FRI-UW-8210
June 1982

VARIANCE ESTIMATES OF SOCKEYE SALMON PREDICTIONS WITH REFERENCE TO THE EGEGIK RIVER SYSTEM OF BRISTOL BAY, ALASKA

by

Linda K. Brannian, Ole A. Mathisen,
and D. A. McCaughran

FINAL REPORT
for the period
January 1, 1982 — June 30, 1982
Alaska Department of Fish and Game
Contract No. 82-0769

Approved

Submitted 23 June 1982

Robert L. Burgner
Director

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

Financial support of this research was provided by the Commercial Fisheries Division of the Alaska Department of Fish and Game. Mr. Charles Meacham and Mr. Henry Yuen deserve special thanks for providing historical data from the Egegik River sockeye salmon run. Finally, thanks are due Dr. John Clark, Chief Fisheries Scientist, for his support of this research.

INTRODUCTION

The magnitude of the sockeye salmon run to Bristol Bay, Alaska, has varied substantially through time, and when this variation is coupled with its short duration, the importance of accurate forecasts to management and industry becomes apparent. The run of sockeye salmon to Bristol Bay is managed for its harvest in five commercial fishing districts which exploit fish bound for eight major river systems (Fig. 1). Harvest levels are controlled to best fulfill predetermined escapement goals set for the major river systems.

The need for accurate forecasts of a year's salmon run to Bristol Bay is steadily gaining in importance due to the increasing costs of catching and processing the fish which emphasize the need for matching effort to expected run size. It has been over 15 years since the economic benefits provided by an individual river system type forecast for Bristol Bay sockeye salmon were analyzed in detail (Mathews 1967). The need is as great today with the addition of limited entry and pre-season price negotiations involving fishermen and processors. Timely forecasts are necessary as industry must commit itself on major processing and cannery preparations well in advance of the run. In addition, the remoteness of many harvest sites requires long-term planning for locating sufficient personnel and equipment.

Fisheries management also benefits from accurate and timely forecasts of the Bristol Bay sockeye salmon run. If managers wish to ensure a predetermined spawning escapement which includes all temporal and racial components of the run, an estimate of total run size is critical. The

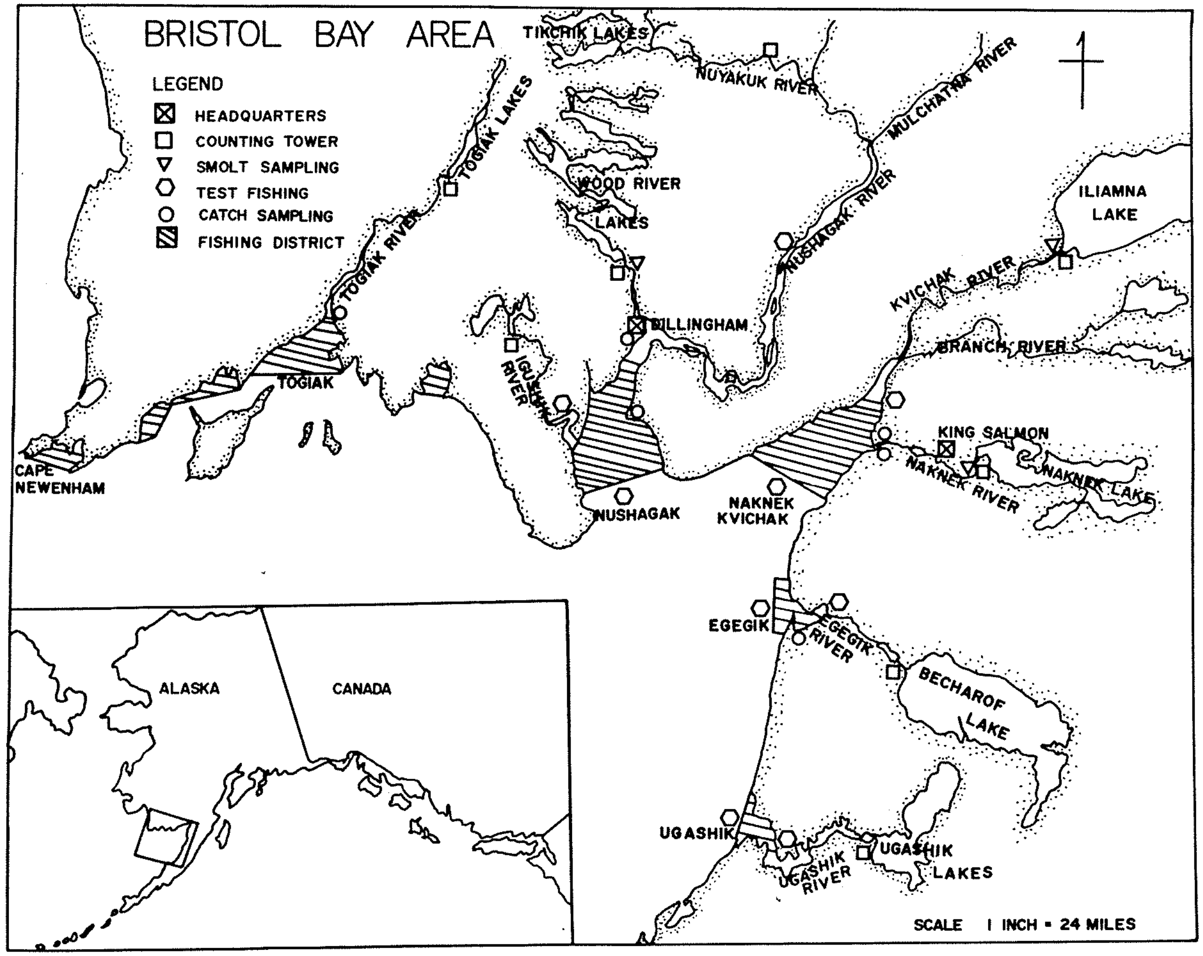

Fig. 1. Fishing districts and sampling programs in Bristol Bay, Alaska.

forecast return, therefore, becomes the basis for management decisions involving the level of harvest early in the season or until other data become available. Unfortunately, the moderate reliability of the forecasts has prompted management to turn to intraseasonal abundance estimation techniques in an attempt to increase the accuracy of harvest control. Yet, some pre-season budgetary decisions such as the use and duration of test fishing programs offshore and inshore should incorporate the forecast return.

Historically, the forecast return of Bristol Bay sockeye salmon has been presented as a point estimate with a range stated without the probability of encompassing the true return. It is comprised of forecasts for the individual river systems which represent the major sockeye salmon producing areas. Forecasts are generally above or below the true return. This is especially true if individual river systems are considered. However, with the point estimate is associated a variance and a probability that a given range set about the estimate will include the true return. This variance can then form the basis for a risk function which explicitly or implicitly is involved in any decision on investment of money or of time. Ideally, management and industry require a forecast with high precision which translates statistically as a forecast with a small variance.

The primary objective of this research was to develop variance estimates for the Bristol Bay sockeye salmon forecasts for future use and possible incorporation into risk functions. Conventional methods, used in the forecasting of sockeye salmon returns, would be investigated as to their models, underlying assumptions, and forecasting reliability.

In addition, alternative methods using readily available data were also investigated and comparisons made between methods. Recommendations followed on further research where additional data would be useful in increasing the precision of the forecast. Variances were estimated for the predictions resulting from these forecasting methods. Confidence intervals placed about the estimates were based on the conventional t-distribution at a 95 percent confidence level for some indication of the resulting variability. Further investigation and recommendations for decreasing the variances were also made. As with any forecasting procedure, the ultimate goal is a point estimate with an associated range. The final objective was to provide recommendations on the best approach for incorporating estimates from available forecasting methods into a final point estimate and its estimated variance.

The forecast of a year's return of sockeye salmon can be based on information obtained at different stages in its life history. Returns can be forecast for individual river systems based on a relationship involving the total return from the parental escapement or the smolt production. Only the class of estimators based on parental escapement or return by age class will be considered here. The Gilbert-Rich system of age class designation for sockeye salmon will be used where the age I(j) refers to fish of total age I with j years of freshwater residence. The difference (I - j) represents the number in years of marine residence. Only the return of the major age classes of 4(2), 5(2), 5(3), and 6(3) are estimated for an individual river system prediction (Fig. 2). A given year's return is therefore composed of age class returns from three brood years each with different amounts of information available. For

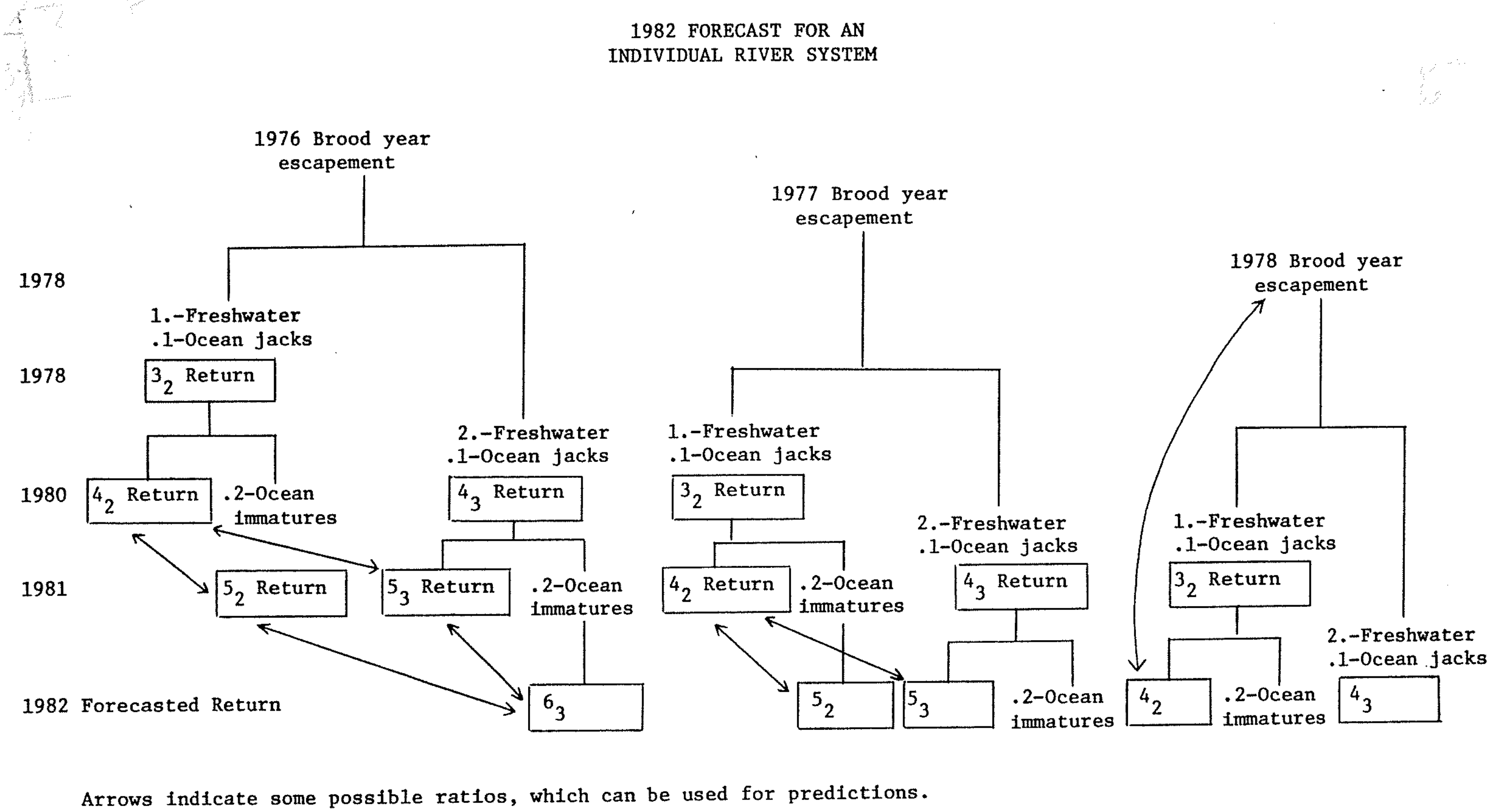

Fig. 2. Brood year components of a forecast of sockeye salmon return for an individual river system using the 1982 forecast in example.

example, the 1982 forecast for an individual river system involves the return as age 6(3) from brood year 1976; as age 5(2) and 5(3) from brood year 1977; and as age 4(2) from brood year 1978 (Fig. 2). Parental escapement is known for all these brood years but additional information on return of younger age classes will differ. The greatest amount of information is available for brood year 1976 where the return of the major age classes 4(2), 5(2), and 5(3) are known, and they will be useful for the prediction of age class 6(3). For brood year 1977 only the return as age 4(2) is known and lastly for brood year 1978 only the return of a minor age class 3(2) is known for use in predicting return as 4(2).

Two prediction methods will be considered in this research. The first is based on a spawner-recruit relationshp where it is generally accepted that one of the major population controls on sockeye salmon is the size of the spawning escapement. The fundamental theory states that a relationship exists between the number of spawners in a particular season, referred to as a brood year, and the total number of mature progeny of all ages that these spawners produce. The progeny or total brood year return mature at four major ages, necessitating for a year's forecast the prediction of the appropriate age class returns of selected brood years. The second class of estimation methods involves the relation between two different ocean-aged fish from the same brood year and smolt migration. The two ocean ages, (N + 1)-ocean and N-ocean, shared the same freshwater history and differ by one year of marine residence. It is then theorized that the observed return of N-ocean fish in a given year will provide survival information on the return of the (N + 1)-ocean fish from the same brood year and smolt migration, the next year. This relationship has been used to forecast the return of (N + 1)-ocean fish as a linear function of the return of N-ocean age fish. Two forms of this relation have been proposed, where one involves the regression of (N + 1)-ocean fish to N-ocean fish from the previous year, both groups of the same freshwater age, and the second proposes a zero-intercept linear model with a ratio to relate (N + 1) to N-ocean age return of sockeye salmon.

Data from the Egegik River sockeye salmon run will be used to evaluate the estimation procedures. This system does not have a continuous history of smolt data, and again that class of estimators will not be considered. The Egegik is a major single-river district, and it has been assumed within the forecasting procedure, that all fish caught in the district were bound for the Egegik River, unlike a multiple-river district, where catch allocation is necessary. Data were provided by the Alaska Department of Fish and Game in their sampling for age composition of catch and escapement, in conjunction with their escapement enumeration from counting towers and catch summaries from fish tickets.

ESTIMATION METHODS USED IN FORECASTING

Spawner-Recruit Estimator

The collection of accurate catch and escapement data for Alaskan sockeye salmon stocks has been facilitated by their anadromous return to major trunk streams. Since the compilation of such data for Bristol Bay sockeye salmon began in the early 1950's, the relationship between the spawning stock and subsequent return of recruits has been apparent. Historically, mean return per spawner has been used assuming an average relationship, constant over environmental conditions and escapement levels. Next, the concept of density regulation of a population was applied where constant production at all levels of escapement did not occur. Concepts of density-dependent or compensatory mortalities were developed to include any factor of mortality whose effectiveness increases with stock density. In salmon, the possibilities of these mortalities were theorized present in their early life history through smoltification. A discussion of such mechanisms and a resulting model incorporating the response in magnitude of the return from a given level of escapement can be found in Ricker's (1954) monumental paper on spawner-recruit relationships. Ricker's model has been extensively used in salmon research to relate stock to escapement. A further modification which is commonly used incorporates depensatory-compensatory sources of mortality, resulting in a three-parameter two-stage reproduction curve,

$$R = A E^{B} \exp(-CE) \qquad [1]$$

where R = return as recruits

E = spawning escapement

A, B, C = model coefficients

(Larkin et al. 1964, Ward and Larkin 1964, Ossiander 1967, Pennoyer 1970).

Estimates of the model coefficients (A, B, C) can be obtained from linear or non-linear regression of historical data. As the objective of this research involves evaluation and comparison of different forecasting estimators, a linear regression approach was taken. Linear regression provides minimum variance unbiased estimates of the model's coefficient when the independent variable is measured without error and provides for tests of significance for the parameters if normality of the errors can be assumed. In contrast, non-linear regression yields estimators with good asymptotic properties, but general tests of significance are not meaningful (Draper and Smith 1980). In this investigation, variance estimates and confidence intervals about the point estimates will be useful in comparing estimators and in combining them for a final estimate.

A linearized form of the spawner-recruit model [1] was obtained through a natural logarithm (log) transformation as

$$\log R = A + B \log E - CE$$

and a multiple linear regression was modeled as

$$Y_i = a + bX_{1i} + cX_{2i} + e_i \qquad [2]$$

with the assumptions

1. That the independent variables (X's) are measured without error.

2. The error term (e_i) has the following properties:

   $E(e_i) = 0$, the mean of the random error is zero.

   $V(e_i) = \sigma^2$, the variance of the errors is constant.

   $E(e_i e_j) = 0$ for all i, j such that $i \neq j$, as there is no covariance or correlation between observations.

Data from the Egegik sockeye salmon run were used to obtained unbiased estimates of (a, b, c) with the above regression and the log of total return was estimated as

$$\hat{Y}_i = \hat{a} + \hat{b}X_{1i} + \hat{c}X_{2i}$$

The variance of our estimate ($\hat{Y}_i$) is derived from regression theory, as $\hat{a}$ is estimated as a function of the mean values of Y, X_1, X_2, resulting in no covariance terms involving $\hat{a}$ and allowing for the covariance between $\hat{b}$ and $\hat{c}$. The variance of the predicted Y_i is estimated as

$$V(\hat{Y}_i) = S^2_{y \cdot x} \left(\frac{n+1}{n}\right) + (X_{1i} - \bar{X}_1)^2 V(\hat{b}) \qquad [3]$$

$$+ (X_{2i} - \bar{X}_2)^2 V(\hat{c})$$

$$+ 2(X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2) \text{ Cov } (\hat{b}, \hat{c})$$

(See Appendix A for full derivation)

where

$S^2_{y \cdot x}$ = the estimated mean square or residual variation of the regression (an unbiased estimate of σ^2)

n = sample size of the regression

X_{1i} = the log of the escapement for brood year i

$\overline{X}_1$ = average for the regression data (X_1's)

$V(\hat{b})$ = estimated variance of the regression coefficient b

X_{2i} = escapement for brood year i

$\overline{X}_2$ = average escapement for the regression data (X_2's)

$V(\hat{c})$ = estimated variance of the regression coefficient c

Cov $(\hat{b}, \hat{c})$ = estimated covariance between regression coefficients b and c

This is a variance for the log of total return, where total return (TR) is derived as $(TR)_i = \exp(\hat{Y}_i)$. No similar transformation of the variance is possible. To view the variability associated with the estimate, 95 percent confidence limits can be obtained for the transformed data, with the additional assumption of a normally distributed random error:

$$\hat{Y}_i \pm t_{(n-2)} [V(\hat{Y}_i)]^{\frac{1}{2}}$$

Here, $t_{(n-2)}$ is from the t-distribution with n - 2 degrees of freedom. The upper and lower limits of total return can then be derived as

$$\text{Upper limit} = \exp [\hat{Y}_i + t_{(n-2)} (V(\hat{Y}_i))^{\frac{1}{2}}]$$

$$\text{Lower limit} = \exp [\hat{Y}_i + t_{(n-2)} (V(\hat{Y}_i))^{\frac{1}{2}}]$$

Note that the 95 percent confidence interval is symmetrical about the estimate of the transformed data. It will not be symmetrical about the estimated total return.

It has been shown how coefficients from the spawner-recruit model [1] are derived and used to estimate total return from a given year's escapement. Yet an objective in sockeye salmon management is the forecasting of a given year's return of adults which is composed of several age classes, each of a particular brood year. It then becomes necessary to divide an estimated total brood year return into its major age classes, which will comprise components of several year of return forecasts. A common approach to the apportionment of brood year return into its major age classes is to model constant proportions as

$$R_{ij} = (TR)_i \, P_j + e_{ij}$$

where

R_{ij} = return at age j from brood year i

$(TR)_i$ = total return from brood year i

P_j = the proportion of TR which matures at age j

e_{ij} = random error

Again, random error may be associated with the process. This relationship is apparent from historical data, and proportions are assumed stable and predictable parameters of the sockeye salmon stock involved in the forecasting. Often the proportions are derived from historical data as an average over brood years (i's) as

$$\overline{P}_j = \sum_{i=1}^{n} P_{ij} = \sum_{i=1}^{n} \frac{R_{ij}}{(TR)_i} \qquad [4]$$

and updated yearly.

To illustrate the procedure, the estimated return of Egegik sockeye salmon from brood year 1970 will be apportioned into age class 5(3) for the 1975 forecast. Coefficients of the spawner-recruit model were estimated by regression of data from 1957 to 1969. The total return for 1970 was then estimated to be 1,796,242 sockeye salmon based on an escapement in 1970 of 919,734. To apportion the total return (TR) into age class 5(3), a mean proportion of 5(3) return was derived as in [4] to be 0.5008 based on data from 1957 to 1969. The return as age class 5(3) in 1975 from brood year 1970 was then estimated as

$$(\hat{R}_{5(3)})_{70} = (\hat{TR})_{70}\, \overline{P}_{5(3)}$$

$$= (1{,}796{,}242)(0.5008)$$

$$= 899{,}588$$

The estimated return of an age class is composed of two variance components, one due to the spawner-recruit model and the other due to apportionment. In deriving both the mean proportion and regression coefficients, identical data sets were used and one would expect covariance between the two terms. While the variance of the mean proportion could be based on ratio methods discussed later, the covariance term between the proportion and total return is not easily identifiable.

To more fully evaluate the predictive nature of this estimation process, age class returns were hindcast for the brood years 1965 to 1977. The historical data set began with brood year 1957 and a prediction was based on a minimum of 8 years, beginning in 1965. Each age class was predicted for a similar number of years with the result that age class 4(2) was hindcast for brood years 1967 - 1977, 5(2) and 5(3) for brood years 1966 - 1976 and age class 6(3) for brood years 1965 - 1975. The hindcasting of these brood years results in the components needed to predict total return as adults for the years 1971 through 1981. For example, the return hindcast for 1971 is composed of the 6(3) return from brood year 1965, the 5(2) and 5(3) return from brood year 1966, and the 4(2) return from brood year 1967.

To illustrate this procedure, begin by estimating the adult return in 1971. Coefficients of the spawner-recruit model are derived from the regression [2] with data from 1957 to 1964. The total return is then estimated for brood years 1965, 1966, and 1967. Next, the proportions of age class return as 4(2), 5(2), 5(3), and 6(3) are derived based on data from 1957 to 1964. Our estimates of return by age class then become the product of total return and the corresponding mean proportions as

$$(\hat{R}_{6(3)})_{65} = (\hat{TR})_{65}\ \overline{P}_{6(3)}$$

$$(\hat{R}_{5(3)})_{66} = (\hat{TR})_{66}\ \overline{P}_{5(3)}$$

$$(\hat{R}_{5(2)})_{66} = (\hat{TR})_{66}\ \overline{P}_{5(2)}$$

$$(\hat{R}_{4(2)})_{67} = (\hat{TR})_{67}\ \overline{P}_{4(2)}$$

The forecast return for 1971 based upon the spawner-recruit relationship becomes the sum of the return by age class derived above (R_{ij}'s).

Figure 3 presents the historical spawner-recruit data from 1957 to 1975. There is substantial scatter of the data points about the 45° replacement line or any spawner-recruit model. Understandably, other factors influence the magnitude of the return than just the level of escapement. As expected, the resulting regression for the spawner-recruit coefficients had only a moderate fit (Table 1) as judged by the coefficient of determination (R^2) which also decreases with the addition of data points. The mean proportions were also derived for hindcasting and are based on the same span of data as the regressions (Table 1). Tables 2 to 5 present the predicted return by age class and brood year. Confidence intervals are based only on the variance due to regression as discussed previously, ignoring for the moment the variance component due to apportionment. In this instance, the proportion is assumed to be measured without error and is treated as a constant. Yet the resulting 95 percent confidence intervals (Tables 2 to 5) are assumed to include the true return 95 percent of the time. This would indicate that on average, when forecasting over 11 years, one would expect the confidence interval not to include the observed return at most 0.55 or one time. Yet from hindcasting, the confidence intervals do not include the observed 4(2) return in three years (Table 2), the 5(2) return in four years (Table 3), and the 5(3) and 6(3) returns in two and one years, respectively (Tables 4 and 5). This would indicate that the variance due to the process of apportionment cannot be ignored. A variance for the mean proportions could be derived based on ratio techniques discussed later.

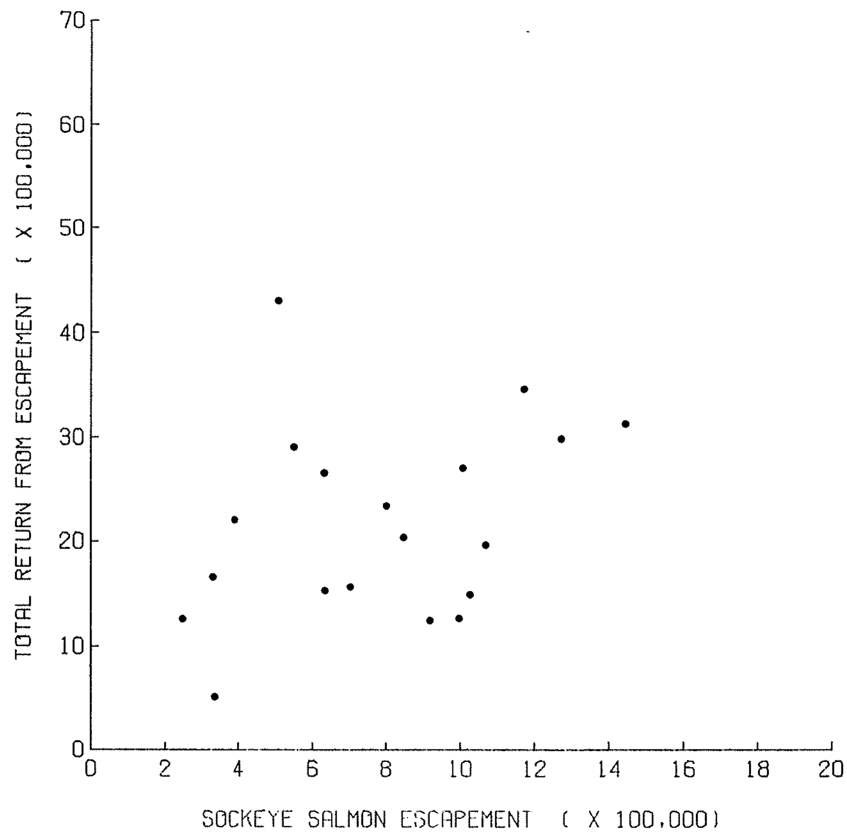

Fig. 3. Total return from brood year escapement for the Egegik sockeye salmon run 1957-1975.

Table 1. The fit of the spawner-recruit model and the mean proportion to return by age used in hindcasting the return of brood years 1965 to 1977.

Last year	Spawner-Recruit model		Mean proportion return at age			
in data set	R^2	$S_{y.x}$	4(2)	5(2)	5(3)	6(3)
1964	0.73	0.363	0.031	0.064	0.556	0.330
1965	0.74	0.335	0.031	0.059	0.566	0.327
1966	0.68	0.346	0.038	0.081	0.535	0.328
1967	0.69	0.324	0.037	0.087	0.526	0.332
1968	0.62	0.433	0.040	0.087	0.501	0.353
1969	0.62	0.419	0.037	0.083	0.498	0.362
1970	0.60	0.416	0.038	0.082	0.513	0.347
1971	0.52	0.442	0.036	0.079	0.517	0.348
1972	0.43	0.472	0.034	0.076	0.518	0.353
1973	0.42	0.459	0.035	0.075	0.508	0.362
1974	0.44	0.443	0.035	0.073	0.521	0.351
1975	0.46	0.435	0.035	0.072	0.538	0.344

Where R^2 = coefficient of determination

$S_{y.x}$ = the standard deviation due to regression.

Table 2. Age class 4(2) estimated as the product of total return from the spawner-recruit relationship and mean proportion expected to return as 4(2).

Brood year	Estimated 4(2)	95% confidence limits Upper limit	Lower limit
1967	27,291	121,422	14,583
1968	44,896	117,106	17,212
1969[1]	76,273	183,194	31,756
1970	67,998	150,908	30,640
1971	53,267	152,117	18,652
1972	46,853	126,981	17,287
1973	42,848	120,771	15,202
1974	103,764	290,527	37,061
1975	89,341	261,722	30,498
1976[1]	49,800	139,587	17,767
1977[1]	46,978	154,599	21,361

[1] Indicates the observed 4(2) return fell outside the confidence interval.

Table 3. Age class 5(2) estimated as the product of total return from the spawner-recruit relationship and mean proportion expected to return as 5(2).

Brood year	Estimated 5(2)	95% confidence limits Upper limit	Lower limit
1966[1]	96,569	277,525	33,603
1967[1]	79,107	195,862	31,950
1968[1]	120,601	313,398	46,409
1969	176,137	390,088	79,532
1970	155,734	442,934	54,756
1971	113,767	308,264	41,986
1972	100,607	266,679	37,954
1973	94,293	279,581	31,802
1974	219,501	651,827	73,916
1975	194,536	548,795	68,959
1976[1]	104,492	280,990	38,857

[1] Indicates the observed 5(2) return fell outside the confidence interval.

Table 4. Age class 5(3) estimated as the product of total return from the spawner-recruit relationship and mean proportion expected to return as 5(3)

Brood year	Estimated 5(3)	95% confidence limits Upper limit	Lower limit
1966	842,026	2,419,857	292,996
1967	759,640	1,880,812	306,811
1968[1]	796,562	2,069,975	306,531
1969	1,010,418	2,348,123	479,738
1970	899,558	2,558,493	316,282
1971	681,368	1,846,247	251,463
1972	628,519	1,666,015	237,115
1973	614,390	1,821,682	207,213
1974	1,502,534	4,461,910	505,973
1975	1,311,375	3,699,447	464,854
1976	747,948	2,011,315	278,140

1 Indicates the observed 5(3) return fell outside the confidence interval.

Table 5. Age class 6(3) estimated as the product of total return from the spawner-recruit relationship and mean proportion expected to return as 6(3).

Brood year	Estimated 6(3)	95% confidence limits Upper limit	Lower limit
1965	1,171,571	3,532,749	388,530
1966	491,465	1,210,888	199,471
1967	475,978	1,160,654	195,196
1968	496,696	1,182,085	208,705
1969	712,833	2,021,921	251,311
1970[1]	673,330	1,813,119	250,051
1971	455,947	1,206,715	171,275
1972	461,091	1,276,737	166,523
1973	414,232	1,421,585	144,203
1974	1,046,609	2,992,060	366,099
1975	909,425	2,446,588	338,044

1 Indicates the observed 6(3) return fell outside the confidence interval.

Yet how this would be combined with the variance of the log of total return and a covariance term derived is unclear. Instead, an alternative model was sought to predict return by age class where separate apportionment was unnecessary.

This alternative method still allowed for compensatory-decompensatory mortalities but included the variability associated with age of maturity. Therefore, an alternative spawner-recruit model was postulated with the substitution into [1] of $R_j = RP_j$. One can then solve for R_j as

$$R_j = \frac{A}{P_j} E^B \exp(-CE) = A' E^B \exp(-CE)$$

Parameter estimates of A', B, and C can again be obtained by a least-squares procedure with transformed data where separate regressions [2] are conducted, each with an age class return as the dependent variable (X's). This relationship results in an estimate based on the same assumptions as the original spawner-recruit model for total return, the difference being that it had previously taken an additional step to apportion total return. An estimated variance for the transformed prediction ($\hat{Y}_i$) can again be obtained [3], the difference being that this variance incorporates both sources of variability and the resulting 95 percent confidence intervals will most likely be wider to include the additional variability associated with age of maturity.

For the investigation, this model would only replace the more common procedure if it hindcast equally well and if the confidence interval included the observed returns in more years than the first model. Hindcasting was again conducted for brood years 1965 to 1967. The linearized form of the spawner-recruit model [2] was fit for each age class (j) and brood year (i). Estimated return of age classes 4(2), 5(2), and 5(3) are now based on additional information. In the case of estimating age class 4(2) from brood year i, total return is unknown for the last two years (i-1, i-2) as the five- and six-year-olds have not returned. Yet with this model using age-class return as the independent variable, the last two years for age class 4(2) can be used in the regression. One additional datum point is available for the estimation of each five-year-old age class. Regressions were conducted by age class and brood year, and they have been categorized (Table 6) by the last year in the data set. Note that the fit is poorer, as judged by the R^2 value, in this second model for those age classes which comprise a small percentage of total return as the 4(2) and 5(2). Yet the R^2 value remains steady through time unlike age class 6(3), where, with the addition of data points, the R^2 value decreases, indicating a less stable relationship. Lastly, in the dominating age class for Egegik of 5(3), the fit of the alternative model is at times better than the original. The return of an age class by brood year was then estimated based on coefficients estimated from regression (Tables 7 through 10). The variance of the transformed data was derived as in [3] and 95 percent confidence limits placed about the estimate. In the majority of years, the confidence interval is wider and includes the observed return in a greater number of years than the previous derivation.

The major evaluation of the spawner-recruit model with age-specific recruits was to be based on its error in prediction. Tables 11 through 14 compare estimates from the spawner-recruit model for total return with

Table 6. Comparison of the fit of the spawner-recruit model with total return versus specific age as recruits.

Data of brood year 1957 to:	Regression with transformed independent variable				
	Total return R^2	4(2) R^2	5(2) R^2	5(3) R^2	6(3) R^2
1964	.73				.47
1965	.74		.28	.71	.49
1966	.68	.27	.11	.73	.42
1967	.69	.27	.09	.74	.30
1968	.62	.29	.18	.57	.44
1969	.62	.23	.16	.57	.42
1970	.60	.22	.15	.57	.34
1971	.51	.23	.15	.49	.27
1972	.43	.22	.16	.40	.19
1973	.42	.21	.13	.41	.17
1974	.44	.23	.11	.45	.14
1975	.46	.24	.13	.47	.15

where R^2 = the coefficient of determination.

Table 7. Age class 4(2) estimated from the spawner-recruit relationship with return of 4(2) as the dependent variable.

Brood year	Estimated 4(2)	95% confidence limits Upper limit	Lower limit
1967	43,710	603,237	3,167
1968	43,036	557,285	3,323
1969	67,956	597,616	7,727
1970	49,131	513,848	4,698
1971	38,725	355,378	4,220
1972	38,113	309,128	4,699
1973	44,236	372,640	5,251
1974	83,156	589,853	11,723
1975	74,937	481,968	11,651
1976[1]	42,067	260,904	6,783
1977[1]	54,845	499,491	6,022

[1] Indicates the observed 4(2) return fell outside the confidence interval.

Table 8. Age class 5(2) estimated from the spawner-recruit relationship with return of 5(2) as the dependent variable.

Brood year	Estimated 5(2)	95% confidence limits Upper limit	Lower limit
1966[1]	86,601	535,076	14,016
1967	108,261	897,510	13,059
1968	85,274	895,189	8,123
1969	148,727	1,166,811	18,957
1970	136,900	949,683	19,735
1971	106,754	682,771	16,691
1972	97,464	562,620	16,884
1973	65,655	399,882	10,786
1974	145,582	735,646	28,810
1975	134,047	633,028	28,385
1976[1]	93,995	427,498	20,667

[1]Indicates that the observed 5(2) return fell outside the 95% confidence interval.

Table 9. Age class 5(3) estimated from the spawner-recruit relationship with return of 5(3) as the dependent variable.

Brood year	Estimated 5(3)	95% confidence limits Upper limit	Lower limit
1966	711,941	2,258,290	224,444
1967	615,879	1,753,992	216,253
1968[1]	744,077	2,030,847	272,621
1969	901,988	4,156,533	195,736
1970	788,449	3,285,968	189,184
1971	536,064	2,059,607	139,524
1972	550,641	2,257,360	134,319
1973	551,434	2,641,982	115,095
1974	1,519,269	6,321,812	365,113
1975	1,358,900	5,281,710	349,659
1976[1]	603,935	2,295,387	158,900

[1] Indicates that the observed 5(3) return fell outside the confidence interval.

Table 10. Age class 6(3) estimated from the spawner-recruit relationship with return of 6(3) as the dependent variable.

Brood year	Estimated 6(3)	95% confidence limits Upper limit	Lower limit
1965	1,026,925	6,179,263	170,664
1966	432,641	1,965,838	95,216
1967	433,270	1,851,923	101,367
1968	470,049	1,974,644	111,891
1969	606,354	2,142,040	171,642
1970	588,220	2,049,191	168,849
1971	404,292	1,698,329	96,243
1972	427,914	1,825,128	100,327
1973	487,386	2,431,601	97,691
1974	813,231	3,618,610	182,762
1975	694,849	1,838,977	262,545

Table 11. Comparison of estimates of age class 4(2) from spawner-recruit relationship with and without apportionment.

		Estimates of age class 4(2) from spawner-recruit relationship			
Brood year	Observed 4(2)	With total return as recruits and apportionment	% error	With 4(2) as recruits	% error
1967	55,739	27,291	-51.0	43,710	-21.6
1968	37,734	44,896	19.0	43,036	14.1
1969	12,429	76,273	513.7	67,956	446.8
1970	58,537	67,998	16.2	49,131	-16.1
1971	44,185	53,267	20.6	38,725	-12.4
1972	54,918	46,853	-14.7	38,113	-30.6
1973	73,702	42,848	-41.9	44,236	-40.0
1974	129,275	103,764	-19.7	83,156	-35.7
1975	144,195	89,341	-38.0	74,937	-48.0
1976	582,296	49,800	-91.5	42,067	-92.8
1977	793,985	46,978	-94.1	54,845	-93.1
		Σ =	218.6	Σ =	70.6
		Σ |error| =	920.4	Σ |error| =	851.1

Table 12. Comparison of estimates of age class 5(2) from spawner-recruit relationship with and without apportionment.

		Estimates of age class 5(2) from spawner-recruit relationship			
Brood year	Observed 5(2)	With total return as recruits and apportionment	% error	With 5(2) as recruits	% error
1966	664,641	96,569	-85.5	86,601	-87.0
1967	220,628	79,107	-64.1	108,261	-50.9
1968	41,940	120,601	187.6	85,274	103.3
1969	105,183	176,137	67.5	148,727	41.4
1970	86,167	155,734	80.7	136,900	58.9
1971	106,255	113,767	7.1	106,754	0.5
1972	58,338	100,607	72.5	97,464	67.1
1973	127,677	94,293	-26.2	65,655	-48.6
1974	97,738	219,501	124.6	145,582	49.0
1975	236,559	194,536	-17.8	134,047	-43.3
1976	775,728	104,492	-86.5	93,995	-87.9
		Σ =	259.9	Σ =	2.5
		Σ |error| =	820.1	Σ |error| =	637.9

Table 13. Comparison of estimates of age class 5(3) from spawner-recruit relationship with and without apportionment.

		Estimates of age class 5(3) from spawner-recruit relationship			
Brood year	Observed 5(3)	With total return as recruits and apportionment	% error	With 5(3) as recruits	% error
1966	604,374	842,026	39.3	711,941	17.7
1967	633,550	759,640	19.9	615,879	- 2.8
1968	118,779	796,562	570.6	744,077	526.4
1969	1,079,491	1,010,418	- 6.4	901,988	- 16.4
1970	787,985	899,588	14.2	788,449	0.1
1971	1,430,466	681,368	- 52.4	536,064	- 62.5
1972	1,516,911	628,519	- 58.6	550,641	- 63.7
1973	568,833	614,390	8.0	551,434	- 3.1
1974	2,178,973	1,502,534	- 31.0	1,519,269	- 30.3
1975	2,281,263	1,311,375	- 42.5	1,358,900	- 30.4
1976	2,886,554	747,948	- 74.1	603,935	- 79.0
		Σ =	387.0	Σ =	255.9
		Σ |error| =	917.0	Σ |error| =	832.5

Table 14. Comparison of estimates of age class 6(3) from spawner-recruit relationship with and without apportionment.

		Estimates of age class 6(3) from spawner-recruit relationship			
Brood year	Observed 6(3)	With total return as recruits and apportionment	% error	With 6(3) as recruits	% error
1965	911,224	1,171,571	28.6	1,026,925	12.7
1966	797,755	491,465	-38.4	432,641	-45.8
1967	555,019	476,978	-14.2	433,270	-21.9
1968	296,573	496,696	67.5	470,049	58.5
1969	1,101,694	712,833	-35.3	606,354	-45.0
1970	169,666	673,330	296.9	588,220	246.7
1971	926,825	455,947	-50.8	404,292	-56.4
1972	1,193,895	461,091	-61.4	427,914	-64.2
1973	842,214	414,232	-50.8	487,386	-42.1
1974	481,564	1,046,609	117.3	813,231	68.9
1975	776,539	909,425	17.1	694,849	-10.5
			Σ = 276.5		Σ = 100.9
		Σ |error| =	778.3	Σ |error| =	672.7

apportionment and the spawner-recruit model with age-specific recruits and no apportionment. In all cases the age class estimates without apportionment had a smaller cumulative percent error and cumulative absolute error. It predicted better or equal to the original method when viewing the error and had a more comprehensive variance estimate. Therefore, the estimate obtained without additional apportionment will be used when comparing estimates from other methods. The variance and 95 percent confidence interval obtained with this method were for the transformed data.

In order to compare these estimates with those from the regression and ratio methods of estimating an age class return, it became necessary to derive a variance estimate of the original data. Earlier, the errors were assumed normally distributed to facilitate the derivation of a confidence interval for the transformed data. This led to the assumption that the original data are log-normally distributed as

$$R_{ij} = \exp(Y_{ij})$$

where

R_{ij} = the return of age class j of brood year i

Y_{ij} = from regression [2] it is the log of R_{ij} and assumed normally distributed

From this, the predicted age class return has a log-normal distribution with a mean and variance which are functions of the mean and variance of the Y_{ij} distribution. Next, there exist unbiased estimates of the mean and variance of Y_{ij} resulting from the regression and from that a variance of $\hat{R}_{ij}$ can be estimated. A method of maximum likelihood was used (see Appendix B) to estimate the variance of each age class estimate (Aitchison and Brown 1957). The resulting values are quite large (Table 15), reflecting the large variability in the regression.

In summary, when viewing the historical data (Fig. 3) and the resulting fit of the spawner-recruit model, is becomes apparent that factors in addition to the level of escapement are influencing the magnitude of the return by age class. Whether these factors are environmental or inherited, some method of incorporating them into the spawner-recruit model would be helpful in reducing the variance associated with the predictions. The great variability in the system resulted in confidence intervals too wide to be useful. Other models have been proposed (Larken et al. 1964) with different shapes. A model acknowledging measurement errors in both spawners and recruits has been derived by Ludwig and Walters (1981), and a comparison with methods presented here can be found in Appendix C.

Regression Estimator for (N + 1)-Ocean Age Return

The estimation of the return as (N + 1)-ocean age adults from a given brood year can be based on a regression involving the magnitude of the return the previous year of ocean age N. Such a linear relationship between N and (N + 1)-ocean age returns is based on two main assumptions. The first one involves a constant or stable apportionment of a smolt migration for a given brood year into its major age classes. The second assumption involves the choice of two age classes of a definable stock which share the same mortality schedule as to length of stay in fresh

Table 15. Variances for the estimates of return by age class for brood years 1965 to 1977.

Brood year	Age class 4(2) ($X10^{10}$)	5(2) ($X10^{10}$)	5(3) ($X10^{11}$)	6(3) ($X10^{11}$)
1965				3.62
1966		0.82	10.02	3.35
1967	3.06	3.05	8.33	2.97
1968	1.92	2.91	7.00	2.49
1969	1.38	2.69	14.04	2.29
1970	1.52	2.20	12.23	2.50
1971	1.78	1.84	10.37	2.83
1972	0.94	1.60	10.30	2.86
1973	0.80	1.42	10.38	3.16
1974	0.72	1.31	9.15	3.04
1975	0.77	1.18	10.56	2.74
1976	0.85	1.29	12.05	
1977	1.81			

water and differ only by one year spent in the ocean. Therefore, the magnitude of the N-ocean return, given a stable apportionment, should provide valuable information about the survival conditions and the resulting magnitude of the (N + 1)-ocean return. Those forecasting (N + 1)-ocean returns have also assumed fairly stable survival for the last year of the (N + 1)-ocean fish in order that this N to (N + 1)-ocean relationship be comparable between brood years.

A linear relationship, allowing for a non-zero intercept, has been assumed between age classes of a given brood year. The major (N + 1)-ocean age classes predicted from the N-ocean ages have been

1. age class 4(2) as a linear function of 3(2) return
2. age class 5(2) as a linear function of 4(2) return
3. age class 5(3) as a linear function of 4(3) return
4. age class 6(3) as a linear function of 5(3) return

where the N and (N + 1)-ocean are of the same brood year. As these pairs share the same freshwater history and differ only by one year of marine residence, the return of N-ocean adults represents the survival conditions experienced up to that point by the (N + 1)-ocean fish. Therefore, if the proportion of N to (N + 1) is a stable parameter and does not vary with run size or environmental conditions, this relationship will have predictive usefulness. In addition, one needs to assume fairly stable survival in the last year for the (N + 1)-ocean return for the resulting relationship to have predictive value when based on historic data. Lastly, these assumptions may only be partly fulfilled, resulting in observed error or scatter of data about a model based upon them. Inappropriateness of these assumptions may also be indicated if a large error in forecasting results using this model or if a large variance is associated with the estimate of (N + 1)-ocean returns. In viewing the relationship between N and (N + 1)-ocean returns (Fig. 4) for the Egegik sockeye salmon run, one notes that the relationships are not straight lines and substantial scatter exists for the data. The usefulness of this model will be judged by its fit to historic data and its ability to accurately forecast.

The regression model commonly used is

$$Y_i = a + bX_i + e_i \quad [5]$$

where

Y_i = (N + 1)-ocean return from brood year i

X_i = n-ocean return from brood year i

a = model intercept

b = slope of the model relating Y_i to X_i

e_i = random error associated with the relationship

This model provides for a non-zero intercept (a) and assumes a non-zero slope (b). The assumptions for the regression are the following:

1. The independent variable (X_i) is measured without error.
2. The error term (e_i) has the properties

   $E(e_i) = 0$, the mean of the random error (e_i) is zero

   $V(e_i) = \sigma^2$, the variance of the errors is constant, though unknown

   $E(e_i e_j) = 0$ for all i, j, $i \neq j$, there is assumed to be no covariance or correlation between observations.

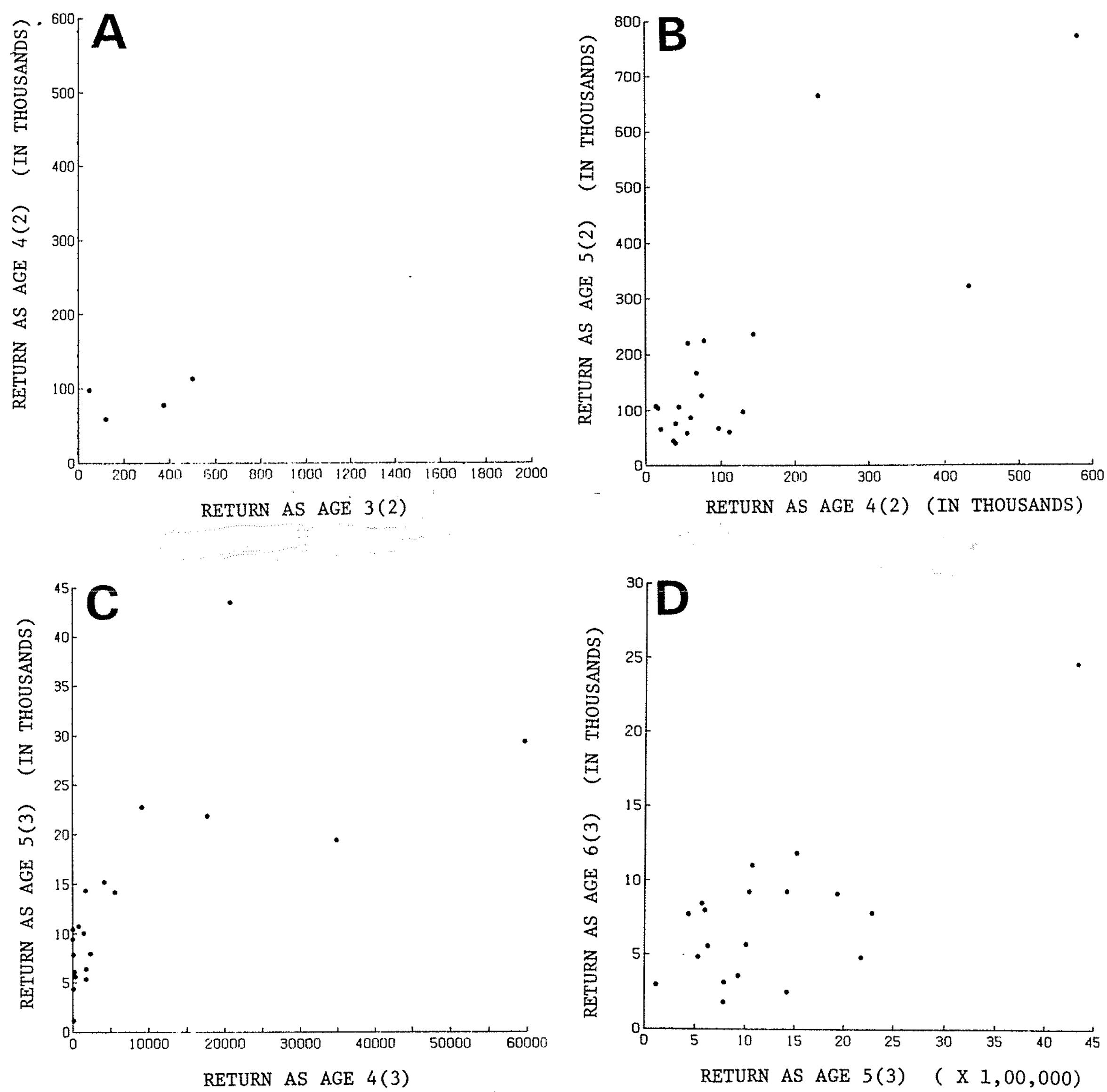

Fig. 4. The (N+1) to N-ocean relationships of Egegik sockeye salmon for 1957-1977. Age class 4(2) versus 3(2) in A, 5(2) versus 4(2) in B, 5(3) versus 4(3) in C, and 6(3) versus 5(3) in D.

Least square estimates of the slope ($\hat{b}$) and intercept ($\hat{a}$) can be obtained using standard regression techniques. The model [5] was fit to historical N-ocean and (N + 1)-ocean return data from the Egegik sockeye salmon run. A given return of (N + 1)-ocean adults from a given brood year is then estimated as

$$\hat{Y}_i = \hat{a} + \hat{b}(X_i) \qquad [6]$$

The estimated variance of $\hat{Y}_i$ is derived as

$$V(\hat{Y}_i) = V(\hat{a} + \hat{b}(X_i) + e_i)$$

$$= S^2_{y \cdot x} \left[1 + \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum_{j=1}^{n} (X_j - \bar{X})^2}\right] \qquad [7]$$

where

$S^2_{y \cdot x}$ = the estimated mean square or residual variation of the regression

n = sample size of the regression

X_i = N-ocean return from which Y_i is estimated

$\bar{X}$ = average N-ocean return for the regression data (X_j's)

(Sokol and Rohlf 1969, p. 425).

The predictive value of the regression estimator was evaluated through hindcasting for the brood years 1965 through 1977. The historic data span began with the 1957 brood year, and a prediction was based on a minimum of 8 years, beginning in 1965. So that each age class was evaluated over a similar number of years, age class 4(2) was hindcast for brood years 1967-1977, age class 5(2) and 5(3) for brood years 1966-1976, and age class 6(3) was hindcast for brood years 1965-1975. This resulted in the components needed to forecast a total year's return for 1971-1981. In order to estimate the (N + 1)-ocean return from brood year i, a regression [5] involving (N + 1) and N-ocean return was conducted (Tables 16-19) using data from 1957 to (i - 1) to estimate model coefficients (a, b). The (N + 1)-ocean return was then estimated as in [6] and the percent error from observed return was calculated.

Substantial scatter is present in the relationship between the N to (N + 1) data for the Egegik sockeye salmon run (Fig. 4). The regression for age class 4(2) had the poorest fit (Table 16) where the R^2 value was less than 0.3, as few observations existed with a non-zero 3(2) return. The regression was not significant in that the slope did not differ from zero, and the cumulative percent error in prediction was the largest over all age classes. The other regression based on the return of a minor age class was the 5(3) - 4(3) relationship (Fig. 4), which had a sligthly better fit (Table 18). Again, there were several years of no 4(3) return, and when the slope differed significantly from zero, so did the intercept. Again, there was substantial error in the predictions. A better fit was possible for the hindcasting of age classes 5(2) and 6(3) where a lower cumulative percent error in prediction was also present (Tables 17, 19).

A variance was estimated as in [7] and a 95 percent confidence interval placed about the estimate of (N + 1)-ocean return to view the associated variability. The 95 percent confidence interval assumes normally distributed errors (e_i) and is defined as

Table 16. Estimates of age class 4(2) based on the regression of 4(2) on 3(2).

Brood year	R^2	$S^2_{Y \cdot X}$	Estimated 4(2)	Observed 4(2)	% error	Sign. regr.	Non-zero intercept
1967	0.03	5.69XE08	92,182	55,739	65	NS	NS
1968	0.03	5.69XE08	92,182	37,734	144	NS	NS
1969	0.03	5.69XE08	92,182	12,429	642	NS	NS
1970	0.03	5.69XE08	92,182	58,537	60	NS	NS
1971	0.19	6.59XE08	73,895	44,185	67	NS	NS
1972	0.19	6.59XE08	73,895	54,918	35	NS	NS
1973	0.19	6.59XE08	73,895	73,702	0	NS	NS
1974	0.19	6.59XE08	73,895	129,275	- 43	NS	NS
1975	0.19	6.59XE08	73,895	144,195	- 49	NS	NS
1976	0.19	6.59XE08	98,229	582,296	- 83	NS	NS
1977	0.29	4.67XE10	995,091	793,985	25	NS	NS
				Σ =	913		
				Σ |error| =	1133		

Where: R^2 = Coefficient of determination

$S^2_{Y \cdot X}$ = Residual variation of the regression

% error = (Est. - Obs.)/Obs.

Sign. regr. = Significant regression which indicates that the slope (b) differed significantly from zero at the 95% level.

NS = Not signficant

S = Significant

Non-zero intercept = A 0.05 level test to determine if the intercept differs significantly from zero.

Table 17. Estimates of age class 5(2) based on the regression of 5(2) on 4(2).

Brood year	R^2	$S^2_{Y \cdot X}$	Estimated 5(2)	Observed 5(2)	% error	Sign. regr.	Non-zero intercept
1966	0.61	3.88XE 9	201,255	664,641	- 70	S	S
1967	0.40	2.49XE10	125,597	220,628	- 43	NS	NS
1968	0.38	2.31XE10	119,524	41,940	185	S	NS
1969	0.39	2.13XE10	87,193	105,183	- 18	S	NS
1970	0.40	1.94XE10	132,338	86,167	54	S	NS
1971	0.40	1.79XE10	115,197	106,255	+ 9	S	NS
1972	0.41	1.65XE10	124,633	58,338	114	S	NS
1973	0.41	1.57XE10	138,070	127,677	+ 8	S	NS
1974	0.41	1.46XE10	190,694	97,738	95	S	NS
1975	0.40	1.42XE10	198,585	236,559	- 16	S	NS
1976	0.40	1.35XE10	616,254	775,728	- 21	S	NS
				Σ =	307		
				Σ |error| =	633		

Where: R^2 = Coefficient of determination

$S^2_{Y \cdot X}$ = Residual variation of the regression

% error = (Est. - Obs.)/Obs.

Sign. regr. = Significant regression which indicates that the slope (b) differed significantly from zero at the 95% level.

NS = Not significant

S = Sigifnicant

Non-zero intercept = A 0.05 level test to determine if the intercept differs significantly from zero.

Table 18. Estimates of age class 5(3) based on the regression of 5(3) on 4(3).

Brood year	R^2	$S^2_{Y \cdot X}$	Estimated 5(3)	Observed 5(3)	% error	Sign. regr.	Non-zero intercept
1966	0.36	1.56XE12	1,002,543	604,374	66	NS	NS
1967	0.41	1.28XE12	1,015,200	633,550	60	NS	NS
1968	0.44	1.08XE12	828,461	118,779	697	NS	NS
1969	0.44	1.08XE12	881,537	1,079,491	- 18	NS	NS
1970	0.44	9.33XE11	859,712	787,985	9	NS	NS
1971	0.44	9.33XE11	977,536	1,430,466	- 32	NS	NS
1972	0.42	8.39XE11	1,187,179	1,516,911	- 22	S	S
1973	0.41	7.56XE11	981,801	568,833	73	S	S
1974	0.43	6.96XE11	2,066,584	2,178,973	- 5	S	S
1975	0.46	6.34XE11	1,524,291	2,281,263	- 33	S	S
1976	0.45	6.25XE11	4,981,574	2,886,554	73	S	S
				Σ =	868		
				Σ |error| =	1088		

Where: R^2 = Coefficient of determination

$S^2_{Y \cdot X}$ = Residual variation of the regression

% error = (Est. - Obs.)/Obs.

Sign. regr. = Significant regression which indicates that the slope (b) differed significantly from zero at the 95% level.

NS = Not significant

S = Significant

Non-zero intercept = A 0.05 level test to determine if the intercept differs significantly from zero.

Table 19. Estimates of age class 6(3) based on the regression of 6(3) on 5(3).

Brood year	R^2	$S^2_{Y \cdot X}$	Estimated 6(3)	Observed 6(3)	% error	Sign. regr.	Non-zero intercept
1965	0.80	1.24XE11	1,084,985	911,224	19	S	NS
1966	0.79	1.10XE11	388,746	797,755	- 51	S	NS
1967	0.75	1.14XE11	460,354	555,019	- 17	S	NS
1968	0.76	1.02XE11	226,277	296,573	- 24	S	NS
1969	0.77	9.26XE10	690,719	1,101,694	- 37	S	NS
1970	0.74	9.84XE10	585,564	169,666	245	S	NS
1971	0.72	1.03XE11	860,691	926,825	- 7	S	NS
1972	0.72	9.58XE10	907,080	1,193,895	- 24	S	NS
1973	0.72	9.44XE10	464,673	842,214	- 45	S	NS
1974	0.69	9.68XE10	1,257,520	481,564	161	S	NS
1975	0.58	1.24XE11	1,208,306	776,539	56	S	NS
				Σ =	276		
				Σ |error| =	686		

Where: R^2 = Coefficient of determination

$S^2_{Y \cdot X}$ = Residual variation of the regression

% error = (Est. - Obs.)/Obs.

Sign. regr. = Significant regression which indicates that the slope (b) differed significantly from zero at the 95% level.

NS = Not significant

S = Significant

Non-zero intercept = A 0.05 level test to determine if the intercept differs signicantly from zero

$$\hat{Y}_i \pm t_{n-2} [V(\hat{Y}_i)]^{\frac{1}{2}}$$

where t_{n-2} represents the value at the 97.5 percentile of the t-distribution with n-2 degrees of freedom. Though the confidence interval is symmetrical about the estimate, it is customary to place a lower bound of zero, as negative values will not occur. The result of the regressions were large variances and wide confidence intervals of up to 300 percent (Tables 20-23).

The moderate to poor fit of the regressions, large variances, and wide confidence intervals for the (N + 1)-ocean estimates would indicate some weakness in this linear N to (N + 1)-ocean relationship. The large residual variation in the system requires explanation by additional factors in order to develop estimates with greater precision. The proportion of (N + 1) to N-ocean return seems to vary through time. It needs to be determined whether this variation can be explained by environmental conditions affecting survival, run size, or varying genetic make-up of the spawning stock. Key environmental measures indicative of survival of sockeye salmon need to be developed and tested as to their usefulness in describing residual variation of the model. An example of an attempt to explain some of the residual variation of the original regression model involves the incorporation of a second variable. Initially the model had assumed constant proportions of each age class to return, where in effect one is assuming these to be inheritable characteristics of the stock. This can be viewed in two cases where each spawner, regardless of its age, has offspring which mature in these proportions or that the age of maturity is itself inheritable such that

Table 20. Age class 4(2) estimated from regression of 4(2) on 3(2) with a 95% confidence interval.

Brood year	Estimated 4(2)	95% confidence limits Upper limit	Lower limit	Standard deviation	n
1967	92,182	542,356	0	35,430	3
1968	92,182	542,356	0	35,430	3
1969	92,182	542,356	0	35,430	3
1970	93,683	482,562	0	30,606	3
1971	73,895	220,541	0	34,080	4
1972	73,895	220,541	0	34,080	4
1973	73,895	220,541	0	34,080	4
1974	73,895	220,541	0	34,080	4
1975	73,895	220,541	0	34,080	4
1976[1]	98,299	241,795	0	33,348	4
1977	995,091	3,426,324	0	764,058	5

[1] Indicates the observed 4(2) return fell outside the confidence interval.

Table 21. Age class 5(2) estimated from regression of 5(2) on 4(2) with a 95% confidence interval.

Brood year	Estimated 5(2)	95% confidence limits Upper limit	Lower limit	Standard deviation	n
1966[1]	201,255	365,632	36,878	69,504	9
1967	125,597	511,370	0	167,291	10
1968	119,524	483,620	0	160,962	11
1969	87,193	433,504	0	155,436	12
1970	132,338	451,623	0	145,041	13
1971	115,197	419,245	0	139,536	14
1972	124,633	412,671	0	133,351	15
1973	138,070	414,874	0	129,046	16
1974	190,694	457,350	0	125,132	17
1975	198,585	460,596	0	123,590	18
1976	616,254	999,149	233,359	181,467	19

[1] Indicates the observed 5(2) return fell outside the confidence interval.

Table 22. Age class 5(3) estimated from regression of 5(3) on 4(3) with a 95% confidence interval.

Brood year	Estimated 5(3)	95% confidence limits Upper limit	Lower limit	Standard deviation	n
1966	1,002,543	4,948,341	0	1,421,397	6
1967	1,015,200	4,196,111	0	1,237,227	7
1968	828,461	3,609,063	0	1,136,331	8
1969	881,537	3,648,404	0	1,130,718	8
1970	859,712	3,323,165	0	1,041,629	9
1971	977,536	3,418,866	0	1,032,275	9
1972	1,187,179	3,409,640	0	963,773	10
1973	981,801	3,069,154	0	922,791	11
1974	2,066,584	4,092,023	41,145	909,084	12
1975	1,524,291	3,344,976	0	827,208	13
1976	4,981,574	8,009,608	1,953,540	1,389,644	14

Table 23. Age class 6(3) estimated from regression of 6(3) on 5(3) with a 95% confidence interval.

Brood year	Estimated 6(3)	95% confidence limits Upper limit	Lower limit	Standard deviation	n
1965	1,084,985	2,012,885	157,085	379,199	8
1966	388,746	1,235,123	0	357,876	9
1967	460,354	1,291,669	0	360,501	10
1968	226,277	1,016,805	0	349,482	11
1969	690,719	1,396,563	0	316,806	12
1970	585,564	1,305,025	0	326,879	13
1971	860,691	1,588,412	132,970	333,970	14
1972	907,080	1,600,799	213,361	321,166	15
1973	464,673	1,152,055	0	320,458	16
1974[1]	1,257,520	1,964,168	550,872	331,604	17
1975	1,208,306	2,002,770	413,842	374,747	18

[1] Indicates that the observed 6(3) return fell outside the confidence interval.

the proportion of age class j in the spawning stock will influence the proportion of age class j in the subsequent return as offspring. The latter proposition was tested in a multiple regression where a second variable representing proportion of age class X was introduced (Tables 24 and 25).

$$Y_i = a + bX_{1i} + cX_{2i} + e_i$$

where an additional variable was added to the original model [2] as

X_{2i} = the proportion of spawners in year i which are of the same age as the X_1 return

The only age classes (X's) with a consistent representation in escapement samples were the 4(2) and 5(3) age classes used to predict 5(2) and 6(3) returns. Two regressions were conducted using 5(2) and 4(2) data from 1957 to 1975. The addition of the variable representing the proportion of 4(2) in the spawning stock explained 30 percent of the residual variance over our original model (Table 24). Similarly, the variable representing the proportion of 5(3) age spawners explained an additional 15 percent of the variation (Table 25). The next step would be to compare this model's hindcasting ability and variance estimates for the predictions. It is illustrative of the need to improve the simple linear regression model in order to improve the precision of our estimates.

Table 24. Comparison of estimates of age class 5(2) from two regressions (in thousands).

Brood year	Observed 5(2)	Estimated 5(2) from 4(2)	% error	from 4(2) and % 4(2) spawners	% error
1957	44	98	123	81	84
1958	75	101	35	89	19
1959	165	128	- 22	147	- 11
1960	321	473	47	316	- 2
1961	224	138	- 38	179	- 20
1962	66	84	27	32	- 52
1963	103	79	- 23	12	- 88
1964	61	171	180	64	- 5
1965	68	157	131	211	68
1966	665	285	- 57	500	- 25
1967	221	117	- 47	134	- 65
1968	42	100	138	41	- 03
1969	105	76	- 28	24	- 77
1970	86	120	40	103	20
1971	106	106	0	91	- 14
1972	58	117	102	122	110
1973	128	134	5	162	27
1974	98	187	115	257	162
1975	237	201	- 15	306	29
		Σ =	713	Σ =	157
		Σ |error| =	1,173	Σ |error| =	881
		R^2 =	0.40	R^2 =	0.70
		S.D. =	116	S.D. =	116

Table 25. Comparison of estimates of age class 6(3) from two regressions (in thousands).

Brood year	Observed 6(3)	Estimated 6(3) from 5(3)	% error	from 5(3) and % 5(3) spawners	% error
1957	928	668	- 28	995	7
1958	312	567	82	449	44
1959	566	654	16	503	- 11
1960	2,475	198	- 92	2,173	- 12
1961	767	425	- 45	640	- 17
1962	356	624	75	556	56
1963	484	463	- 4	480	- 1
1964	249	819	229	740	66
1965	911	102	- 88	673	- 26
1966	798	493	- 38	749	- 6
1967	555	504	- 9	479	- 14
1968	297	300	1	269	- 9
1969	1,102	681	- 38	479	- 57
1970	170	565	70	283	66
1971	927	820	- 12	939	1
1972	1,194	854	- 28	841	- 30
1973	842	478	- 43	793	- 6
1974	482	1,117	132	893	85
1975	777	1,158	49	1,296	67
		Σ =	229	Σ =	203
		Σ |error| =	1,079	Σ |error| =	581
		R^2 =	0.55	R^2 =	0.70
		S.D. =	355	S.D. =	300

Ratio Estimator for (N + 1)-Ocean Age Return

The estimation of the return from a given brood year of ocean age (N + 1) from that of the return of ocean age (N) can be based on regression or ratio estimation techniques. Like estimation using a linear regression equation, ratio estimation assumes a linear relationship between ocean age (N + 1) and (N). Two age classes from the same brood year are chosen which share the same freshwater history and differ only by one year spent in the ocean. This relationship is appealing for two age classes that share the same mortality schedules up to the return of the N-ocean fish. An important assumption underlying this relationship is that a constant ratio of (N + 1) to N-ocean age recruits will return from a given brood year and river system. Under this assumption, the return of N-ocean fish from a given brood year will provide important information on the survival conditions in both the freshwater and oceanic environments. The magnitude of its return can be used to predict the magnitude of the following year's (N + 1)-ocean return. The appropriateness of this assumption could be viewed in relation to the ability to predict (N + 1)-ocean return. The resulting residual between observed and predicted return should be random and uncorrelated between observations. In addition, the residuals should be uncorrelated with environmental conditions or other factors thought to possibly affect this ratio and negate the assumption.

Traditionally, the major age classes have been forecast using this relationship where information on abundance of N and (N + 1)-ocean return is available. The major age classes used in the prediction of the Egegik sockeye salmon run are

1. The return of age class 4(2) as predicted from the previous year's return of age class 3(2), when available.
2. The return of age class 5(2) as predicted from the previous year's return of age class 4(2).
3. The return of age class 5(3) as predicted from the previous year's return of age class 4(3), when available.
4. The return of age class 6(3) as predicted from the previous year's return of age class 5(3).

Again, the N and (N + 1)-ocean returns have the same freshwater age and differ only in one year spent in the ocean.

The underlying model for predicting (N + 1)-ocean from N-ocean of a given brood year is the same for the raio and regression estimators. It differs only in the assumption of the intercept and assumptions involving parameters for the distribution of the random error. Ratio estimation is based on the model

$$Y_i = RX_i + e_i$$

where

Y_i = (N + 1)-ocean return from brood year i

X_i = N-ocean return from brood year i

R = ratio relating Y to X

e_i = random error

This model assumes a zero intercept where the regression model had no such condition.

The mean ratio is commonly used to estimate the true ratio (R) of the above model and makes the following assumptions on the random error (e_i):

1. $E(e_i) = 0$
2. $V(e_i) = X_i^2 \sigma^2$, the variance of the error is proportional to the square of X_i
3. $E(e_i e_j) = 0$ for all i, j, $i \neq j$, there is assumed to be no covariance between observations

Making appropriate transformations for X and Y in the model and using least squares theory (see Appendix D), an unbiased estimate of R becomes

$$\hat{R} = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i}{X_i}$$

In addition, $S^2_{Y \cdot X}$ is an unbiased estimate of σ^2, where

$$S^2_{Y \cdot X} = \frac{1}{n-1} \left[\sum_{i=1}^{n} \left(\frac{Y_i}{X_i}\right)^2 - n\hat{R}^2 \right]$$

To derive an estimated variance for $\hat{R}$, note that

$$\hat{R} - R = \frac{1}{n} \left(\sum_{i=1}^{n} \frac{e_i}{X_i} \right)$$

which implies that

$$V(\hat{R}) = \frac{1}{n^2} \sum_{i=1}^{n} \frac{V(e_i)}{X_i^2}$$

$$= \frac{S^2_{Y \cdot X}}{n}$$

Given an estimate of R based on historical Egegik sockeye salmon data from 1957 to (i-1), age class Y of brood year (i) is predicted as

$$\hat{Y}_i = \hat{R} X_i$$

The variance for the estimated return must also include the variance of the random error associated with that observation and is estimated as

$$V(\hat{Y}_i) = V(\hat{R} X_i + e_i)$$

$$= X_i^2 \frac{S^2_{Y \cdot X}}{n} + X_i^2 S^2_{Y \cdot X}$$

A desirable property of this procedure is the unbiasedness of the estimates of the ratio (R) and the error variance (σ^2). Yet strong assumptions have been made to impose a zero intercept on the relationship and that the error variance be proportional to the square of the N-ocean observation (X_i).

The predictive nature of the mean ratio estimator was evaluated through hindcasting for the returns from brood years 1965 through 1977. Data began with the 1957 brood year, and a prediction was based on a minimum of eight years. Table 26 presents the average ratios used in predicting the return for the major age classes, along with the estimated standard deviation and the sample size on which it is based. The ratio of a given (N + 1)-ocean return from brood year (i) is based on an average over the years 1957 to i-1. The ratio for 1965 (Table 26) of 6(3) to 5(3) was 0.71 based on 8 observations (n) over the years 1957 to 1964. The return of 6(3) from the 1965 brood year is then predicted as

Table 26. Average ratio of (N+1) to N-ocean return.

Brood year	4(2) to 3(2) Ratio	S_R	n	5(2) to 4(2) Ratio	S_R	n	5(3) to 4(3) Ratio	S_R	n	6(3) to 5(3) Ratio	S_R	n
1965										0.71	0.18	8
1966				2.319	0.68	9	297.7	76.8	6	0.68	0.16	9
1967	809.6	752.6	3	2.373	0.61	10	609.0	318.2	7	0.74	0.15	10
1968				2.517	0.57	11	576.4	277.3	8	0.76	0.14	11
1969				2.400	0.53	12	576.4	277.3	8	0.90	0.19	12
1970				2.866	0.68	13	665.2	260.2	9	0.91	0.18	13
1971	725.1	429.1	4	2.766	0.63	14	665.2	260.2	9	0.86	0.17	14
1972				2.742	0.59	15	679.3	233.2	10	0.85	0.16	15
1973				2.637	0.56	16	650.7	213.0	11	0.84	0.15	16
1974				2.584	0.53	17	708.3	202.6	12	0.88	0.15	17
1975				2.482	0.51	18	663.3	192.0	13	0.84	0.14	18
1976				2.438	0.49	19	633.5	180.0	14			
1977	810.2	343.1	5									

Where: S_R = Standard deviation of the ratio

n = Sample size on which ratio is based

$\hat{Y}_{65} = \hat{R}_{65} X_{65}$ and a variance derived for the estimate $(V(\hat{Y}_i))$ using the above-mentioned formula.

Estimation was done for age classes 4(2), 5(2), 5(3), and 6(3) based on the mean ratio estimator (Tables 27-30). Few observations exist on the abundance of age class 3(2), minimizing its usefulness in estimating 4(2) return (Table 27). A 95 percent confidence interval was placed about each estimate as

$$\hat{Y}_i \pm t_{(n-1)} [V(\hat{Y}_i)]^{\frac{1}{2}}$$

where the $\hat{Y}_i$ are assumed normally distributed such that $\hat{Y}_i/[V(\hat{Y}_i)]^{\frac{1}{2}}$ conforms to a t-distribution with n-1 degrees of freedom (t_{n-1}). Substantial variability is evident in the width of the confidence intervals (Tables 27-30) about the predicted returns at times greater than 200 percent. By the equation above, the confidence interval is symmetrical about the estimate, though it is customary to place a lower limit of zero, as negative values do not occur.

There exists an alternative method for estimating the ratio, which is the more common method in sampling theory (Cochran 1977) and results in the ratio of the mean of Y and X. This method is based on the same model and differs only in the assumptions made on the random error (e), where

$$Y_i = RX_i + e_i$$

and the assumptions become

Table 27. Age class 4(2) estimated from average ratio 4(2)/3(2).

Brood year	Estimated 4(2)	95% confidence limits Upper limit	Lower limit	Standard deviation
1967				
1968				
1969				
1970	100,393	735,325	0	147,556
1971				
1972				
1973				
1974				
1975				
1976	366,896	1,911,720	0	485,488
1977	1,415,467	5,491,627	0	1,468,357

Table 28. Age class 5(2) estimated from average ratio 5(2)/4(2).

Brood year	Estimated 5(2)	95% confidence limits Upper limit	Lower limit	Standard deviation
1966	539,923	1,688,779	0	498,203
1967	132,269	385,897	0	112,125
1968	94,976	260,234	0	74,173
1969[1]	28,830	82,192	0	23,790
1970	167,767	490,080	0	147,918
1971	122,216	356,248	0	108,348
1972	150,585	428,584	0	129,603
1973	194,352	558,125	0	170,705
1974	334,046	950,798	0	290,921
1975	357,891	1,034,645	0	320,737
1976	1,419,638	4,071,846	0	1,262,355

[1] Indicates observed run fell outside confidence interval.

Table 29. Age class 5(3) estimated from average ratio 5(3)/4(3).

Brood year	Estimated 5(3)	95% confidence limits Upper limit	Lower limit	Standard deviation
1966[1]	72,631	200,017	0	49,547
1967	1,107,771	5,109,601	0	1,635,403
1968	0			
1969	452,482	1,996,289	0	652,773
1970	0			
1971	1,179,995	4,545,991	0	1,459,690
1972	2,823,170	10,089,782	0	3,212,472
1973	275,901	972,495	0	312,655
1974	12,511,766	40,920,498	0	12,907,193
1975	5,850,966	20,293,423	0	6,628,021
1976	37,777,000	127,534,922	0	41,557,834

[1]Indicates the observed 5(3) return fell outside the confidence interval.

Table 30. Age class 6(3) estimated from average ratio 6(3)/5(3).

Brood year	Estimated 6(3)	95% confidence limits Upper limit	Lower limit	Standard deviation
1965	1,373,919	3,800,977	0	1,026,240
1966	410,974	1,101,439	0	299,421
1967	468,827	1,205,315	0	325,592
1968[1]	90,272	220,183	0	58,308
1969	971,542	2,628,437	0	752,792
1970	717,066	1,857,440	0	523,348
1971	1,230,201	3,277,055	0	947,617
1972	1,289,374	3,372,874	0	971,329
1973	477,847	1,227,584	0	351,824
1974	1,917,496	4,769,504	0	1,345,287
1975	1,916,260	4,932,933	0	1,429,703

[1]Indicates observed run fell outside confidence interval.

$$E(e_i) = 0$$

$$V(e_i) = X_i \sigma^2$$

$$E(e_i e_j) = 0 \text{ for all } i, j, i \neq j$$

Under the conditions that the relation between Y_i and X_i is a straight line through the origin and that the variance of Y_i about this line is proportional to X_i, this becomes the "best linear estimator" (Cochran 1977). Here "best" refers to the property of having the smallest variance in that class of estimators. In this case, after appropriate transformation and use of least square theory (see Appendix D), R is estimated from observed data as

$$\hat{R} = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i}$$

and the error variance (σ^2) is estimated as

$$S^2_{Y \cdot X} = \frac{1}{n-1} \left[\sum_{i=1}^{n} \frac{Y_i^2}{X_i} - \hat{R} \sum_{i=1}^{n} Y_i \right]$$

The variance for the estimated ratio becomes

$$V(\hat{R}) = S^2_{Y \cdot X} \Big/ \sum_{i=1}^{n} X_i$$

Given an estimate of the ratio (R) based on historical sockeye salmon data from Egegik, brood year 1957 to (i-1), an age class return from brood year i can be predicted as

$$\hat{Y}_i = \hat{R} X_i$$

The variance for the forecast return must also incorporate the variance of the random error involved in the prediction. The estimated variance for the prediction becomes

$$V(Y_i) = V(\hat{R} X_i + e_i)$$

$$= \frac{X_i^2 \, S^2_{Y \cdot X}}{\sum_{i=1}^{n} X_i} + X_i \, S^2_{Y \cdot X}$$

To compare the precision in the two ratio estimators, returns from brood years 1965 through 1977 were also hindcast using the ratio of means to estimate R (Table 31). Note that the ratio of means (Table 31) is constantly less than that estimated by the mean ratio (Table 26). The ratio of means in effect weights a year's data by their magnitude unlike the mean ratio in giving each observation pair (Y_i, X_i) equal weight. In the last few years in Egegik, with radically changing relationships for the 4(2) and 6(3) age classes, the run has been of such a magnitude to affect the ratio much quicker in the ratio of means. The predicted returns for the brood years 1965 to 1977 (Tables 32-35) are in general similar to those of the mean ratios (Tables 27-30). Yet the variance for the predicted return based on the ratio of means is consistently smaller than that from the mean ratio, resulting in estimates of greater precision. The 95 percent confidence interval placed about each estimate is

Table 31. Ratio of mean (N+1)-ocean to mean N-ocean return.

Brood year	4(2) to 3(2) Ratio	S_R	n	5(2) to 4(2) Ratio	S_R	n	5(3) to 4(3) Ratio	S_R	n	6(3) to 5(3) Ratio	S_R	n
1965										0.58	0.34	8
1966				1.26	0.39	9	163.7	104.6	6	0.57	0.10	9
1967	310.1	282	3	1.59	0.39	10	172.1	110.0	7	0.60	0.11	10
1968				1.70	0.40	11	176.8	102.0	8	0.61	0.15	11
1969				1.68	0.37	12	176.8	102.0	8	0.63	0.16	12
1970	329.2	218	4	1.75	0.41	13	176.3	95.6	9	0.66	0.15	13
1971				1.74	0.38	14	176.3	95.6	9	0.64	0.14	14
1972				1.76	0.36	15	191.9	94.8	10	0.64	0.13	15
1973				1.73	0.34	16	201.4	88.7	11	0.65	0.12	16
1974				1.73	0.33	17	207.7	88.1	12	0.67	0.12	17
1975				1.65	0.31	18	191.8	76.8	13	0.63	0.12	18
1976				1.65	0.29	19	196.8	70.8	14			
1977	596.0	247	5									

Where: S_R = Standard deviation of the ratio

n = Sample size on which ratio is based

Table 32. Age class 4(2) as estimated from the ratio of means 4(2)/3(2).

Brood year	Estimated 4(2)	95% confidence limits Upper limit	Lower limit	Standard deviation
1967	0			
1968	0			
1969	0			
1970	38,454	362,665	0	101,889
1971	0			
1972	0			
1973	0			
1974	0			
1975	0			
1976	166,675	705,052	0	193,940
1977	1,041,264	2,569,168	0	594,284

Table 33. Age class 5(2) as estimated from the ratio of means 5(2)/4(2).

Brood year	Estimated 5(2)	95% confidence limits Upper limit	Lower limit	Standard deviation
1966	293,809	745,788	0	199,814
1967	88,685	312,815	0	100,597
1968	64,249	251,288	0	84,979
1969	20,935	121,533	0	46,167
1970	102,600	343,421	0	111,491
1971	76,882	275,415	0	92,556
1972	96,766	310,186	0	100,150
1973	127,826	368,054	0	113,315
1974	224,197	555,899	0	147,726
1975	238,627	564,154	0	154,939
1976	962,941	1,663,717	262,164	334,819

Table 34. Age class 5(3) as estimated from the ratio of means 5(3)/4(3).

Brood year	Estimated 5(3)	95% confidence limits Upper limit	Lower limit	Standard deviation
1966	39,952	1,042,180	0	423,775
1967	313,089	3,157,473	0	1,233,471
1968	0			
1969	138,754	1,848,337	0	755,784
1970	0			
1971	312,714	2,716,084	0	1,078,712
1972	797,391	4,502,550	0	1,683,398
1973	85,372	1,184,298	0	504,326
1974	3,668,807	11,423,918	0	3,590,329
1975	1,772,314	6,861,736	0	2,372,691
1976	11,729,377	26,588,697	0	6,972,933

Table 35. Age class 6(3) as estimated from the ratio of means 6(3)/5(3).

Brood year	Estimated 6(3)	95% confidence limits Upper limit	Lower limit	Standard deviation
1965	1,127,885	2,499,074	0	579,784
1966	341,758	1,019,621	0	293,956
1967	380,362	1,112,576	0	323,702
1968	72,825	736,044	0	198,371
1969	679,325	2,056,803	0	625,841
1970	518,215	1,643,170	0	516,271
1971	908,961	2,406,178	0	693,156
1972	965,474	2,439,290	0	687,094
1973	369,019	1,215,109	0	397,039
1974	1,467,254	3,201,995	0	818,274
1975	1,430,979	3,215,549	0	845,768

symmetrical,

$$\hat{Y}_i \pm t_{(n-1)} [V(\hat{Y}_i)]^{\frac{1}{2}}$$

and based on a t-distribution (t_{n-1}) with n-1 degrees of freedom. Again, a lower limit of zero has been defined for the confidence interval.

The predicted returns based on the two ratios were also compared in relation to the observed return (Tables 36-39). The predictions based on the ratio of the means are consistently smaller than the mean ratio estimates. In addition, the percent error between observed and predicted return is less for the ratio of means estimators in all major age classes.

Lastly, there exists an additional estimation procedure used in forecasting the 6(3) age class return of a given brood year. Like the spawner-recruit relationship, it assumes a constant proportionment of total return into the major age classes. Then, instead of predicting total return from spawning escapement, it predicts total return from the return to date (RTD) of three, four, and five-year-olds, where the proportion

$$P = \frac{RTD}{RTD + 6(3)}$$

is based on historical data and total predicted return of brood year (i) becomes

$$(\text{Total Return})_i = \frac{RTD_i}{P}$$

The difference between the total and the return to date becomes an

Table 36. Comparison of estimates of age class 4(2) from the mean ratio and the ratio of means 4(2)/3(2).

Brood year	Observed 4(2)	Estimated 4(2) Mean ratio	% error	Ratio of means	% error
1967					
1968					
1969					
1970	58,537	100,393	71.5	38,454	-34.3
1971					
1972					
1973					
1974					
1975					
1976	582,296	366,896	-37.0	166,675	-71.4
1977	793,985	1,415,467	78.3	1,041,264	31.1
			Σ = 112.8		Σ = -74.6
			Σ |error| = 186.8		Σ |error| = 136.8

Table 37. Comparison of estimates of age class 5(2) from the mean ratio and the ratio of means 5(2)/4(2).

Brood year	Observed 5(2)	Estimated 5(2) Mean ratio	% error	Ratio of means	% error
1966	664,641	539,923	-18.8	293,809	-55.8
1967	220,628	132,269	-40.1	88,685	-59.8
1968	41,940	94,976	126.5	64,249	53.2
1969	105,183	28,830	-71.6	20,935	-80.1
1970	86,167	167,767	94.7	102,600	19.1
1971	106,255	122,216	15.0	76,882	-27.6
1972	58,338	150,585	158.1	96,766	65.9
1973	127,677	194,352	52.2	127,826	0.1
1974	97,738	334,046	241.8	224,197	129.4
1975	236,559	357,891	51.3	238,627	0.9
1976	775,728	1,419,638	83.0	962,941	24.1
			Σ = 692.1		Σ = 69.4
			Σ |error| = 953.1		Σ |error| = 516.0

Table 38. Comparison of estimates of age class 5(3) from the mean ratio and the ratio of means 5(3)/4(3).

Brood year	Observed 5(3)	Estimated 5(3) Mean ratio	% error	Ratio of means	% error
1966	604,374	72,631	-88.0	39,952	-93.4
1967	633,550	1,107,771	75.9	313,089	-50.6
1968	118,779	0			
1969	1,079,491	452,482	-58.1	138,754	-87.1
1970	787,985	0			
1971	1,430,466	1,179,994	-17.5	312,714	-78.1
1972	1,516,911	2,823,170	86.5	797,391	-47.4
1973	568,833	275,901	-51.5	85,372	-85.0
1974	2,178,973	12,511,766	474.2	3,668,807	68.4
1975	2,281,263	5,850,966	156.5	1,772,314	-22.3
1976	2,886,554	37,777,000	1,209.0	11,729,377	306.3
		Σ =	1,787.3	Σ =	-89.2
		Σ |error| =	2,217.2	Σ |error| =	838.6

Table 39. Comparison of estimates of age class 6(3) from the mean ratio and the ratio of means 6(3)/5(3).

Brood year	Observed 6(3)	Estimated 6(3) Mean ratio	% error	Ratio of means	% error
1965	911,224	1,373,919	50.8	1,127,885	23.8
1966	797,755	410,974	-48.5	341,758	-57.2
1967	555,019	468,827	-15.5	380,362	-31.5
1968	296,573	90,272	-69.6	72,825	-75.4
1969	1,101,694	971,542	-11.8	679,325	-38.3
1970	169,666	717,066	322.6	518,215	205.4
1971	926,825	1,230,201	32.7	908,961	-1.9
1972	1,193,895	1,289,374	8.0	965,474	-19.1
1973	842,214	477,847	-43.3	369,019	-56.2
1974	481,564	1,917,496	298.2	1,467,254	204.7
1975	776,539	1,916,260	146.8	1,430,979	84.3
		Σ =	670.4	Σ =	238.6
		Σ |error| =	1,047.8	Σ |error| =	797.8

estimate of the 6(3) age class:

$$6(3)_i = \left(\frac{RTD_i}{P}\right) - RTD_i$$

Upon algebraic manipulation (see Appendix E), an equivalent formulation can be derived which follows the ratio model.

$$6(3)_i = RTD_i \ (R)$$

where

R = ratio relating 6(3) return to RTD

This estimator can then be evaluated in terms of a ratio estimate based on the model first presented in this section. The mean ratio then used (Table 40) to estimate return from brood year (i) is based on historical data from 1957 to i-1 as

$$\hat{R}_i = \frac{1}{n} \sum_{j=1957}^{(i-1)} \frac{6(3)_j}{RTD_j}$$

where

n = the number of observations from brood year 1957 to i-1

Similarly, the ratio of the means (Table 40) used to predict 6(3) return from brood year (i) is derived as

$$R_i = \frac{\sum_{j=1957}^{i-1} 6(3)_j}{\sum_{j=1957}^{i-1} RTD_j}$$

Table 40. The mean ratio and the ratio of the means for the ratio of 6(3) return to total return to date (RTD).

Brood year	Mean ratio			The ratio of the means		
	Ratio	S_R	n	Ratio	S_R	n
1965	0.55	0.10	8	0.49	0.08	8
1966	0.54	0.09	9	0.48	0.07	9
1967	0.54	0.08	10	0.49	0.07	10
1968	0.54	0.08	11	0.49	0.06	11
1969	0.62	0.10	12	0.51	0.07	12
1970	0.64	0.10	13	0.53	0.07	13
1971	0.61	0.10	14	0.52	0.07	14
1972	0.61	0.09	15	0.52	0.06	15
1973	0.62	0.09	16	0.54	0.08	16
1974	0.64	0.08	17	0.51	0.08	17
1975	0.62	0.08	18	0.49	0.08	18

Where: S_R = Standard deviation of the ratio

n = Sample size on which the ratio is based

The variance for the ratio estimates ($\hat{R}$) and the errors (σ^2) then follow from the earlier discussions of ratio estimates. Hindcasting for the brood years 1965 through 1975 using the mean ratio (Table 41) produced larger estimates with less precision than those based on the ratio of means (Table 42). The predicted returns based on ratio estimates were then compared with the observed return (Table 43) and the percent error in prediction was derived. The smallest cumulative error of the RTD estimation process is the forecast based on the ratio of means (Table 43). Overall, the prediction of 6(3) based on return to date was more precise when viewing the cumulative error or the magnitude of the variance as that based on the 6(3) to 5(3) ratios.

In summary, two methods of estimating the ratio used in predicting (N + 1)-ocean return have been presented. They differ in the assumption of the error variance and the resulting formulation of an estimate of the ratio and its variances. The mean ratio gave equal weight to every observation pair and in hindcasting was consistently larger than the other method. In contrast, the ratio of the means gave more weight to years of larger runs. There is also a smaller variance associated with this procedure. Lastly, the ratio of the means is a preferred method when the two components of the ratio are highly correlated.

The choice between these two methods should be based on which set of assumptions is more appropriate. In viewing the N-ocean versus (N + 1)-ocean return by brood year (Fig. 4), there exists substantial scatter in the data points, and they would vary greatly about any line fit through the origin with these data. One cannot, from this, determine whether the scatter in the data increases with N-ocean return or by its

Table 41. Age class 6(3) estimated from average ratio of 6(3) to total return to date (RTD).

Brood year	Estimated 6(3)	95% confidence limits Upper limit	Lower limit	Standard deviation
1965	1,171,967	2,743,952	0	664,687
1966	804,591	1,818,832	0	439,827
1967	488,533	1,055,786	0	250,775
1968	107,423	222,507	0	51,654
1969	743,529	1,741,049	253,991	453,212
1970	600,452	1,355,169	9	346,360
1971	966,543	2,258,830	0	598,281
1972	996,065	2,271,281	0	594,506
1973	475,174	1,053,085	0	271,192
1974	1,562,285	3,413,829	0	873,369
1975	1,657,265	3,718,157	0	976,726

Table 42. Age class 6(3) estimated from the ratio of mean 6(3) return to mean return to date (RTD).

Brood year	Estimated 6(3)	95% confidence limits upper limit	lower limit	Standard deviation
1965	1,054,378	2,158,324	0	466,785
1966	727,484	1,552,716	0	357,863
1967	446,041	1,029,751	0	258,050
1968	98,156	349,294	0	112,719
1969	606,316	1,290,360	0	310,788
1970	497,286	1,126,297	0	288,669
1971	816,887	1,650,219	0	385,802
1972	852,653	1,664,494	40,812	378,481
1973	401,811	1,139,292	0	346,073
1974	1,312,958	2,683,781	0	646,615
1975	1,355,969	2,777,105	0	673,524

Table 43. Comparison of estimates of age class 6(3) from the mean ratio and the ratio of the means 6(3) to return to date.

Brood year	Observed 6(3)	Mean ratio	% error	Ratio of means	% error
1965	911,224	1,171,967	28.6	1,054,378	15.7
1966	797,755	804,591	0.9	727,484	- 8.8
1967	555,019	488,533	- 12.0	446,041	- 19.6
1968	296,573	107,423	- 63.8	98,156	- 66.9
1969	1,101,694	743,529	- 32.5	606,316	- 45.0
1970	169,666	600,452	253.9	497,286	193.1
1971	926,825	966,543	4.3	816,887	- 11.9
1972	1,193,895	996,065	- 16.6	852,653	- 28.6
1973	842,214	475,174	- 43.6	401,811	- 52.3
1974	481,564	1,562,285	224.4	1,312,958	172.6
1975	776,539	1,657,265	113,4	1,355,969	74.6
		Σ =	457.0	Σ =	22.9
		Σ |error| =	794.0	Σ |error| =	689.1

Table 44. The correlation coefficient matrix for N-ocean and (N +1) - ocean return from brood years 1957 to 1977.

(N+1)-ocean return	N -ocean return 3(2)	4(2)	4(3)	5(3)
4(2)	.85			
5(2)		.65		
5(3)			.67	
6(3)				.76

square needed to differentiate between the error assumptions. One can comment on the strong correlation between the (N + 1) and N-ocean return (Table 44), which would indicate that the mean ratio is quite well-suited for this application. In addition, the decrease in percent error and absolute percent error with the use of the ratio of means (Tables 36-39, 43) would support its preferred use. Lastly, it did have the smaller variance for the estimate of (N + 1)-ocean return.

COMPARISON OF ESTIMATION PROCEDURES USED IN FORECASTING

Estimation procedures widely used in the forecasting of sockeye salmon returns were discussed in the previous sections as to their models and underlying assumptions. Variances and confidence intervals for the predictions were developed, and the estimators were evaluated by their hindcasting reliability for the brood years 1965 through 1977. The spawner-recruit model either estimated total return as recruits with age class apportionment necessary, or it estimated age-specific recruits. Only the latter resulted in a comprehensive variance estimate and will be used unless otherwise specified. The second type of estimator was based on a relationship between two age classes which share the same freshwater history but differ by one year of marine residence. The relation was developed in two ways. Both assumed a linear model where the $(N+1)$-ocean return is estimated as a linear function of the N-ocean age return the year before. A regression estimator accommodated a non-zero intercept; constant error variance and model coefficients were estimated by a least squares procedure. In contrast, the ratio method assumed a zero intercept in the linear model, and a ratio was estimated to relate $(N+1)$ to N-ocean return. Two formulations for estimating the ratio were presented based on a different assumption for the error variance, resulting in different variance formulae for the predicted values. The average ratio and the ratio of the means are both discussed in this section.

Assumptions common to all procedures involved a random error with a mean of zero and which will therefore cancel through time. Also, they assume that no covariance between terms exists, and it may be questioned how this suits the biological system involved, which is subject to trends in environmental conditions and consists of overlapping age classes of a fish stock which share a genetic makeup. In addition, the least squares procedure used to estimate the model's coefficients requires that the observed returns be measured without error. In most salmon management programs, sampling is conducted to apportion a given year's return into age classes, and the sample should be of a size to ensure some agreement with this assumption. On the question of zero mean error and no covariance between observations, a close evaluation of the residual resulting when the model is fit to historic data could detect any appreciable deviation which would suggest that alternative models should be sought.

In contrast, the assumption made on the error variance ($V(e_i)$) varies between models. The important case involves the (N + 1) to N-ocean relationship for Egegik sockeye salmon. Here the regression assumes a constant error variance (σ^2) where the ratio method assumes that it varies with the return of N-ocean ($X\sigma^2$) for the ratio of means model and its square ($X^2 \sigma^2$) for the mean ratio model. If the true variance structure were known, those models in violation could be eliminated. Yet as discussed in the section on ratio estimation, the data observed thus far give no strong indication of which assumption is more appropriate, and the choice returns to other considerations such as ability to forecast.

With all the estimators, there was substantial scatter about their underyling model when fitted to historic data. This was accounted for in defining the random error in the linear models and in the statement by Ricker (1954) that the spawner-recruit model was based on average conditions. Therefore, given that these models are biologically plausible, they are describing the system in too simple terms. They may each describe important factors influencing the magnitude of an age class return, but substantial variability in the system is left unexplained. The next step needed to increase precision (as defined by the decrease in the variance) is to incorporate additional factors into the models which are found to influence the level of return of adult sockeye salmon.

In summary, assumptions on which each estimator is based are often not verifiable with available data being limited. Initially residuals in fitting the models to historic data were evaluated to determine if they trend, indicating covariance, or do not cancel, indicating a non-zero mean error. No strong evidence against the support of a particular model's assumption was present. Therefore at this time no estimator can be discarded on the basis of its assumptions.

Historically, the ultimate outcome of this procedure is a single forecast point for a season's return of adult sockeye salmon. Different approaches can be taken in the use of the hindcasting information developed thus far. One approach would be to use all prediction methods available from previous sections to construct estimates for each year being hindcast. The previous estimates would be combined by some averaging technique. Yet there is a high degree of similarity in the assumed model and input data for the linear (N + 1) to N-ocean estimators, especially with the two ratio methods. In result, only one estimator from each of the three general methods would be used as one from the spawner-recruit model, one from the regression, and one from the ratio model. Another approach would be to use only that estimator which was most reliable. One would need to compare the performance of the different prediction methods in terms of their ability to hindcast returns and use only that method with the greatest reliability. Indicators of reliability would be the percent error in observed from predicted or the residual variance of the forecast. Yet this could result in different estimators being used each year, and, in effect, information is being ignored.

One objective of this research was to develop the best approach for deriving a single point forecast with an estimated variance based on commonly used procedures and incorporating available data. Thus far, several estimators have been presented along with their variance properties. It now becomes necessary to develop criteria to be used in deciding which of and how these estimators will be incorporated into a final point forecast. Previously cumulative percent error and cumulative absolute error were derived for the hindcasts of the return from brood years 1967-1977 based on each estimator. These will be valuable in comparing estimators, yet they do not reflect the magnitude of the error, giving equal weight to small deviations of small returns and large deviations of large returns. Therefore, a measure of the reliability of an estimate was sought which was independent of the size of the estimate but which incorporated the absolute magnitude of the deviation of predicted from observed. One such measure of the reliability of an estimation process involves the measure of the residual variance about the observed. This would be defined for estimator j as the residual variance of forecast (RVF_j)

$$RVF_j = \frac{1}{n-1} \sum_{i=k}^{n+k-1} (\hat{R}_{ij} - R_i)^2 \qquad [9]$$

where

R_i = observed return from brood year i

$\hat{R}_{ij}$ = predicted return from brood year i using estimation procedure j

n = the number of years for which the return can be hindcast or for which historic forecast data are available

In summary, criteria evaluating the estimators would be developed to determine the approach taken in deriving a final point forecast. The estimators would be evaluated by developed criteria involving previous hindcasting done for brood years 1965-1977.

As previously stated, two criteria of estimator performance involve the percent error between the prediction of an age class return and that observed. Values of the cumulative percent errors and the cumulative absolute errors were presented within the discussion of each estimator and its hindcasting ability. The cumulative errors from the spawner-recruit model with age-specific recruits can be found in Tables 11 through 14. The cumulative errors from the regression estimation of the four major age classes are in Tables 16-19. The two ratio estimators are compared by percent error in Tables 36-39 with the additional ratio for age class 6(3) involving return to date presented in Table 43. The remaining criterion of performance involves the residual variance of forecast (RVF). Like percent error, it measures the deviation between predicted and observed. In contrast, the RVF incorporates the magnitude of the deviation and is thus sensitive to the dominance by a few large departures.

The RVF was calculated for each estimator and age class (Tables 45-48) based on the hindcasting for brood years 1965-1977. Note that few estimators have a constant RVF through time, while those of age class 6(3) appear the most stable. It is evident that, beginning with brood year 1975, the system is changing, as seen in the more than doubling of the RVF in age classes 4(2) and 5(3). The final RVF for the hindcasting is represented by that of the last brood year estimate of each age class. The estimator with the minimum final RVF differed little between age classes. The regression estimator had a minimum RVF for age classes 4(2), 5(3), and 6(3), while the ratio of the means minimized the RVF's for the 5(2) age class. This result must be tempered with knowledge of the large deviation present in just the last two brood years, it being most evident in the ratio methods.

The criteria of performance for the estimators based on hindcasting became

1. the estimator which minimized the cumulative percent error,
2. the estimator which minimized the cumulative absolute value of the errors,
3. the estimator which maximized the number of years with lowest absolute error,
4. the estimator which maximized the number of years with minimum RVF.

Ideally, the best estimator would be one that fulfilled all the above criteria. Yet only for age class 4(2) was the choice clear (Table 49),

Table 45. Comparison of the residual variance in forecasting (RVF) for the various estimators of age class 4(2).

Last brood year in RVF	Residual variance in forecasting ($x10^9$) for: Spawner-recruit	Regression	Ratio: Average	Ratio: Of the means
1967	0.14	1.33		
1968	0.17	4.29		
1969	1.63	5.33		
1970	1.12	3.96	1.75	0.40
1971	0.84	3.19		
1972	0.73	2.61		
1973	0.75	2.18		
1974	0.95	2.30		
1975	1.43	2.63		
1976	33.70	28.38	48.15	17.31
1977	84.96	29.58	21.72	11.71

Table 46. Comparison of the residual variance in forecasting (RVF) for the various estimators of age class 5(2).

Last brood year in RVF	Residual variance in forecasting ($x10^{10}$) for: Spawner-recruit	Regression	Ratio: Average	Ratio: Of the means
1966	33.52	21.56	1.58	13.75
1967	34.78	22.46	2.36	15.49
1968	17.49	11.53	1.32	7.77
1969	11.72	7.70	1.07	5.42
1970	8.85	5.83	0.97	4.07
1971	7.08	4.66	0.80	3.27
1972	5.93	3.96	0.79	2.75
1973	5.14	3.96	0.74	2.36
1974	4.52	3.08	1.35	2.26
1975	4.14	2.75	1.36	2.01
1976	4.65	27.32	5.37	2.16

Table 47. Comparison of the residual variance in forecasting (RVF) for the various estimators of age class 5(3).

Last brood year in RVF	Residual variance in forecasting (x10^{11}) for: Spawner-recruit	Regression	Ratio: Average	Ratio: Of the means
1966	0.12	1.59	2.83	3.19
1967	0.12	3.04	5.08	4.21
1968	1.60	4.04	5.08	4.21
1969	1.17	2.82	4.50	6.53
1970	0.88	2.13	4.50	6.53
1971	2.30	2.11	3.21	8.52
1972	3.48	1.94	6.68	7.68
1973	2.98	1.91	5.51	6.61
1974	3.15	1.69	18.25	9.21
1975	3.75	2.14	17.47	8.27
1976	8.58	6.31	167.50	104.98

Table 48. Comparison of the residual variance if forecasting (RVF) for the various estimators of age class 6(3).

Last brood year in RVF	Residual variance in forecasting (x10^{11}) for: Spawner-recruit	Regression	Ratio: Average	Ratio: Of the means	RTD: Average ratio	RTD: Ratio of the means
1965	0.13	0.30	2.14	0.47	0.68	0.20
1966	1.47	1.98	3.64	2.55	0.68	0.25
1967	0.81	1.03	1.86	1.43	0.36	0.19
1968	0.64	0.70	1.38	1.12	0.36	0.26
1969	1.09	0.95	1.08	1.29	0.59	0.81
1970	1.22	1.11	1.46	1.27	0.84	0.86
1971	1.48	0.93	1.37	1.06	0.71	0.74
1972	2.10	0.91	1.19	0.98	0.66	0.80
1973	2.14	0.98	1.21	1.14	0.75	0.94
1974	2.02	1.54	3.36	2.09	1.96	1.60
1975	1.83	1.57	4.33	2.31	2.54	1.78

Table 49. Comparison of the performance of the different estimators for age class return of the Egegik sockeye salmon run.

Estimator of age class return which --	Age			
	4(2)	5(2)	5(3)	6(3)
Minimized cumulative percent error	S-R	S-R	ROM	RTD-ROM
Minimized cumulative absolute error	S-R	ROM	S-R	S-R
Maximized the number of years with lowest error	S-R	ROM	S-R	S-R
Maximized the number of years with minimum RVF	S-R	MR	Regr	M-RTD

Where the estimators are:

- S-R = Spawner-recruit model with age specific recruits.
- Regr = Regression of (N+1) - ocean on N-ocean returns.
- MR = Mean ratio of (N+1) to N-ocean return.
- ROM = Ratio of the mean (N+1) to N-ocean return.
- M-RTD = Mean ratio of age class 6(3) to return to date.
- RTD-ROM = Ratio of the mean 6(3) return over the mean return to date.

as the spawner-recruit estimator fulfilled all criteria. For the other age classes the choice was less apparent, though the spawner-recruit model did dominate in each. Both the ratio and regression methods fulfilled some criteria. It is apparent from Table 49 that the spawner-recruit model for estimating age class return compares well with the other methods and should be included in the final estimate. Following it, different linear estimators fulfilled different criteria, and no one estimator dominated.

In that no one procedure fulfilled all criteria and some estimators failed to fulfill any, a combination of the two approaches for deriving a final point forecast is proposed. Again the objective is a yearly forecast which best predicts the upcoming year's return and has defined variance properties such that a statement of range can be made with some probability. In result, methods which consistently perform poorly in terms of a large percent error and a large RVF in hindcasting will not be used. The remaining estimates will be incorporated into the final forecast as a weighted average. If the variance of an estimate was independent of that estimate and the estimation methods were independent, weighting by the reciprocal of the variance would produce an estimate with minimum variance. Yet the variances derived in all methods incorporated a component of the estimate itself. In both the spawner-recruit and (N + 1)-ocean regressions, the variance of the estimate was a function of the level of the independent variable (see equations 3 and 7) from which the estimate was derived. A similar dependence was observed for the variance of the ratio estimates of (N + 1)-ocean return. The error variance (see pages 46 and 52) was assumed proportional to the level of the N-ocean return which is used in the estimate itself. In result, only an independent measure of the reliability of an estimate could be used as a weight.

A weighted average would be based on those estimates which best fulfilled the performance standards of Table 49. The weight for the averaging would then become the reciprocal of the RVF which is an independent measure of the variability of that forecasting method. The weighted average for age class (R) of brood year (i) was calculated as

$$R = \sum_{j=1}^{h} (R_{ij}/RVF_j) / \sum_{j=1}^{h} (RVF)_j^{-1} \qquad [10]$$

To be consistent with the forecasting procedure, the RVF used to average for brood year i was based on the forecasts developed over brood year 1965 through (i-1). Therefore the first year hindcast for each age class is an unweighted average where the RVF is taken to equal one.

Taking an unweighted average of the three widely used estimators is a commonly used procedure, and an estimate was developed for comparison with the weighted average estimate. The averages included estimates from the spawner-recruit model with apportionment, the regression estimator, and the average ratio estimate. The weighted average was also based on three estimates where the age-specific spawner-recruit estimator was used. Tables 50 through 53 compare these two approaches for the hindcasting of brood years 1965 through 1977. In all cases, the cumulative errors of the weighted average were similar to, or markedly less than, those of the unweighted average. A visible improvement is demonstrated in the weighting for point estimates of age classes 4(2) (Table 50) and 5(3) (Table 52). Yet the averages involving these particular estimators were

Table 50. Comparison of average estimates of age class 4(2), averaged over the spawner-recruit, regression, and mean ratio estimates.

Brood year	Observed 4(2)	Average est.	% error	Weighted average est.	% error
1967	55,739	59,737	7	67,946	22
1968	37,734	68,538	82	47,863	27
1969	12,429	84,228	578	68,894	454
1970	58,537	87,359	49	59,573	2
1971	44,185	63,581	44	46,450	5
1972	54,918	60,374	10	45,578	- 17
1973	73,702	58,371	- 21	50,726	- 31
1974	129,275	88,830	- 46	80,773	- 38
1975	144,195	81,618	- 43	74,633	- 48
1976	582,296	171,665	- 71	167,427	- 71
1977	793,985	805,845	2	771,363	- 3
		Σ =	591	Σ =	302
		Σ |error| =	953	Σ |error| =	718

Table 51. Comparison of average estimates of age class 5(2), averaged over the spawner-recruit, regression, and mean ratio estimates.

Brood year	Observed 5(2)	Average est.	% error	Weighted average est.	% error
1966	665,541	279,924	- 58	275,926	- 58
1967	220,628	112,324	- 49	130,830	- 41
1968	41,940	111,700	166	96,610	130
1969	105,183	97,720	- 7	42,890	- 59
1970	86,167	151,946	76	161,436	87
1971	106,255	117,060	10	119,972	13
1972	58,338	125,275	114	142,615	144
1973	127,677	142,238	11	173,068	36
1974	97,738	248,080	154	291,045	198
1975	236,559	250,337	6	279,217	18
1976	775,728	713,461	- 8	962,431	24
		Σ =	415	Σ =	492
		Σ |error| =	659	Σ |error| =	808

Table 52. Comparison of average estimates of age class 5(3), averaged over the spawner-recruit, regression, and mean ratio estimates.

Brood year	Observed 5(3)	Average est.	% error	Weighted average est.	% error
1966	604,374	639,066	6	595,705	- 1
1967	633,550	960,870	52	660,105	4
1968	118,779	812,512	585	747,259	529
1969	1,079,491	781,479	- 28	814,430	- 25
1970	787,985	879,650	12	809,339	3
1971	1,430,466	946,299	- 34	727,293	- 49
1972	1,516,911	1,546,289	2	1,235,848	- 19
1973	568,833	624,031	10	741,133	30
1974	2,178,973	8,040,442	269	3,711,267	70
1975	2,281,263	2,895,544	27	1,492,803	- 35
1976	2,886,554	14,502,174	402	3,657,688	27
		Σ =	1303	Σ =	534
		Σ |error| =	1427	Σ |error| =	792

Table 53. Comparison of average estimates of age class 6(3), averaged over the spawner-recruit, regression, and mean ratio estimates.

Brood year	Observed 6(3)	Average est.	% error	Weighted average est.	% error
1965	911,224	1,227,771	35	1,161,943	28
1966	797,755	546,156	- 32	463,362	- 42
1967	555,019	485,923	- 13	468,958	- 16
1968	296,573	233,029	- 21	208,228	- 30
1969	1,101,694	788,841	- 28	732,613	- 34
1970	169,666	653,959	285	618,254	264
1971	926,825	894,858	- 4	862,288	- 7
1972	1,193,895	931,359	- 22	927,529	- 22
1973	842,214	467,343	- 45	481,154	- 43
1974	481,564	1,469,094	205	1,461,768	204
1975	776,539	1,443,364	86	1,232,070	59
		Σ =	446	Σ =	361
		Σ |error| =	776	Σ |error| =	749

derived for comparison with commonly used procedures and may not represent the best choice resulting from this investigation. In summary, weighting appears to provide a similar or better estimate than the unweighted average. Furthermore, the same weighting scheme can be used to derive a variance for the weighted average which will be smaller than if the variances were averaged with no weighting.

In the first derivation of the linear estimators which describe the (N + 1) to N-ocean relationship, the similarity between the regression and ratio estimators was apparent. It therefore may not be appropriate to average one spawner-recruit estimate with two (N + 1)-ocean estimators as previously done. In the case of the unweighted average, weight is given to the (N + 1)-ocean type estimator at two to one over the spawner-recruit. The ratio is less clear in the weighted average, but still additional influence is given the (N + 1)-ocean estimators. Most importantly, the variance of a weighted average incorporates no covariance terms which result if estimators are not independent. The similarity between the regression and mean ratio estimator becomes obvious when viewing the estimated correlation between procedures (Tables 54-57). Here the correlation coefficient between the regression and ratio estimates ranges from 0.97 in age class 4(2) to 0.999 with age class 5(3). In addition, the low correlation between the spawner-recruit estimator and the (N + 1)-ocean estimators is noteworthy, being under 0.2 for age classes 4(2), 5(2), and 5(3).

It follows from the correlation matrices that a better method for deriving point estimates would result from the use of a weighted average involving a spawner-recruit estimate and one additional (N + 1)-ocean

Table 54. Correlation coefficient matrix for estimates of age class 4(2).

	Observed	Spawner-recruit estimate	Regression estimate	Ratio estimate
Observed				
S-R	-0.144			
Regr.	0.796	-0.186		
Ratio	0.909	-0.201	0.970	

Table 55. Correlation coefficient matrix for estimates of age class 5(2).

	Observed	Spawner-recruit estimate	Regression estimate	Ratio estimate
Observed				
S-R	-0.310			
Regr.	0.805	-0.114		
Ratio	0.876	-0.135	0 .988	

Table 56. Correlation coefficient matrix for estimates of age class 5(3).

	Observed	Spawner-recruit estimate	Regression estimate	Ratio estimate
Observed				
S-R	0.425			
Regr.	0.780	0.077		
Ratio	0.793	0.112	0.999	

Table 57. Correlation coefficient matrix for estimates of age class 6(3).

	Observed	Spawner-recruit estimate	Regression estimate	Ratio estimate	RTD estimate
Observed					
S-R	0.045				
Regr.	0.364	0.748			
Ratio	0.368	0.698	0.991		
RTD	0.363	0.698	0.942	0.951	

estimate. The choice of the linear model estimator should be allowed to vary by age class and could vary by year forecast. The objective would be to use the method which had performed best in the past as determined by minimizing the RVF or cumulative errors for the hindcasting of the Egegik sockeye salmon run. This would result in final point estimates for age classes 4(2) and 5(3) being the weighted average of the spawner-recruit estimate and the regression estimate (Tables 58 and 60). Estimates of age class 5(2) would result from the weighted average of the spawner-recruit estimate and ratio of the means estimate (Table 59) and lastly age class 6(3) uses the weighted average of the spawner-recruit and ratio of the means of 6(3) to return to date estimates (Table 61). Because of the low correlation between the spawner-recruit and linear estimators, covariance terms have been neglected and the variance for the weighted average (Tables 58-61) of return at age (R) from brood year i were estimated as

$$V(\hat{R}_i) = \frac{\sum_{j=1}^{2} V(\hat{R}_{ij}) / (RVF_j)^2}{\left[\sum_{j=1}^{2} (RVF_j)^{-1} \right]^2}$$

The variance of the predicted return of age class by brood year and estimator are presented in Tables 62 to 65.

The final point estimates for year of return for 1971 through 1981 became the sum of the appropriate returns by brood year of four, five, and six-year-olds. Again, a comparison was made of the three approaches (Table 66), and the weighted average incorporating two estimates minimized

Table 58. Estimates of age class 4(2) derived as a weighted average of the spawner-recruit and regression estimates.

Brood year	Observed 4(2)	Estimated 4(2)	% Error	Variance ($x10^9$)
1967	55,739	67,946	22	7.97
1968	37,734	47,863	27	15.72
1969	12,429	68,894	454	12.78
1970	58,537	59,213	1	8.95
1971	44,185	46,450	5	7.23
1972	54,918	45,578	- 17	5.95
1973	73,702	50,726	- 31	5.81
1974	129,275	80,773	- 38	4.08
1975	144,195	74,633	- 48	3.72
1976	582,296	61,848	- 89	3.70
1977	793,985	565,256	- 29	175.82
		Σ =	257	
		Σ |error| =	761	

Table 59. Estimates of age class 5(2) derived as a weighted average of the spawner-recruit and ratio of the means estimates.

Brood year	Observed 5(2)	Estimated 5(2)	% Error	Variance ($x10^{10}$)
1966	665,541	279,924	- 58	10.04
1967	220,628	94,379	- 57	0.77
1968	41,940	70,728	69	0.62
1969	105,183	60,244	- 43	0.36
1970	86,167	113,446	32	0.80
1971	106,255	86,292	- 19	0.59
1972	58,338	96,987	66	0.63
1973	127,677	108,129	- 15	0.74
1974	97,738	199,459	104	1.15
1975	236,559	203,767	- 14	1.20
1976	775,728	678,944	- 12	5.22
		Σ =	53	
		Σ |error| =	489	

Table 60. Estimates of age class 5(3) derived as a weighted average of the spawner-recruit and regression estimates.

Brood year	Observed 5(3)	Estimated 5(3)	% Error	Variance ($x10^{11}$)
1966	604,374	857,242	42	7.55
1967	633,550	643,902	2	7.16
1968	118,779	747,281	529	6.50
1969	1,079,491	896,186	- 17	8.21
1970	787,985	809,346	3	7.08
1971	1,430,466	665,132	- 54	6.12
1972	1,516,911	882,622	- 42	4.89
1973	568,833	827,758	46	4.84
1974	2,178,973	1,852,807	- 15	4.49
1975	2,281,263	1,466,541	- 36	4.17
1976	2,886,554	3,391,057	17	9.41

Σ = 475

Σ |error| = 803

Table 61. Estimates of age class 6(3) derived as a weighted average of the spawner-recruit and ratio of the means for 6(3) to return to date estimates.

Brood year	Observed 6(3)	Estimated 6(3)	% Error	Variance ($x10^{11}$)
1965	911,224	1,040,651	14	1.45
1966	797,755	548,791	- 31	1.43
1967	555,019	444,185	- 20	0.54
1968	296,573	168,816	- 43	0.18
1969	1,101,694	606,327	- 45	0.68
1970	169,666	536,053	216	0.73
1971	926,825	646,295	- 30	1.00
1972	1,193,895	711,073	- 40	0.95
1973	842,214	425,418	- 49	0.87
1974	481,564	1,160,554	141	2.30
1975	776,539	1,063,761	37	1.95

Σ = 150

Σ |error| = 666

Table 62. Comparison of variances for the different estimates of age class 4(2).

Brood year	SP-Recruit 4(2)	SP-Recruit Variance ($x10^9$)	Regression 4(2)	Regression Variance ($x10^9$)	Average ratio 4(2)	Average ratio Variance ($x10^9$)	Ratio of means 4(2)	Ratio of means Variance ($x10^9$)
1967	43,710	30.60	92,182	1.26				
1968	43,036	19.19	92,182	1.26				
1969	67,956	13.81	92,182	1.26				
1970	49,131	15.18	93,683	0.94	100,393	21.77	38,454	10.38
1971	38,725	11.80	73,895	1.16				
1972	38,113	9.42	73,895	1.16				
1973	44,236	7.96	73,895	1.16				
1974	83,156	7.23	73,895	1.16				
1975	74,937	7.75	73,895	1.16				
1976	42,067	8.48	98,299	1.11	366,896	23.57	166,675	37.61
1977	54,845	18.11	995,091	583.78	1,415,467	215.16	1,041,264	353.17

Table 63. Comparison of variances for the different estimates of age class 5(2).

Brood year	SP-recruit 5(2)	SP-recruit Variance ($x10^{10}$)	Regression 5(2)	Regression Variance ($x10^{10}$)	Average ratio 5(2)	Average ratio Variance ($x10^{10}$)	Ratio of means 5(2)	Ratio of means Variance ($x10^{10}$)
1966	86,601	0.82	201,255	0.48	539,923	24.82	293,809	3.99
1967	108,261	3.05	125,597	2.80	132,269	1.26	88,685	1.01
1968	85,274	2.91	119,524	2.59	94,976	0.55	64,249	0.72
1969	148,727	2.69	87,193	2.42	29,830	0.06	20,935	0.21
1970	136,900	2.20	132,338	2.10	167,767	2.19	102,600	1.24
1971	106,754	1.84	115,197	1.95	122,216	1.17	76,882	0.86
1972	97,464	1.60	124,633	1.78	150,585	1.68	96,766	1.00
1973	65,655	1.42	138,070	1.67	194,352	2.91	127,826	1.28
1974	145,582	1.31	190,694	1.57	334,046	8.46	224,197	2.18
1975	134,047	1.18	198,585	1.53	357,891	10.29	238,627	2.40
1976	93,995	1.29	616,254	3.29	1,419,638	159.35	962,941	11.21

Table 64. Comparison of variances for the different estimates of age class 5(3).

Brood year	SP-recruit 5(3)	SP-recruit Variance ($x10^{12}$)	Regression 5(3)	Regression Variance ($x10^{12}$)	Average ratio 5(3)	Average ratio Variance ($x10^{12}$)	Ratio of means 5(3)	Ratio of means Variance ($x10^{12}$)
1966	711,941	1.00	1,002,543	2.02	72,631	0.003	39,952	0.18
1967	615,879	0.82	1,015,200	1.53	1,107,771	2.67	313,089	1.52
1968	744,077	0.70	828,461	1.29				
1969	901,988	1.40	881,537	1.28	452,482	0.43	138,754	0.57
1970	788,449	1.23	859,712	1.09				
1971	536,064	1.04	977,536	1.07	1,179,994	2.13	312,714	1.16
1972	550,641	1.03	1,187,179	0.93	2,823,170	10.32	797,391	2.83
1973	551,434	1.04	981,801	0.85	275,901	0.10	85,372	2.54
1974	1,519,269	0.92	2,066,584	0.83	12,511,766	166.60	3,668,807	12.89
1975	1,358,900	1.06	1,524,291	0.68	5,850,966	43.93	1,772,314	5.63
1976	603,935	1.20	4,981,574	1.93	37,777,000	1,727.10	11,729,377	48.62

Table 65. Comparison of variances for the different estimates of age class 6(3).

Brood year	SP-recruit 6(3)	SP-recruit Variance ($x10^{11}$)	Regression 6(3)	Regression Variance ($x10^{11}$)	Average ratio 6(3)	Average ratio Variance ($x10^{11}$)	Ratio of means 6(3)	Ratio of means Variance ($x10^{11}$)	RTD mean ratio 6(3)	RTD mean ratio Variance ($x10^{11}$)	RTD ratio of means 6(3)	RTD ratio of means Variance ($x10^{11}$)
1965	1,026,925	3.62	1,084,985	1.44	1,373,919	10.53	1,127,885	3.36	1,171,967	4.42	1,054,378	2.18
1966	432,641	3.35	388,746	1.28	410,974	0.90	341,758	8.64	804,591	1.93	727,484	1.28
1967	433,270	2.97	460,354	1.30	468,827	1.06	380,362	1.05	488,533	0.63	446,041	0.66
1968	470,049	2.49	226,277	1.22	90,272	0.03	72,825	0.39	107,423	0.27	98,156	0.13
1969	606,354	2.29	690,719	1.00	971,542	5.67	679,325	3.92	743,529	2.05	606,316	0.97
1970	588,220	2.50	585,564	1.07	717,066	2.74	518,215	2.67	600,452	1.20	497,286	0.83
1971	404,292	2.83	860,691	1.12	1,230,201	8.98	908,961	4.80	966,543	3.58	816,887	1.49
1972	427,914	2.86	907,080	1.03	1,289,374	9.43	965,474	4.72	996,065	3.53	852,653	1.43
1973	487,386	3.16	464,673	1.03	477,847	1.24	369,019	1.58	475,174	0.74	401,811	1.20
1974	813,231	3.04	1,257,520	1.10	1,917,496	18.10	1,467,254	6.70	1,562,285	7.63	1,312,958	4.18
1975	694,849	2.74	1,208,306	1.40	1,916,260	20.44	1,430,979	7.15	1,657,265	9.54	1,355,969	4.54

Table 66. Comparison of total run estimates of the Egegik sockeye salmon run from 1971-1981.

Year of return	Observed return	Unweighted average return	% error	Weighted average return: With two linear estimators	% error	With one linear estimator	% error
1971	1,941,000	2,206,498	14	2,101,520	8	2,245,763	16
1972	1,256,000	1,687,888	34	1,302,160	4	1,334,935	6
1973	546,000	1,494,363	174	1,381,721	153	1,331,088	144
1974	1,438,000	1,199,587	- 17	1,125,121	- 22	1,184,459	- 18
1975	2,113,000	1,884,018	- 11	1,749,838	- 17	1,575,569	- 25
1976	1,814,000	1,777,692	- 2	1,511,097	- 17	1,333,055	- 27
1977	2,438,000	2,624,793	8	2,291,477	- 6	1,676,630	- 31
1978	2,098,000	1,786,458	- 15	1,922,503	- 8	1,727,733	- 18
1979	3,286,000	8,837,483	169	4,558,099	39	2,552,317	- 22
1980	3,467,000	4,786,640	38	4,084,429	18	2,892,710	- 17
1981	5,176,000	17,464,844	237	6,623,552	28	5,699,018	10
		Σ =	629	Σ =	186	Σ =	18
		Σ |error| =	719	Σ |error| =	326	Σ |error| =	334

the cumulative error. The variance for the total return estimate from the two-estimator weighted average procedure becomes just the sum of the variance of each age class estimate (Table 67). Again, normal errors were assumed and a confidence interval based in the t-distribution was placed about the estimate. An 80 percent confidence level was thought sufficient, as it was desirable that the confidence interval be narrow enough to be useful rather than using the conventional 95 percent bounds which too often included zero. Only in the last few years has the confidence interval narrowed to a useful width (Table 67). Yet the confidence interval contained the observed return in all years hindcast, indicating the reliability of the variance estimates. Still, it is desirable to narrow these confidence intervals, which reconfirms the need to better tailor the models to the complexity of the system being described by the addition of factors not included in the forecast procedures commonly used today.

Table 67. Estimates of total run and an 80 percent confidence interval based on a weighted average of two estimates for each age class return.

Year of return	Estimated total return	80% confidence limits Upper limit	Lower limit	Standard deviation
1971	2,245,763	3,666,672	824,854	1,004,176
1972	1,334,935	2,647,238	22,632	939,372
1973	1,331,088	2,507,028	155,148	850,282
1974	1,184,459	2,450,533	0	922,795
1975	1,575,569	2,787,973	363,165	889,511
1976	1,333,055	2,465,010	201,100	834,775
1977	1,676,630	2,723,302	629,958	775,313
1978	1,727,733	2,761,267	694,199	768,427
1979	2,552,317	3,547,932	1,556,702	742,442
1980	2,892,710	3,981,113	1,804,307	814,064
1981	5,699,018	7,255,846	4,142,190	1,167,913

SUMMARY AND CONCLUSIONS

Three estimation procedures were investigated as to their variance components and forecast reliability using data from the Egegik sockeye salmon run. A spawner-recruit model which estimated age-specific recruits was proposed, circumventing the need to apportion total return into age classes and possessing a more comprehensive variance estimate. The regression of (N + 1)-ocean return with N-ocean of the same brood year often resulted in a poor fit to historic data. An improvement was seen in the incorporation of a variable representing the proportion of N-ocean fish in the escapement of that brood year. Lastly two methods of estimating the ratio used to relate (N + 1) to N-ocean age return were presented. The average ratio is that commonly used, yet the ratio of the means proved to be of greater predictive value with lower cumulative percent error and smaller variances for the estimates. The 95 percent confidence intervals placed about hindcast estimates from each estimator and age class were too wide to be of any predictive value and reconfirm the need to develop models which explain more of the variation in the level of return of a sockeye salmon stock.

Criteria of estimator performance were defined to aid in the choice of estimates to be incorporated into a final point forecast. The residual variance of the forecast (RVF) was defined and its reciprocal used in developing a weighted average as a final estimate. Again, hindcasting data were used to evaluate the different averaging methods. The best approach in developing a single point forecast was a weighting procedure based on two estimates. One involved the spawner-recruit estimator and the other an (N + 1)-ocean estimator chosen on the basis of its hindcasting performance. This approach resulted in total run estimates with the lowest cumulative error. Eighty percent confidence intervals about the estimates were of a useful width only in the last few years.

The estimation methods and evaluation techniques developed within this research can be applied to any sockeye salmon stock or salmon species where forecasting is presently based on spawner-recruit, regression, or ratio methods. Yearly evaluation of the RVF of each estimator could also reveal reactions to a changing system in extreme changes in magnitude or the continued poor performance of an estimator, which would suggest its removal from the final estimation process.

LITERATURE CITED

Aitchison, J., and J. Brown. 1957. The log-normal distribution. Cambridge: Cambridge Univ. Press, 170 pp.

Cochran, W. G. 1977. Sampling techniques. New York: John Wiley & Sons, Inc., 3rd edition, 428 pp.

Draper, N. R., and H. Smith. 1980. Applied regression analysis. New York: John Wiley & Sons, Inc., 2nd edition, 709 pp.

Larkin, P. A., R. F. Raleigh, and N. J. Wilimovsky. 1964. Some alternative premises for constructing theoretical reproduction curves. J. Fish. Res. Bd. Can. 21(3):477-484.

Ludwig, D., and C. J. Walters. 1981. Measurement error and uncertainty in parameter estimates for stock and recruitment. Can. J. Fish. and Aquat. Sci. 38(6):711-720.

Mathews, S. B. 1967. The economic consequences of forecasting sockeye salmon (Oncorhynchus nerka, Walbaum) runs to Bristol Bay, Alaska: a computer simulation study of the potential benefits to a salmon canning industry from accurate forecasts of the runs. Ph.D. Dissertation. Univ. of Washington, Seattle, 238 pp.

Ossiander, F. J. (ed.). 1967. Bristol Bay red salmon forecast of run for 1967. Alaska Dept. Fish and Game, Info. Leaflet No. 105, 52 pp.

Pennoyer, S. (ed.). 1970. Final forecast of 1970 Bristol Bay sockeye run. Alaska Dept. Fish and Game, Info. Leaflet No. 146, 65 pp.

Ricker, W. E. 1954. Stock and recruitment. J. Fish. Res. Bd. Can. 11:559-623.

Sokol, R. R., and F. J. Rohlf. 1969. Biometry. San Francisco: W. H. Freeman and Co., 776 pp.

Ward, F. J., and P. A. Larkin. 1964. Cyclic dominance in Adams River sockeye salmon. Intern. Pac. Salm. Fish. Comm., Prog. Rept. No. 11, 116 pp.

APPENDIX A

Derivation of the Estimated Variance for the Log of Predicted Total Return from the Spawner-Recruit Model

The spawner-recruit model and underlying assumptions were discussed in the section involving the spawner-recruit estimator. Let the model notation persist here for the derivation of the variance of our estimate Y, where

$$Y = a + bX_1 + cX_2 + e$$

From regression theory, a is defined as

$$a = \overline{Y} - b\overline{X}_1 - c\overline{X}_2$$

then

$$Y = \overline{Y} - b\overline{X}_1 - c\overline{X}_2 + bX_1 + cX_2 + e$$

$$= \overline{Y} + b(X_1 - \overline{X}_1) + c(X_2 - \overline{X}_2) + e$$

The variance of Y becomes

$$V(Y) = V[\overline{Y} + b(X_1 - \overline{X}_1) + c(X_2 - \overline{X}_2) + e]$$

$$= V(\overline{Y}) + (X_1 - \overline{X}_1)^2 V(b) + (X_2 - \overline{X}_2)^2 V(c) + \sigma^2$$
$$- 2(X_1 - \overline{X}_1)(X_2 - \overline{X}_2) \text{Cov}(b, c)$$

$$= \sigma^2 \left[\frac{1}{n} + 1\right] + (X_1 - \overline{X}_1)^2 V(b) + (X_2 - \overline{X}_2)^2 V(c)$$
$$- 2(X_1 - \overline{X}_1)(X_2 - \overline{X}_2) \text{Cov}(b, c)$$

Note that

$$\mathrm{Cov}(\overline{Y},\ b) \;=\; \mathrm{Cov}(\overline{Y},\ c) \;=\; 0$$

as does any covariance term involving the error (e). The coefficients, their variances and covariances are then estimated by regression and the variance of the predicted Y becomes

$$V(\hat{Y}) \;=\; S^2_{Y\cdot X}\,[1 + \frac{1}{n}] + (X_1 - \overline{X}_1)^2\, V(\hat{b}) + (X_2 - \overline{X}_2)^2\, V(\hat{c})$$
$$- \; 2(X_1 - \overline{X}_1)(X_2 - \overline{X}_2)\ \mathrm{Cov}(\hat{b},\ \hat{c})$$

APPENDIX B

A Maximum Likelihood Estimate for the Variance of the Log-Normal Distribution

If $\log(Y_i)$ is normally distributed as $N(M, \sigma^2)$, then Y_i has a log-normal distribution of log-normal (A, B^2). The variance of Y_i (B^2) can then be approximated by the following maximum likelihood estimate.

$$\hat{b}^2 = \exp(2m)[f_n(2S_m^2) - f_n(\frac{n-2}{n-1} S_m^2)]$$

where

$$m = \frac{1}{n} \sum_{i=1}^{n} \log(Y_i)$$

$$S_m^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\log(Y_i) - m)^2$$

and

$$f_n(t) = \exp(t)[1 - t(\frac{t+1}{n}) + \frac{t^2(3t^2 + 22t + 21)}{6n^2}]$$

(Aitchison and Brown 1957).

APPENDIX C

Comparison of Spawner-Recruit Models

The usual spawner-recruit model, the so-called Ricker curve, can be written as follows:

$$r_i = s_{i-1} e^{a+bs_{i-1}}$$

or sometimes modified as a three-parameter model

$$r_i = s_{i-1}^c e^{a+bs_{i-1}}$$

The basic observations used with either version are a set of observed spawners entering a river (s_i) and a set of catches from the same stock. The spawners plus the catch make up the recruits (r_i). The data are the pairs (r_i, s_{i-1}) where the s_{i-1} produce the r_i's.

There are two types of error associated with the model, (1) errors in measuring the number of returning spawners and catches and (2) the error of lack of fit of the model, this is called process error. The process error results from the failure of the model to perfectly account for all aspects of the biological process. The two errors have been modelled in several different ways:

I $$r_i = s_{i-1}e^{a + bs_{i-1} + e_i}$$

$e_i \sim N(0, \sigma^2)$ models the process error,

r_i and s_{i-1} are assumed to be measured without error.

II $$r_i = s_{i-1}e^{a + bs_{i-1}} + e_i$$

$e_i \sim N(0, \sigma^2)$ models the process error,

r_i and s_{i-1} are considered measured without error.

The only difference in the two models is that I has log-normal error and II has normal error. Process error can never be replicated, since only one observation per year is available; the choice of distribution is strictly arbitrary. In fact, while there are components to the error that may be modelled with a random variable, there is certainly a large amount of systematic error, since the model is an oversimplification of the true biological situation. Modeling the error is more a matter of convenience for deriving estimators for the parameters than for depicting real biological phenomena.

III Ludwig and Walters (1981).

Ludwig and Walters (1981) have added measurement error to the spawner counts to the model. Their model is formulated as

$$r_i = s_{i-1}e^{a+bs_{i-1} + u_i}$$

u_i are independent identically distributed (I.I.D.N.) normal random variables with variance σ_u^2 and mean zero.

The observed counts are

$$R_i = r_i e^{v_i} \quad \text{and} \quad S_i = s_i e^{v_i}$$

v_i are I.I.D.N. with mean zero and variance σ_v^2.

Modeling measurement error, which is known to occur, adds a further element of realism to the general model, but there is no particular reason why the measurement error should be modeled with the log-normal distribution. A variety of other distributions could have been used with equal acceptability of depicting the true state of nature. The particular structure used by Ludwig and Walters complicates the estimation of the parameters. This is because with normal random variables the parameters of the model are not identifiable. That is, different parameter values can occur with the identical probability distribution of the data. When this happens there is no point in estimating the parameters.

In order to overcome this problem, Ludwig and Walters assume the same variance σ_v^2 for counts from different years, and assume the process error and measurement error, as measured by σ_u^2 and σ_v^2 have a known ratio,

$$\lambda = \sigma_v^2 / \sigma_u^2$$

With these assumptions, a and b are identifiable.

Since λ is never known and can only be guessed at, there is the same uncertainty as regards their parameter estimates.

Estimation

Model I parameters are estimated by taking the logarithm of both sides of the equation and using least squares (regression) estimation. Since no measurement error is acknowledged, and yet measurement error is known to be present, the good properties of least squares do not apply. The measurement errors must be modelled before any study of the departure from these properties can be made.

Model II parameters are estimated by non-linear least squares. The estimators are also maximum likelihood estimators if no measurement error is present. The properties of these estimators under a measurement error structure is unknown.

Model III (Ludwig and Walters)

The estimators developed by Ludwig and Walters are unusual. They claim that v_i and u_i are random variables when specifying the model but call them parameters when they begin estimation. They use a likelihood function to estimate the random variables v_i and u_i, and then they use the estimates to estimate σ_u^2. This is not usual maximum likelihood technique; the properties of this procedure are entirely unknown and are not necessarily any better than least squares with no measurement error.

Summary

The statistical properties of the estimators of these three models are unknown. To determine which method might be preferable, a large-scale monte carlo simulation should be undertaken.

In this report we used the regression approach and Ludwig and Walters procedure (C1 to C4) to estimate the number of returning age classes. These estimates were compared to the observed returns; Model I with parameters estimated by regression produced the best results.

Table C1. Comparison of estimates of age class 4(2) from a spawner-recruit relationship using the procedure of Ludwig and Walters (1981).

		Estimates of age class 4(2) from spawner-recruit relationship			
Brood year	Observed 4(2)	With $\lambda = 1.0$	% Error	With $\lambda = 1.5$	% Error
1967	55,739	56,813	1.9	54,843	-1.6
1968	37,734	33,335	-11.7	31,721	-15.9
1969	12,429	97,020	680.6	96,520	676.6
1970	58,537	87,743	49.9	87,240	49.0
1971	44,185	64,695	46.4	63,466	43.6
1972	54,918	52,358	-4.7	50,862	-7.4
1973	73,702	32,267	-56.2	30,700	-58.3
1974	129,275	103,135	-20.2	105,772	-18.2
1975	144,195	93,937	-34.9	95,551	-33.7
1976	582,296	53,530	-90.8	51,989	-91.1
1977	793,985	67,978	-91.4	67,692	-91.5
		$\Sigma =$	468.9	$\Sigma =$	451.5
		Σ |error| =	1,088.7	Σ |error| =	1,086.9

Table C2. Comparison of estimates of age class 5(2) from a spawner-recruit relationship using the procedure of Ludwig and Walters (1981).

		Estimates of age class 5(2) from spawner-recruit relationship			
Brood year	Observed 5(2)	With $\lambda = 1.0$	% Error	With $\lambda = 1.5$	% Error
1966	664,641	137,870	-79.3	135,011	-79.7
1967	220,628	107,041	-51.5	103,705	-53.0
1968	41,940	87,554	108.8	83,419	98.9
1969	105,183	219,756	108.9	219,734	108.9
1970	86,167	194,148	84.6	193,995	84.4
1971	106,255	134,950	27.0	131,748	24.0
1972	58,338	112,291	92.5	108,438	85.9
1973	127,677	72,143	-43.5	66,426	-48.0
1974	97,738	217,531	122.6	223,526	128.7
1975	236,559	205,859	-13.0	205,435	-13.2
1976	775,728	111,624	-85.6	110,398	-85.8
		$\Sigma =$	271.5	$\Sigma =$	251.1
		Σ |error| =	817.3	Σ |error| =	810.5

Table C3. Comparison of estimates of age class 5(3) from a spawner-recruit relationship using the procedure of Ludwig and Walters (1981).

Brood year	Observed 5(3)	Estimates of age class 5(3) from spawner-recruit relationship			
		With $\lambda = 1.0$	% Error	With $\lambda = 1.5$	% Error
1966	604,374	1,202,145	98.9	1,177,219	94.8
1967	633,550	1,027,887	62.2	995,846	57.2
1968	118,779	578,290	386.9	550,977	363.9
1969	1,079,491	1,322,817	22.5	1,322,701	22.5
1970	787,985	1,121,447	42.3	1,120,560	42.2
1971	1,430,466	808,241	-43.5	789,064	-44.8
1972	1,516,911	701,514	-53.8	677,441	-55.3
1973	568,833	470,066	-17.4	432,814	-23.9
1974	2,178,973	1,489,340	-31.6	1,530,086	-29.8
1975	2,281,263	1,387,700	-39.2	1,384,845	-39.3
1976	2,886,554	799,003	-72.3	790,228	-72.6
		Σ =	355.0	Σ =	314.9
		Σ |error| =	870.6	Σ |error| =	846.3

Table C4. Comparison of estimates of age class 6(3) from a spawner-recruit relationship using the procedure of Ludwig and Walters (1981).

Brood year	Observed 6(3)	Estimates of age class 6(3) from spawner-recruit relationship			
		With $\lambda = 1.0$	% Error	With $\lambda = 1.5$	% Error
1965	911,224	1,024,184	12.4	1,059,707	16.3
1966	797,755	593,725	-25.6	691,954	-13.3
1967	555,019	600,289	8.2	582,934	5.0
1968	296,573	347,566	17.2	333,931	12.6
1969	1,101,694	860,013	-21.9	864,649	-21.5
1970	169,666	816,094	381.0	809,757	377.3
1971	926,825	542,760	-41.4	527,937	-43.0
1972	1,193,895	501,440	-58.0	473,280	-60.4
1973	842,214	348,356	-58.6	325,711	-61.3
1974	481,564	1,034,147	114.7	1,036,319	115.2
1975	776,539	958,943	23.5	972,174	25.2
		Σ =	351.5	Σ =	352.1
		Σ |error| =	762.5	Σ |error| =	751.4

APPENDIX D

Transformations of the Ratio Model for Conformation to the Gauss-Markov Theorem

The Mean of Ratios Estimator

Consider the model discussed in the ratio estimator section:

$$Y_i = R\,X_i + e_i$$

where

$$E(e_i) = 0$$

$$V(e_i) = X_i^2\,\sigma^2$$

$$E(e_i\,e_j) = 0 \quad \text{for all } i,\ j,\ \ i \neq j$$

The transformation to conform to the Gauss-Markov (G-M) Theorem is

$$e_i' = e_i / X_i$$

$$Y_i' = Y_i / X_i$$

$$X_i' = X_i / X_i = 1$$

General theory then provides the following unbiased estimators (all sums over the sample of size n):

$$\hat{R} = \frac{\Sigma\, X_i'\, Y_i'}{\Sigma\, X_i'^2} = \frac{1}{n}\,\Sigma\,\frac{Y_i}{X_i}$$

and

$$S^2_{Y \cdot X} = \frac{1}{n-1} \Sigma (Y_i' - \hat{R} X_i')^2$$

$$= \frac{1}{n-1} \left(\Sigma \left(\frac{Y_i}{X_i}\right)^2 - n \hat{R}^2 \right)$$

The Ratio of Means Estimator

Consider the model discussed in the ratio estimator section:

$$Y_i = R X_i + e_i$$

where

$$E(e_i) = 0$$

$$V(e_i) = X_i \sigma^2$$

$$E(e_i e_j) = 0 \quad \text{for all } i, j, \; i \neq j$$

The appropriate transformation to allow the use of general least square theory (G-M Theorem) is

$$e_i' = e_i / \sqrt{X_i}$$

$$Y_i' = Y_i / \sqrt{Y_i}$$

$$X_i' = \sqrt{X_i}$$

General theory then provides the following unbiased estimates (all sums over the sample of size n):

$$\hat{R} = \frac{\Sigma X_i' Y_i'}{\Sigma X_i'^2} = \frac{\Sigma Y_i}{\Sigma X_i}$$

and

$$S^2_{Y \cdot X} = \frac{1}{n-1} [\Sigma (Y_i' - \hat{R} X_i')^2]$$

$$= \frac{1}{n-1} (\Sigma \frac{Y_i^2}{X_i} - \hat{R} \Sigma Y_i)$$

APPENDIX E

Transformation of the Return to Date Estimator for Conformation to Ratio Estimation Theory

The return to date (RTD) method for estimating age class 6(3) was derived as

$$RTD_i = 3 + 4 + 5 \text{ year old return from brood year } (i)$$

Model: $$\hat{6}(3)_i = \frac{RTD_i}{P} - RTD_i$$

If P is a function of the proportion of RTD to total return as

$$P = f\left(\frac{RTD}{RTD + 6(3)}\right)$$

therefore

$$6(3)_i = RTD_i \left(\frac{1}{p} - 1\right)$$

$$= RTD_i \left(\frac{1}{\frac{RTD}{RTD + 6(3)}} - 1\right)$$

$$= RTD_i \frac{(RTD + 6(3) - RTD)}{RTD}$$

$$= RTD_i \frac{(6(3))}{RTD}$$

which implies $P = \frac{(6(3))}{RTD}$ and can be treated as a ratio (R) in the section on ratio estimators.