N.I. Semenets. 1983. To the method of chum salmon *Oncorhynchus keta*, Walbaum (Salmonidae) identification from structure of first year scale zone. Vopr. Ichthiologii 23: 735–748.

Pogodayev, Y.G. 1995. Comparative description of the ichthiofauna of Paratunka's Lakes. Studies of biology and stock abundance dynamics of fishery species of Kamchatkan shelf. Transactions Petropavlovsk-Kamchatsky: KamchatNIRO. Issue IV. pp. 61–66.

Schwartzberg, M., and J.K. Fryer. 1993. Identification of hatchery and naturally spawning stocks of Columbia basin spring chinook salmon by scale pattern analyses. N. Am. J. Fish. Manage. 13: 263–271.

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.

Temnykh, O.S. 1998. Regional variability of Asian pink salmon sclerogrammes. Izv. TINRO 124: 375–391.

Zaporozhets, O.M., and G.V. Zaporozhets. 2000. Particularity and prospects for differentiating wild and hatchery salmon stocks from their scale structure characteristics. Russian-American Conf. on Salmon Conservation. October 1999. Khabarovsk 2000. pp. 64–65.